Regulating Multiparty Persuasion with Bipolar Arguments: Discussion and Examples

Dionysios Kontarinis⋆ Elise Bonzon⋆ Nicolas Maudet† Pavlos Moraitis⋆

⋆LIPADE
Université Paris Descartes
45 rue des Saints-Pères
75270 Paris 06, France
dionysios.kontarinis@parisdescartes.fr
{elise.bonzon,pavlos}@mi.parisdescartes.fr

†LAMSADE
Université Paris Dauphine
75775 Paris Cedex 16, France
maudet@lamsade.dauphine.fr

Résumé :
Réguler un débat entre plus de deux agents est un problème difficile. Certains travaux récents ont proposé des protocoles adaptés à des agents dont le raisonnement est modélisé comme un système abstrait d'argumentation de Dung. Dans cet article, nous discutons informellement des questions qui se posent lorsque l'on considère des systèmes plus expressifs permettant de représenter par exemple attaque et support envers un argument, en particulier dans un cadre d'évaluation graduelle de ces arguments. Nous voyons en particulier que la notion de pertinence peut être interprétée de différentes manières, donnant lieu à différents types de protocoles. La discussion est illustrée par de nombreux exemples.

Mots-clés : Systèmes multi-agents, argumentation, persuasion

Abstract :
The problem of regulating multiparty persuasion among several (more than two) self-interested agents has recently been put forward in a series of papers. Most of these works are settled in the original Dung's framework of argumentation. In this paper, we discuss informally the challenges that arise when one wants to consider more expressive bipolar argumentation frameworks using a gradual evaluation of arguments. We see in particular that the notion of relevance can be interpreted in different ways, giving rise to different types of protocols. The discussion is illustrated by a number of examples.

Keywords : Multi-agent systems, argumentation, persuasion

1 Introduction

Argumentation is a sophisticated and flexible approach to model an agent's (non-monotonic) reasoning. In particular, the abstract approach advocated by Dung [11] as a way to identify sets of acceptable arguments has given rise to a huge amount of work. While there are concerns regularly expressed that this abstract view may not be adequate to model "real argumentation", and that simply plugging this on top of some logic-based language may not be appropriate [1], the emergence of platforms allowing users or agents to directly express their view under the form of argument graphs (see for instance DebateGraph [1]) is a strong motivation to further study this framework. Until recently the main focus has been on *single agent* reasoning, with few works investigating protocols for bilateral interaction, in particular negotiation [13]. Recently, in a series of papers [16, 17], Rahwan and Larson have put forward a research agenda at the interface of game theory and argumentation. A related line of work is also pursued by Caminada and Pigozzi [5]. Finally, Konieczny et al. [9] have studied the problem of merging a set of argumentation systems, hence giving a clean way to define what could be a social outcome in a society of agents equipped with (potentially different) Dung's systems. This leaves open the question of how such debates should be regulated in practice. This is not a trivial question in a multiparty setting [10, 19]. Very recently, an intuitive protocol building on the idea of relevance of moves introduced by Prakken [14] was proposed [4]. It was shown in particular that the protocol allows debates that are not fully determined from the initial situation to occur. However, the setting remains unsatisfying in the sense that it sticks to the basic version of abstract argumentation systems, and does not allow for instance to express that an argument supports another argument (while this is typically allowed in DebateGraph). Several works have extended the framework of Dung by allowing a support relation to be explicitly represented. We follow the works of Cayrol and Lagasquie-Schiex [8] and consider a gradual evaluation setting, where a value can be assigned to each argument. This value is supposed to capture

1. see http ://debategraph.org

"the relative strength of an argument taking into account the undercuts, undercuts to undercuts, and so on" [3, p.24]. It is important to understand that this constitutes an important differenciation from the "acceptability-based" approach *à la* Dung, as indeed the valuations may not be in line with the output provided by the different semantics of Dung (we refer to [7] for a detailed discussion on these aspects). For what is our main concern here, we will see that the valuation-based approach is challenging : in brief, it may not be so easy in this case to classify agents as being "for" (PRO) or "against" (CON) an issue as it is done in [4]. This questions how to better regulate such debates. The purpose of this paper is simply to put forward these issues, and to do this with the help of many illustrative examples. We discuss the many different design choices that are possible for such protocols, and deliberately chose two of them as support for our examples.

The remainder of this paper is as follows. In Section 2 we set up the background for this paper, introducing in particular how bipolar argumentation frameworks can be naturally handled with gradual evaluation techniques. We also define what it means to merge several such argumentation systems. In Section 3 we get to the core of this paper and discuss various design options for multiparty protocols in the setting described. We give the details of two protocols that are then illustrated in Section 4 through different examples. Section 5 concludes and discusses the perspectives opened by this work.

2 Background

In this section, we briefly recall some elements of abstract bipolar argumentation frameworks using a gradual evaluation of arguments as defined in [6, 8, 2].

