On the Outcomes of Multiparty Persuasion

Elise Bonzon
LIPADE, Université Paris Descartes
France
elise.bonzon@parisdescartes.fr

Nicolas Maudet
LAMSADE, Université Paris Dauphine
France
maudet@lamsade.dauphine.fr

ABSTRACT

In recent years, several bilateral protocols regulating the exchange of arguments between agents have been proposed. When dealing with persuasion, the objective is to arbitrate among conflicting viewpoints. Often, these debates are not entirely predetermined from the initial situation, which means that agents have a chance to influence the outcome in a way that fits their individual preferences. This paper introduces a simple and intuitive protocol for multiparty argumentation, in which several (more than two) agents are equipped with argumentation systems. We further assume that they focus on a (unique) argument (or issue) —thus making the debate two-sided— but do not coordinate. We study what outcomes can (or will) be reached if agents follow this protocol. We investigate in particular under which conditions the debate is pre-determined or not, and whether the outcome coincides with the result obtained by merging the argumentation systems.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: Multiagent systems

General Terms

Theory

Keywords

Argumentation, persuasion protocols, multiagent systems

1. INTRODUCTION

Protocols for persuasion [11] regulate the exchange of arguments to arbitrate among conflicting viewpoints. Depending on the underlying objective, such protocols can be more or less flexible. When conceived as *argument games* (or *disputes*) between a proponent and an opponent, proof theoretical counterparts of argumentation semantics must leave no room for uncertainty in the result. On the other hand, when the ambition is to regulate some interaction between different agents, it is often desirable that the outcome of the dialogue is not entirely predetermined from the initial situation [7, 8]. This means that agents have a chance to influence the outcome of the game depending on how they play.

Recently, different properties of these protocols have been studied with the help of game-theoretical concepts (see [13] for a survey). This paper follows this line of work, and develops an analysis which builds on very similar assumptions. In particular, we shall take for granted, as [12] do for instance, that agents' argument moves should immediately improve their satisfaction with respect to the current situation of the debate. This work however departs from these previous proposals, in the sense that we address a case of *multiparty argumentation*. In this context, a number ($n > 2$) of agents exchange arguments on a common gameboard. No central computation of the whole system takes place, and no coordination between agents is assumed (even if they share the same view). The motivating applications we have in mind are for example online platforms allowing users to asynchronously modify the content of a collective debate. We want to study what outcomes will be reached with these type of interactions. This situation has received so far little attention and there are good reasons for that (see [4] for a discussion on the challenges raised by multiparty dialogues, and [16] for a recent study of multiparty persuasion in a specific framework). Firstly, it is not obvious to identify what would be the "correct" collective outcome in this case. In this paper we rely on a specific (natural in our case) *merged solution* [3] to assess the quality of the outcome. Secondly, the design of these protocols is made very difficult by the number of parameters to consider (think of several agents focused on possibly different issues), and renders the analysis of their formal properties challenging.

To keep things as simple as possible in this study, the following assumptions are made: (i) all the 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 status of this specific argument); (ii) all the agents make use of the same argumentation semantics to evaluate both their private argumentation system and the situation on the common gameboard (specifically, we rely on Dung's grounded semantics [5]); and (iii) all the agents share the same set of arguments, but they may have different views on the attack relations between these arguments (this may result, *e.g.*, from agents being equipped with value-based argumentation systems [1] and ranking differently the values). While these restrictions are arguably severe, we will see that the resulting framework is already sufficiently rich to illustrate the variety of results that may be derived in the

Cite as: On the Outcomes of Multiparty Persuasion, E. Bonzon and N. Maudet, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, Tumer, Yolum, Sonenberg and Stone (eds.), May, 2–6, 2011, Taipei, Taiwan, pp. XXX-XXX.

study of multiparty argumentation protocols.

The remainder of this paper is as follows. In the next section we provide the necessary background on argumentation semantics. Section 3 sets up the basic elements of our framework. The properties of the proposed protocol are studied in Section 4. Finally, Section 5 discusses related works and concludes, discussing possible extensions of the preliminary study proposed here.

2. BACKGROUND

2.1 Argumentation Systems

In this section, we briefly recall some key elements of abstract argumentation frameworks as proposed by Dung [5]. The exact content of arguments is left unspecified here. Instead, a (finite) set of arguments is given, as well as the different conflicts among them.