2.1 Bipolar argumentation

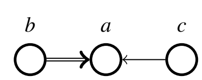
The principle of bipolar argumentation is to use two types of relations between arguments. The first one is, as in the classical framework proposed by [11], a *defeat* relation (denoted R). This binary relation between arguments indicates that the first argument attacks the second one. If argument b defeats argument a (bRa) then b is a reason to believe that a does not hold. The second type of binary relation between arguments introduced in this framework is a *support* relation (denoted S), which indicates that the first argument supports the second one. So, if argument b supports argument a (bSa) then b is a reason to believe that a holds. It must be noted that the support relation is different than the defence relation defined in [11], because the notion of defence depends on the notion of defeat, while on the other hand, the notions of support and defeat are independant. Support relations between arguments are often used in everyday life and this makes them particularly interesting. An argumentation framework using both defeat and support relations is called *bipolar argumentation framework (BAF)* [6, 8, 2].

Abstract Bipolar Argumentation Framework. An abstract bipolar argumentation framework (denoted BAF because we are only using abstract frameworks in the present work) is an extension of the basic abstract argumentation framework introduced in [11] in which both defeat and support relations are used. As in [11], we are not interested in the structure of its arguments.

A BAF is a triplet $\langle A, R, S \rangle$ which consists of a set of arguments A, a defeat relation R on A, and a support relation S on A. If $a, b \in A$, then aRb (resp. aSa) means that a defeats (resp. supports) b.

A BAF may be represented by a directed graph $\mathcal{G}_b$ called *bipolar interaction graph*. This graph has two types of edges. The double arrows edges will represent the defeat relation whereas the simple arrows will represent the support relation between arguments.

Example 1 *The* $BAF = \langle \{a, b, c\}, \{(b, a)\}, \{(c, a)\} \rangle$ *is represented by the following bipolar interaction graph. The argument* a *is defeated by* b *and supported by* c*.*



A BAF is said *well-founded* if and only if there is no infinite sequence of arguments $a_0, a_1, \dots a_n, \dots$ such that $\forall i, a_i \in A$ and $(a_{i+1}, a_i) \in R$ or $(a_{i+1}, a_i) \in S$.

Notations : Let $a \in A$. The set $\{b \in A | bRa\}$ is denoted by $R^-(a)$. Its elements are called *direct defeaters* of a. The set $\{b \in A | aRb\}$ is denoted by $R^+(a)$. In the same way, we define $S^-(a)$ (the *direct supporters of* a) and $S^+(a)$.

Valuation in a Bipolar Argumentation Framework. The authors of [6] propose two different processes for gradual valuation. In the first process, called *local gradual* valuation, the value of an argument only depends on the values of its direct defeaters and supporters. In the second process, called *global gradual* valuation, the value of an argument represents the set of all the branches leading to this argument. In our present work we will be focusing on local gradual valuation processes.

According to [6], there are three basic principles for local gradual valuation processes :

P1 The valuation of an argument is a function of its direct defeaters and of its direct supporters.

P2 If the quality (see Definition 1) of the support (resp. defeat) increases then the value of the argument increases (resp. decreases).

P3 If the quality of the supports (resp. defeats) increases then the quality of the support of an argument increases (resp. decreases).

In respect of the previous principles, [6] assumes that there exists a completely ordered set $\mathcal{V}$ with a minimum element (V_{Min}) and a maximum element (V_{Max}) and proposes the following formal definition for a local gradual valuation.

Definition 1 [6]

Let $\langle A, R, S\rangle$ *be a bipolar argumentation framework. Consider* $a \in A$ *with* $R^{-}(a) = \{b_1, \ldots, b_n\}$ *and* $S^{-}(a) = \{c_1, \ldots, c_p\}$*. A* local gradual valuation *on* $\langle A, R, S\rangle$ *is a function* $v\colon A \rightarrow \mathcal{V}$ *such that :* $v(a) = g(h_S(v(c_1), \ldots, v(c_p)), h_R(v(b_1), \ldots, v(b_n)))$ *where*

- *The function* h_R *(resp.* h_S*) :* $\mathcal{V}^* \rightarrow H_R$ *(resp.* $\mathcal{V}^* \rightarrow H_S$*) valuates the quality of the defeat (resp. support) on* a*.* $\mathcal{V}^*$ *denotes the set of the finite sequences of elements of* $\mathcal{V}$*, including the empty sequence, and* H_R*,* H_S *are ordered sets.*
- *The function* $g : H_S \times H_R \rightarrow \mathcal{V}$ *with* $g(x, y)$ *increases on* x *and decreases on* y*.*

Note that the definition above produces a generic local gradual valuation. There exist several instances for this generic valuation. We will use the following on the remainder of this paper :

- $\mathcal{V} = [-1, 1]$ interval of reals ;
- $H_R = H_S = [0, \infty]$ interval of reals ;
- $h_R(x_1, \ldots, x_n) = h_S(x_1, \ldots, x_n) = \sum_{i=1}^{n} \frac{x+1}{2}$;
- $g(x, y) = \frac{1}{1+y} - \frac{1}{1+x}$.

The functions h_R and h_S represent the two axes of an unipolar bivariate scale. Then, with the function g, we obtain a bipolar univariate scale. So, after using g, we cannot distinguish the case "balance between defeats and supports" and the case "no defeat and no support".

2.2 Merged Bipolar Argumentation Frameworks

The notion of *merged argumentation system* has been introduced in [9] for classical abstract argumentation frameworks. A meaningful way to merge several argumentation systems which share exactly the same arguments but with possible conflicting views on the attack relations between them is to take the *majority argumentation system*. In this system, defined in [9], attacks supported by the majority of agents are kept, and ties are broken in favour of the absence of the attack.