Definition 1. *An* **argumentation system (AS)** *is a pair* $\langle A, R\rangle$ *of a set* A *of arguments and a binary relation* R *on* A *called the* **attack relation**. $\forall a, b \in A$, aRb *(or* $(a, b) \in R$*) means that* a **attacks** b *(or* b *is* **attacked** *by* a*). An AS may be represented by a directed graph, called the* **argumentation graph**, *whose nodes are arguments and edges represent the attack relation.*

From this argumentation graph, we can introduce some notions related to graph theory in order to characterize some properties of the argumentation system.

Definition 2. *Let AS be an argumentation system, and* G *be the argumentation graph associated. A* **path** *in* G *is a sequence of nodes such that from each node there is an edge to the next node in the sequence. A finite path has a first and a last node. An edge* (b, c) *is an* **attack edge** *(resp.* **defense edge***) for an argument* a *iff there is an even-length (resp. odd-length) path from* c *to* a.

Note that an edge can be both an attack and a defense edge. In Dung's framework, the *acceptability of an argument* depends on its membership to some sets, called extensions. These extensions characterize collective acceptability.

Definition 3. *Let* $AS = \langle A, R\rangle$ *be an argumentation system. Let* $S \subseteq A$. S *is* **conflict-free** *for AS iff there exists no* a, b *in* S *such that* aRb. S **collectively defends** *an argument* a *iff* $\forall b \in A$ *such that* bRa, $\exists c \in S$ *such that* cRb.

A set of arguments is *admissible* when it is conflict-free and each argument of the set is collectively defended by the set itself. Several *semantics for acceptability* have been defined in [5]. In what follows, we concentrate on the notion of *grounded semantics* which can be defined as follows:

Definition 4. *Let* $AS = \langle A, R\rangle$ *be an argumentation system. Let* $S \subseteq A$. S *is a* **grounded extension** *of AS iff* S *is the least fixed point of the characteristic function of AS (*F*:* $2^A \rightarrow 2^A$ *with* $F(S) = \{a$ *such that* S *collectively defends* $a\}$*).*

Intuitively, a *grounded extension* contains all arguments which are not attacked, as well as the arguments which are defended (directly or not) by non-attacked arguments. There always exists a unique grounded extension. We shall denote by $\mathcal{E}(AS)$ the grounded extension of the system AS.

2.2 Merged Argumentation System

We now consider a set N of n agents. Each agent holds an argumentation system $AS_i = \langle A, R(i)\rangle$, sharing the same arguments A, but with possible conflicting views on attack relations between arguments (coming for instance from different underlying preferences). What should be the collective view in that case? To tackle this problem, we rely on the notion of a *merged* argumentation system [3]. In the specific case we discuss here, it turns out that a meaningful way to merge is to take the *majority argumentation system* where attacks supported by a majority of agents are kept (this corresponds to minimizing the sum of the edit distances between the AS_i and the merged system, see Prop. 41 in [3]). Assuming, on top of that, that ties are broken in favour of the absence of an attack allows to ensure the existence of a single such merged argumentation system, that we denote MAS_N.

Definition 5. *Let* N *be a set of agents and* $\langle AS_1 \dots AS_n\rangle$ *be the collection of their argumentation systems. The* **majority argumentation system** *is* $MAS_N = \langle A, M\rangle$ *where* $M \subseteq A \times A$ *and* xMy *when* $|\{i \in N | (x, y) \in R_i\}| > |\{i \in N | (x, y) \notin R_i\}|\}$.

The corresponding *merged outcome* is denoted by $\mathcal{E}(MAS_N)$.

3. A PROTOCOL FOR FOCUSED AGENTS

We now turn to the following question: supposing 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 answer this problem, we need of course to design a specific protocol and to make some assumption regarding agents' preferences regarding the outcome.

3.1 Agents' Preferences

We assume that agents are *focused* [14], 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 [11]. Unsurprisingly, agents want to see the acceptability status (under the grounded semantics) of the issue coincide in the debate and in their individual system. Thus we can see the debate as opposing two groups of agents: $CON = \{a_i \in N | d \notin \mathcal{E}(AS_i)\}$ and $PRO = \{a_i \in N | d \in \mathcal{E}(AS_i)\}$. If $X = PRO$ (resp. CON), we have $\overline{X} = CON$ (resp. PRO).

3.2 The Gameboard

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 argumentation system* [6] where the weight is simply a number equal in the difference between the number of agents who asserted a given attack and the number of agents who opposed it. We denote by $xR_\alpha y$ the fact that the attack 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 argumentation system $\langle A(GB), M\rangle$ where $M \subseteq A(GB) \times A(GB)$ and $xMy = \{xR_\alpha y | \alpha > 0\}$. In words, we only retain those attacks 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).