We propose here a similar process for the bipolar argumentation systems. We consider a set N of n agents, each one having an abstract bipolar argumentation system BAF_i=$\langle A, R(i), S(i)\rangle$. The bipolar argumentation systems share exactly the same arguments, but with possible conflicting views on the attack and support relations between them. The merged bipolar argumentation framework (denoted $MBAS_N$) is then defined in the following way :

Definition 2 *Let* N *be a set of* n *agents and* $\langle BAF_1, \ldots, BAF_n\rangle$ *be the collection of their bipolar argumentation systems. Let* $Defeat(a, b) = \{i \in N | (a, b) \in R(i)\}$ *and* $Support(a, b) = \{i \in N | (a, b) \in S(i)\}$*. Then, the* merged bipolar argumentation system *is* $MBAS_N = \langle A, M, T\rangle$ *where* $M \subseteq A \times A$*,* $T \subseteq A \times A$ *and*

- aMb *iff* $|Defeat(a, b)| > n - |Defeat(a, b)|$
- aTb *iff* $|Support(a, b)| > n - |Support(a, b)|$

3 Designing Protocols

We now assume that we are in the same line of questions as [4] : first, we assume that all agents are *focused* on the same single issue (argument) of the debate (that is, agents evaluate how good is a state of the debate on the sole basis of the valuation of this single argument). Then, we suppose that the agents of the system would not report to a central authority their whole argumentation system, but instead contribute step-by-step in the debate, guided by their individual assessment of the current state of the discussion, and without coordination with other agents. What would be the outcome they would reach ? For instance, can we guarantee that the merged outcome would always be reachable ? To be able to formally give an answer to this problem, we need of course to design a specific protocol and to make some assumptions regarding agents' preferences on the outcome. A key feature of multi-party protocol is to come up with a well-defined way to allocate turn-taking among agents. In fact, when we inspect more carefully the issue, we see that with gradual valuations different approaches can be chosen. Suppose indeed that the issue of the debate is d, and that we are in presence of four agents such that $v_1(d) = -0.9, v_2(d) = -0.1, v_3(d) = +0.1, v_4(d) = +0.4$.

Taking account of these assumptions, we propose two main types of protocols :

– in *category-based protocols*, the society of agents can be divided in k predefined categories $K_1 \ldots K_k$ (we avoid the word "team" because, again, no coordination takes place among agents of the same group) partitioning the range of values. In the extreme case where $k = 2$ we can see the debate as opposing PRO and CON such that : $CON = \{a_i \in N | v_i(d) < \alpha\}$ and $PRO = \{a_i \in N | v_i(d) \geq \alpha\}$. By setting $\alpha = 0$ we would have $CON = \{a_1, a_2\}$ and $PRO = \{a_3, a_4\}$.
– in *cluster-based protocols*, no category of agents is defined beforehand, but agents may be clustered in k groups $C_1 \ldots C_k$ (not necessarily partitioning the range of values) on the basis of the similarity of their views (in the sense of minimizing the max difference between two agents of a group). For instance, with $k = 3$ we have three clusters $C_1 = \{a_1\}, C_2 = \{a_2, a_3\}, C_3 = \{a_4\}$ (but note that for $k = 2$ we have $C_1 = \{a_1\}, C_2 = \{a_2, a_3, a_4\}$). Here, an interesting extreme case is that of n clusters, meaning that each group is a singleton, and that each agent is solely guided by the prospect of matching his value with that of the gameboard.

We make no claim on the relative merits of these approaches, and believe that they may be justified in different contexts. If there is an exogenously given meaning to threshold, then dispatching agents in the adequate categories seems appropriate. If the purpose is instead to regulate the dialogue in such a way that the diversity of views is fairly represented, then it may be more appropriate to opt for a cluster-based approach. For illustrative purposes, we will focus our attention in this paper on two extreme versions of these types of protocols, as already mentioned : a category-based protocol with $k = 2$ (π_0), and a cluster-based protocol with $k = n$ (π_1). Note that π_0 is more in line with persuasion dialogues which confront two opposing views (although Prakken explicitely mentions the possibility to have a third party of undecided agents), and clearly follows from the fact that these protocols are based on the notion of *acceptability* of an argument. On the other hand, π_1 is more coherent with the view of the value as appropriate measure for the problem considered.

3.1 The Gameboard (GB)

Here, we introduce the notion of gameboard, already presented in [4]. Agents will exchange arguments via a common gameboard. The issue will be assumed to be present on this gameboard when the debate begins. The "common" argument system is therefore a *weighted (bipolar) argumentation system* [12] where the weight is simply a number equal to the difference between the number of agents who asserted a given attack and the number of agents who opposed it. We denote by $xR_\alpha y$ (resp. $xS_\alpha y$) the fact that the attack relation (resp. support relation) has a weight α. Let $A(GB)$ be the set of all the arguments present on the gameboard. The collective outcome is obtained by applying the semantics used on the bipolar argumentation system $\langle A(GB), M, T\rangle$ where $M, T \subseteq A(GB) \times A(GB), xMy = \{xR_\alpha y | \alpha > 0\}$ and $xTy = \{xS_\beta y | \beta > 0\}$. In words, we only retain those attacks and support relations supported by a (strict) majority of agents having expressed their view on this relation. Observe that following our tie-breaking policy we require the number of agents supporting the relation to strictly overweight the number of agents who oppose it (*i.e* in case of tie, the relation does not hold).