3.3 A Relevance-based Protocol

We now introduce our simple protocol which allows agents to exchange their arguments in order to agree on the status of a specific argument d, the issue of the dialogue. Let $AS^t(GB)$, $A^t(GB)$ and $R^t(GB)$ be respectively the argumentation system, the set of arguments and the set of attack relations on the gameboard after round t. The protocol indeed proceeds in rounds which alternate between the two groups of agents (PRO and CON). Within these groups though, no coordination takes place: the agents may for instance play asynchronously and the authority simply picks the first permitted and relevant move before returning the token to the other side. *Permitted* moves are simply positive assertions of attacks xRy (with $y \in A^t(GB)$), or contradiction of (already introduced) attacks (with $(x, y) \in R^t(GB)$). Note that arguments are progressively added on the gameboard via these attacks, and that it may not contain the whole set of arguments when the debate concludes. A move is *relevant* [10] at round t for a PRO agent (resp. CON agent) if it puts the issue back in (resp. drops the issue from) $\mathcal{E}(AS^t(GB))$. 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 $RP_i^t \subseteq \{(x, y)|x, y \in A\}$ which contains the attack relations or the *non*-attack relation he has added on the gameboard at time t, in order to prevent him from adding twice the same relation. The proposed protocol is as follows:

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

(2) The first round starts with the issue on the gameboard and the turn given to CON.

(3) Until a group of agents cannot move, we have:

    (a) agents independently propose moves to the central authority;

    (b) the central authority picks the first (or at random) relevant move from the group of agents whose turn is active, update the gameboard, and passes the turn to the other group

When a (relevant) move is played on the gameboard, the following update operation takes place:

(1) after an assertion xRy

- if $xR_\alpha y \in R^t(GB)$ then $\alpha := \alpha + 1$
- if $xR_\alpha y \notin R^t(GB)$ 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, we have $\alpha := \alpha - 1$

Note the asymetry here: introducing a new argument can only be done via a positive assertion, since it can never be relevant to contradict an attack refering to an argument that was not introduced already. The reader may remark that the value of α is binary if agents obey this protocol; however we discuss in Section 4.3 an extension where this is not necessarily the case.

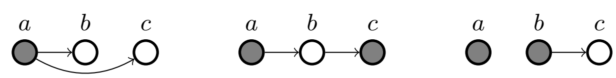
When (after a sequence σ of moves) a group of agents cannot move, we say that the gameboard is stable and we refer to $\mathcal{E}(AS(GB^\sigma_{t\to\infty}))$ (or simply $\mathcal{E}(AS(GB))$ when clear from the context) as the *outcome of the debate.*

3.4 Properties

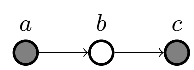
The outcome $\mathcal{E}(AS(GB^\sigma_{t\to\infty}))$ resulting from a specific sequence of moves σ obeying this protocol will typically be compared with the result which would be obtained by merging the argumentation systems ($\mathcal{E}(MAS)$). We may want to ensure different properties, but we typically 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, that is $\forall\sigma\ d \in \mathcal{E}(AS(GB^\sigma_{t\to\infty})) \leftrightarrow d \in \mathcal{E}(MAS_N)$
- *Reachability of the merged outcome*— requires *at least* one possible sequence of moves to reach the merged outcome, that is $\exists\sigma\ d \in \mathcal{E}(AS(GB^\sigma_{t\to\infty})) \leftrightarrow d \in \mathcal{E}(MAS_N)$

EXAMPLE 1. *Let three agents with their argumentation systems, and the following merged argumentation framework:*



a_1 $\mathcal{E}(AS_1) = \{a\}$ a_2 $\mathcal{E}(AS_2) = \{a, c\}$ a_3 $\mathcal{E}(AS_3) = \{a, b\}$



$\mathcal{E}(MAS) = \{a, c\}$

The issue of the dialogue is the argument c. We have $CON = \{a_1, a_3\}$, $PRO = \{a_2\}$. At the begining, we have $RP_1^0 = RP_2^0 = RP_3^0 = \{\}$, $AS^0(GB) = \langle\{c\}, \{\}\rangle$ and $\mathcal{E}(AS^0(GB)) = \{c\}$.