We abuse notation and denote by $v^t_{GB}(d)$ the value, at time t, of the argument d given the system present on the gameboard. When a group of agents cannot move, we say that the gameboard is stable and we refer to $v_{GB}(d)$ as the *outcome of the debate*.

3.2 Agents' Preferences

We assume that agents are *focused* [18], that is, they concentrate their attention on a specific (same for all) argument. This argument is referred to as the *issue* d of the debate [15]. Now, depending on the type of debate considered, preferences of agents have to be defined accordingly. We will define preferences with respect to *attraction values* representing the value which the agents of a group prefer.

In principle, in category-based protocols, agents of category K_j prefer any state of the debate such that the value of the issue is in K_j, regardless of the actual value reached. Alternatively, since the objective is to make sure that the value reached lies in the interval [2] $K_j = [K_j^l, K_j^u]$, agents of this group might prefer $(K_j^u + K_j^l)/2$ (except for K_1 and K_k for which the extreme values -1 and 1 are preferred) as the attraction value, since it maximizes the distance to any other interval.

In cluster-based protocols, the interpretation of the attraction value is different : they seek to make the value of the issue on the board as close as possible from the *representative value* of their cluster. That may be defined in different ways : eg. the mean or the median of agents' values, etc.

Note that a move which occurs at round t yields a value $v_{GB}^{t+1}(d)$ after occurrence on the gameboard GB. Getting back to our protocols π_o and π_1, the question is how shall we instantiate these notions ?

For π_1, the representative value is just $v_i(d)$, hence we simply have that a state at $t+1$ is preferred over a state at t (by agent i) when $|v_{GB}^{t+1}(d) - v_i(d)| < |v_{GB}^{t}(d) - v_i(d)|$. In words, each agent seeks to make the value of the GB as close as possible to its own value.

For π_0 (where we have only two categories), if attraction values are considered, we simply have that a state at $t+1$ is preferred over a state at t by PRO (resp. by CON) if $v_{GB}^{t+1}(d) > v_{GB}^{t}(d)$ (resp. $v_{GB}^{t+1}(d) < v_{GB}^{t}(d)$).

These definitions yield very different behaviours : for instance, suppose an agent i holds the view that $v_i(d) = -0.1$. If the threshold α is set up to 0, then under the protocol π_0 i prefers a state of the debate with $v_{GB}(d) = -0.9$ to a state of the debate with $v_{GB}(d) = +0.1$.

2. K_j^l (resp. K_j^u) denotes the *lowest* (resp. *highest*) value of the interval K_j

3.3 Two protocols

We now introduce two specific protocols which allow agents to exchange their arguments in order to agree on the valuation of a specific argument d, the issue of the dialogue. These two protocols are built on the same underlying principle.

Let $AS^t(GB)$, $A^t(GB)$, $R^t(GB)$ and $S^t(GB)$ be respectively the argumentation system, the set of arguments, the set of attack relations and the set of support relations on the gameboard after round t. The protocols indeed proceed in rounds which alternate between the groups of agents. Within these groups though, no coordination takes place : the agents may for instance play asynchronously.

Permitted moves are simply positive assertions of attacks xRy (resp. support xSy) with $y \in A^t(GB)$, or contradiction of (already introduced) attacks (resp. supports) with $(x, y) \in R^t(GB)$ (resp. $(x, y) \in S^t(GB)$). Note that arguments are progressively added on the gameboard via these attacks or supports, and that it may not contain the whole set of arguments when the debate concludes.

How about relevance ? In a sense, any move which simply changes the current value of the issue may be deemed relevant, as it brings a modification of the current status of the issue under discussion. More specifically however, a move will be said to be *relevant* at round t if the updated gameboard is preferred by the group the agent who played the move belongs to. In other words, a move is relevant for a group if it makes the valuation closer to the attraction value of the group. Furthermore, the protocol prevents the repetition of similar moves from the same agent. To account for this, each agent a_i is equipped with a set $RP_i^t \subseteq \{(x, y) | x, y \in A\}$ (resp. $SP_i^t \subseteq \{(x, y) | x, y \in A\}$) which contains the attack relations (resp. support relations) or the *non*-attack relations [3] (resp. non-support relations) he has added on the gameboard until time t, in order to prevent him from adding twice the same relation.

When a (relevant) move is played on the gameboard, the update operation given in Tab. 1 takes place.

Note the asymmetry here : introducing a new

3. The non-attack and the non-support relations are the relations the agent does not possess in its own system.

(1) after an assertion xRy (resp. xSy)
  - if $xR_\alpha y \in R^t(GB)$ (resp. $xS_\alpha y \in S^t(GB)$) then $\alpha := \alpha + 1$
  - if $xR_\alpha y \notin R^t(GB)$ (resp. $xS_\alpha y \notin S^t(GB)$) and $x, y \in A^t(GB)$, then the edge is created with $\alpha := 1$
  - otherwise (x is not present), then the node of the new argument is created and the edge is created with $\alpha := 1$

(2) after a contradiction of xRy or of xSy, we have $\alpha := \alpha - 1$

TABLE 1 – Update of the gameboard

argument can only be done via a positive assertion, since it can never be relevant to contradict an attack referring to an argument that was not introduced already.

We now give the details of our two illustrative protocols. As mentioned before, π_0 is a two-sided category-based protocol. It is described in Tab. 2. On the other hand, π_1 is a cluster-based protocol with n agents playing with the objective to make the value of the issue matching their own valuation. It is described in Tab. 3.

It is important to notice the difference with the two-sided acceptability-based protocol proposed in [4]. In this latter protocol, the notion of relevance coincides with the condition for turn-taking : each relevant move passes the turn (by definition) to the other side. With π_0 and π_1, this is not the case. A move may be relevant but not sufficient to switch the turn. Similarly, a move may be relevant for a group which doesn't currently have the possibility to play.

Furthermore, we clearly see that different design choices can be made in many steps of these protocols. Relevant moves correspond to moves that update the gameboard in a way that is in line with the preferences of the agent, as previously discussed. When several such moves are proposed however, how shall the authority *select* which relevant move to pick ? Here, we chose for π_0 to let the central authority pick the first (or at random) move from *RelevantMoves*. For π_1, the function *select* chooses the move which changes the most the value of the issue on the gameboard. Of course, these choices are debatable and a lot of other choices could have been made.

Another important question is about the turn-taking : when the turn has to be given to a different side, and which one should be chosen (the problem doesn't occur with only two groups of course) ? Observe also that at the end of a two-sided protocol, the side winning the debate is $\overline{ToPlay}$.

(1) Agents report their individual view on the issue to the central authority, which then assigns (privately) each agent to PRO or CON.

(2) The first round starts with the issue on the gameboard ;

(3) $ToPlay \leftarrow CON$

(4) While $Winner = \{\,\}$ loop
  (a) agents of $ToPlay$ can independently propose moves to the authority (*ProposedMoves*) ;
  (b) the authority filters relevant moves only (*RelevantMoves*) ;
  (c) if $RelevantMoves = \{\,\}$ then $Winner \leftarrow \overline{ToPlay}$ else $PickedMove \leftarrow select(RelevantMoves)$
  (d) update the GB with *PickedMove*
  (e) if $sign(v^{t+1}_{GB}) \neq sign(v^t_{GB})$ then $ToPlay \leftarrow \overline{ToPlay}$

TABLE 2 – Category-based protocol with $k = 2$

(1) While $Winner = \{\,\}$ loop
  (a) agents can independently propose moves to the authority (*ProposedMoves*) ;
  (b) the authority filters relevant moves (*RelevantMoves*) ;
  (c) if $RelevantMoves = \{\,\}$ then $Winner \leftarrow argmin_i |v_i(d) - v_{GB}(d)|$ else $PickedMove \leftarrow select(RelevantMoves)$
  (d) update the GB with *PickedMove*.

TABLE 3 – Cluster-based protocol with $k = n$

3.4 Properties

We now present some properties that we wish to study on these protocols. The outcome $v(d_{AS(GB^\sigma_{t\to\infty})})$ resulting from a specific sequence of moves σ obeying one of these protocols will typically be compared with the result which would be obtained by merging the argumentation systems ($v(d_{MBAS})$). We may want to ensure different properties, but we typically have :

- *Termination*— trivially guaranteed by assuming finite argument systems and preventing move repetition.
- *Guaranteed convergence to the merged outcome*— requires *all* possible sequences of moves (in particular, regardless of the specific choice of the agent and of the move to pick, when several relevant moves are proposed to the authority) to converge to the merged outcome.
- *Reachability of the merged outcome*— requires *at least* one possible sequence of moves to reach the merged outcome.

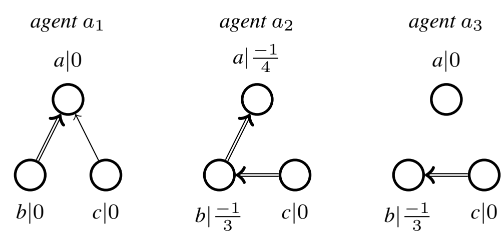
4 Illustrative examples

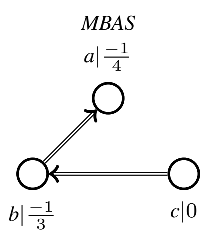
4.1 Protocol π_0

We recall that this protocol is the two-sided category-based one, with a threshold arbitrarily fixed at 0.

The first example shows that the outcome obtained in the MBAS can be unreachable with this protocol. Recall that double arrows represent the attack relations, whereas simple arrows reprensent the support relations.

Example 2 *Let three agents with their argumentation systems, and the following merged argumentation framework :*

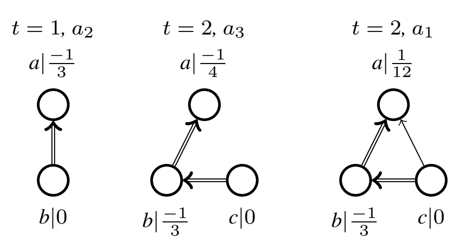


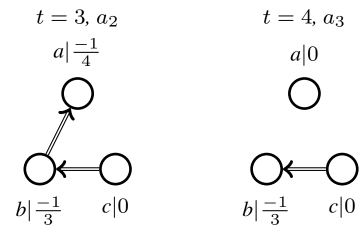


The issue of the dialogue is the argument a.

We have $PRO = \{a_1, a_3\}$, $CON = \{a_2\}$. At the begining, we have $RP_1^0 = RP_2^0 = RP_3^0 = \{\,\}$, $SP_1^0 = SP_2^0 = SP_3^0 = \{\,\}$ and $AS^0(GB) = \langle\{a\}, \{\,\}, \{\,\}\rangle$.

A sequence of moves allowed by this protocol is the following :



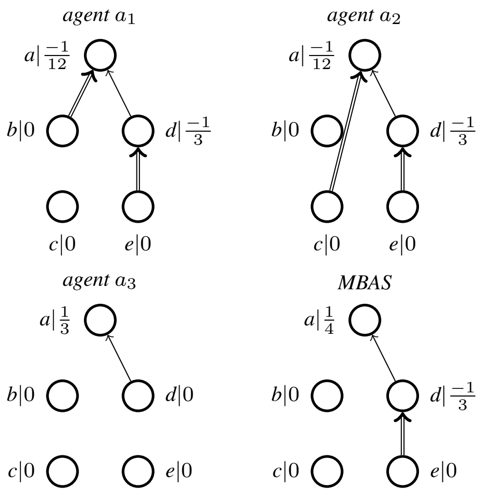


At $t = 1$, a_2 plays for CON and attacks the argument a with b. As the valuation of a is now negative, the token is given to PRO. a_3 attacks the argument b with c. This move increases the valuation of a, but this valuation remains negative. So, a_1 plays another move for PRO and adds a support relation between c and a. The valuation of a is positive, the token is given to CON. a_2 removes the support relation (c, a), and finally a_3 removes the attack between b and a. a_2 cannot move anymore. The gameboard is stable, the protocol stops with $v(a) = 0$.

The first interesting thing to observe on this simple example is the fact that CON agents cannot ensure the valuation of a to be negative. We leave to the reader to check that in this example, it is impossible to reach a gameboard in which the sign of the valuation of the issue is the same as the one obtained in the merged argumentation system. This is due to the fact that agent a_1 has no interest to play the attack relation (b, a), which appears in the MBAS. As studied in a different context by [16], as well as in [4], this can be seen as a strategic manipulation by withholding an argument or an attack between arguments.

The second example shows that the majority of agents does not always win, and that it is possible to have a garanteed convergence to the merged outcome.

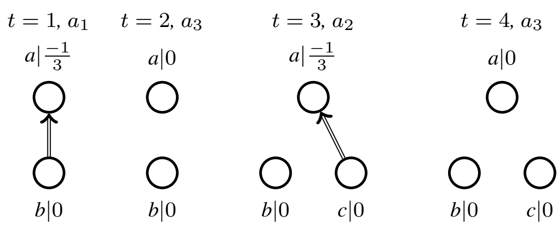
Example 3 *Let three agents with their argumentation systems, and the following merged argumentation framework :*



The issue of the dialogue is the argument a.

We have $PRO = \{a_3\}$, $CON = \{a_1, a_2\}$. At the begining, we have $RP_1^0 = RP_2^0 = RP_3^0 = \{\,\}$, $SP_1^0 = SP_2^0 = SP_3^0 = \{\,\}$ and $AS^0(GB) = \langle\{a\}, \{\,\}, \{\,\}\rangle$.

A sequence of moves allowed by this protocol is the following :

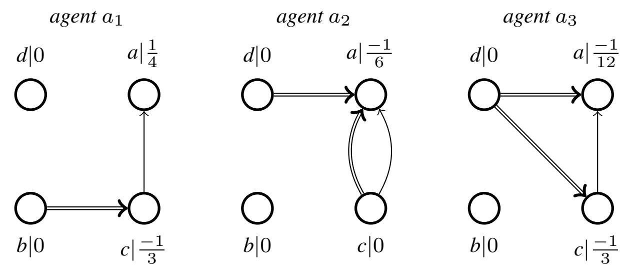


At $t = 1$, a_1 plays for CON and attacks the argument a with b. As the valuation of a is now negative, the token is given to PRO. a_3 removes this attack relation. The valuation of a is positive, the token is given to CON. a_2 attacks a with c, and a_3 removes this attack. a_3 has no relevant moves left, the gameboard is stable, the protocol stops with $v(a) = 0$.

We let the reader check that in this example, every possible dialogue ends with a non-negative valuation for a. We have here garanteed convergence to the merged outcome.

However, this protocol can also lead to more complex examples in which the issue is not predetermined, and can depend on the moves of the agents.

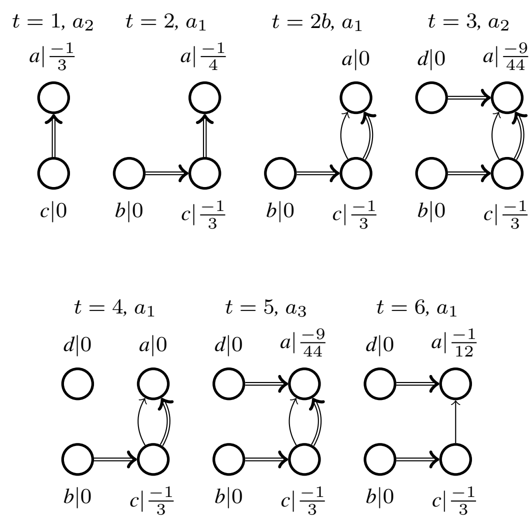
Example 4 *Let three agents with their argumentation systems.*



The issue of the dialogue is the argument a.

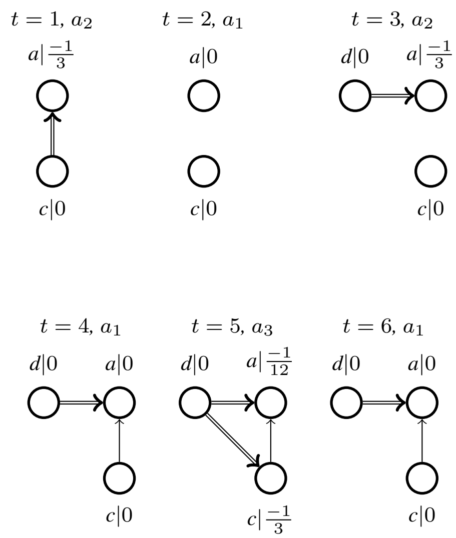
We have $PRO = \{a_1\}$, $CON = \{a_2, a_3\}$. At the begining, we have $RP_1^0 = RP_2^0 = RP_3^0 = \{\,\}$, $SP_1^0 = SP_2^0 = SP_3^0 = \{\,\}$ and $AS^0(GB) = \langle\{a\}, \{\,\}, \{\,\}\rangle$.

This first sequence of moves allows players in CON to win :



At $t = 1$, a_2 plays for CON and attacks the argument a with c. Then, a_1 attacks the argument c with b. This move increases the valuation of a, but this valuation remains negative. So, a_1 plays another move for PRO and adds a support relation between c and a. The valuation of a is 0, the token is given to CON. a_2 attacks a with d, a_1 removes this attack relation but a_3 puts it back at $t = 5$. Finally, a_1 removes the attack between c and a, but this move does not allow the valuation of a to reach 0. a_1 has no relevant move left, and cannot raise the valuation of a more. The gameboard is stable, the protocol stops with $v(a) = \frac{-1}{12}$.

This second sequence of moves allows PRO to win :



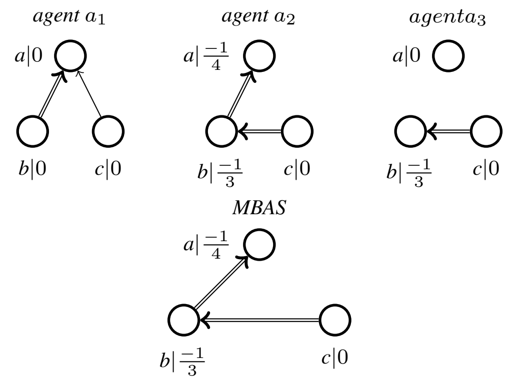
At $t = 1$, a_2 plays for CON and attacks the argument a with c, but a_1 removes this attack. At $t = 3$, a_2 attacks a with d. a_1 adds a support relation between c and a, then a_3 attacks c with d. Finally, a_1 removes the attack between d and c, and puts the valuation of a to 0. Players in CON have no relevant move left, the gameboard is stable and the protocol stops with $v(a) = 0$.

Note that in the previous example, a_1 can play the attack relation (b, c), but this prevents his group from obtaining a positive value for the issue on the gameboard, as it lowers the influence of his support relation and players in CON have no interest to remove it.

4.2 Protocol π_1

In this protocol, agents want to have the valuation of the issue in their own argumentation system as close as possible to the one on the gameboard. We do not have "groups" of agents anymore and agents can play anytime. The central authority picks the move which amends the most the valuation of the issue.

Example 5 *Let three agents with their argumentation systems, and the following merged argumentation framework :*

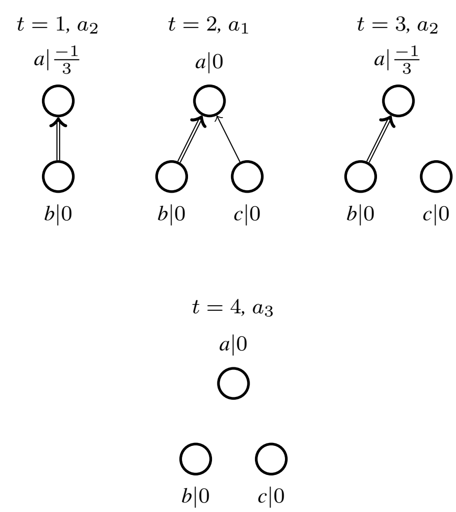


The issue of the dialogue is the argument a.

At the begining, we have $RP_1^0 = RP_2^0 = RP_3^0 = \{\}$, $SP_1^0 = SP_2^0 = SP_3^0 = \{\}$ and $AS^0(GB) = \langle\{a\}, \{\}, \{\}\rangle$.

We can notice that at the beginning, a_1 and a_3 are perfectly happy with the gameboard as it is. The only one who has incentive to move is a_2.

A sequence of moves allowed by this protocol is the following :



At $t = 1$, a_1 and a_3 are perfectly happy with a valuation of a at 0. a_2 proposes to add the attack (b, a) to move the valuation of a closer to his own.
At $t = 2$, each agent proposes a move : a_1 proposes to add the support relation (c, a) to set the valuation of a back to 0, a_3 proposes to remove the attack between b and a for the same purpose, whereas a_2 proposes to attack b with c to obtain his own argumentation framework. As this latter move has a smaller effect on the valuation of a, the central authority chooses randomly a move

from the two former. Assume that he chooses the support relation (c, a) proposed by a_1.
At $t = 3$, a_2 is the only agent to propose a move, and his most relevant one is to remove the support relation between c and a.
Then, at $t = 4$, a_1 cannot propose any move as he already has added the support relation. The two remaining relevant moves are for a_2 to add (c, b) and for a_3 to remove (b, a). As this latter move has the biggest effect on the valuation of a, it is picked by the central authority.
a_2 cannot propose any other move. The gameboard is stable and the protocol stops with $v(a) = 0$.

In this example, all the possible sequences of moves will lead to a valuation of 0 for argument a. Examples in which the final valuation of the issue depends on the sequence of moves are difficult to construct, due to the strict rule of selection of the most "relevant" move, that is the move which amends the most the valuation of the issue. The large number of agents and arguments necessary to do so prevent us from presenting such an example here.

5 Conclusion

This paper is a preliminary study of the design of multiparty protocols for regulating debates among agents equipped with expressive argumentation systems allowing to distinguish supports and attacks and relying on the notion of valuation of arguments (instead of acceptability). Our main objective has been to emphasize the problems and questions that one faces when designing such protocols. When several (more than two) agents are involved, a natural approach is to "group" agents, so as to make sure for instance that the diversity of opinions is well represented in the debate (we may want to avoid a bias of the protocol towards agents that have the technical ability to respond more quickly). We have argued that at least two distinct types of protocols may be introduced, depending on domain-specific settings : category-based and cluster-based protocols. Furthermore, setting a regulation policy is a difficult issue, and once the protocol is fixed its specific properties have to be studied carefully. With the help of several examples, we have illustrated the consequences of different design choices instantiated in two specific protocols.

Références

[1] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems (regular paper). In *International Conference on Scalable Uncertainty Management (SUM), Toulouse, 01/09/2010-02/09/2010*, pages 42–55, http ://www.springerlink.com/, septembre 2010. Springer-Verlag.

[2] L. Amgoud, C. Cayrol, M. C. Lagasquie-Schiex, and P. Livet. On bipolarity in argumentation frameworks. *Int. J. Intell. Syst.*, 23 :1062–1093, 2008.

[3] P. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artif. Intell.*, 128(1-2) :203–235, 2001.

[4] E. Bonzon and N. Maudet. On the outcomes of multiparty persuasion. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2011)*, May 2011. To appear.

[5] M. Caminada and G. Pigozzi. On judgment aggregation in abstract argumentation. *Journal of Autonomous Agents and Multiagent Systems*, 22 :64–102, 2011.

[6] C. Cayrol and M.-C. Lagasquie-Schiex. Gradual valuation for bipolar argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, pages 366–377, 2005.

[7] C. Cayrol and M.-C. Lagasquie-Schiex. Graduality in argumentation. *J. Artif. Intell. Res. (JAIR)*, 23 :245–297, 2005.

[8] C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation frameworks. In *European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'05)*, pages 378–389, 2005.

[9] S. Coste-Marquis, C. Devred, S. Konieczny, M.-C. Lagasquie-Schiex, and P. Marquis. On the Merging of Dung's Argumentation Systems. *Artificial Intelligence*, 171 :740–753, 2007.

[10] F. Dignum and G. Vreeswijk. Towards a testbed for multi-party dialogues. In *Proc. of the Workshop on Agent Communication Languages*, pages 212–230, 2003.

[11] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and

n-persons games. *Artificial Intelligence*, 77 :321–357, 1995.

[12] P. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Inconsistency tolerance in weighted argument systems. In *Proc. of AAMAS'09*, 2009.

[13] S. Parsons, M. Wooldridge, and L. Amgoud. Properties and complexity of some formal inter-agent dialogues. *Journal of Logic and Computation*, 13(3) :347–376, 2003.

[14] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15 :347–376, 2005.

[15] H. Prakken. Formal systems for persuasion dialogue. *Knowledge Engineering Review*, 15 :1009–1040, 2005.

[16] I. Rahwan and K. Larson. Pareto optimality in abstract argumentation. In *Proc. of AAAI'08*, pages 150–155, 2008.

[17] I. Rahwan and K. Larson. *Argumentation and Game Theory*, chapter Argumentation in Artificial Intelligence, pages 321–339. Springer, 2009.

[18] I. Rahwan and F. A. Tohmé. Collective argument evaluation as judgement aggregation. In *Proc. of AAMAS'10*, pages 417–424, 2010.

[19] M. Wardeh, T. Bench-Capon, and F. Coenen. Multi-party argument from experience. In *Proc. of Argumentation in Multiagent Systems (ArgMas'09)*, pages 216–235, 2009.