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

$t = 1$ - a_1 *plays for CON:* $RP_1^1 = \{(a, c)\}$

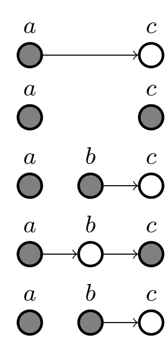
$t = 2$ - a_2: $RP_2^2 = \{(a, c)\}$

$t = 3$ - a_3 *plays for CON:* $RP_3^3 = \{(b, c)\}$

$t = 4$ - a_2: $RP_2^4 = \{(a, b), (a, c)\}$

$t = 5$ - a_3: $RP_3^5 = \{(b, c), (a, b)\}$

$t = 6$ - a_2 *cannot add c in the extension*

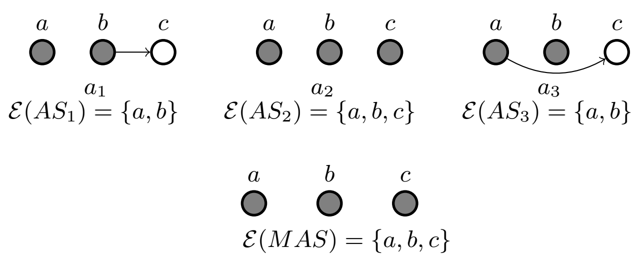


The game board is stable, we obtain $\mathcal{E}(AS(GB)) = \{a, b\}$.

The first interesting thing to observe on this simple example is the fact than the status of an issue in the merged argumentation system can contradict the opinion of the majority. This is discussed in [3]: if agents vote on extensions, the attack relations from which extensions are characterized are not taken into consideration, and a lot of significant information is not exploited.

Another important thing to note in this example is that PRO agents cannot ensure c in $\mathcal{E}(AS(GB))$. It is then impossible to guarantee convergence to the status of the issue obtained in the merged argumentation system. This is due to the fact that agent a_1 has no interest to play the attack relation (a, b), which appears in the MAS. As studied in a different context by [12], this can be seen as a strategic manipulation by witholding an argument or an attack between arguments. But is it even possible to reach the merged outcome in this case? We leave it to the reader to check that this is not the case here. One may then think that the group with the highest number of agents will always win with our protocol. It is not the case, as shown by the fairly simple following example.

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



The issue of the dialogue is the argument c. We have $CON = \{a_1, a_3\}$, $PRO = \{a_2\}$. Agents in CON can attack c in two ways: either a_1 can play bRc; or a_3 can play aRc. But a_2 will be able to remove either attack, and CON agents will not have the possibility to counter-attack. We will obtain $\mathcal{E}(AS(GB)) = \{a, b, c\}$.

The two previous examples show that the characterization of the result obtained by debates following this protocol is not as simple as one can believe at first glance. We now introduce some useful and more sophisticated notions.

3.5 Global arguments-control graph

In order to characterize the status of the issue obtained by our protocol we will need the notion of *global arguments-control graph* (ACG). The idea here is to gather the attacks of all agents in the same argumentation graph, and then determine which group, PRO or CON, have the control over some path of this graph, and thus a possible way to reach its preferred outcome. To do so, we first need to define the notion of *control* over an attack relation:

Definition 6. *Let N be a set of agents, $\langle AS_1 \dots AS_n\rangle$ be the collection of their argumentation systems, and $L = \cup_{i \in 1\dots n} R(i)$ be the union of all attack relations. Let $X \in \{CON, PRO\}$. Finally, let $add_{(a,b)} = \{a_i \in N | (a, b) \subseteq R(i)\}$, and $rem_{(a,b)} = \{a_i \in N | (a, b) \not\subseteq R(i)\}$.*

- *X has the* **constructive control** *of $(a, b) \in L$, denoted by $X^+(a, b)$, iff $|add_{(a,b)} \cap X| > |rem_{(a,b)} \cap \overline{X}|$, that is if the number of agents in X who can add (a, b) is greater than the number of agents in $\overline{X}$ who can remove it.*
- *X has the* **destructive control** *of $(a, b) \in L$, denoted by $X^-(a, b)$, iff $|rem_{(a,b)} \cap X| \geq |add_{(a,b)} \cap \overline{X}|$, that is if the number of agents in X who can remove (a, b) is greater or equal than the number of agents in $\overline{X}$ who can add it.*

The following remarks are simple but useful: (1) It is impossible to have both $X^+(a, b)$ and $\overline{X}^-(a, b)$; (2) It is possible to have both $X^+(a, b)$ and $X^-(a, b)$; (3) A minority group cannot have constructive and destructive control of an edge: if $|X| < |\overline{X}|$, it is impossible to have both $X^+(a, b)$ and $X^-(a, b)$; (4) If there is not $X^-(a, b)$ (resp. $X^+(a, b)$), then there is $\overline{X}^+(a, b)$ (resp. $\overline{X}^-(a, b)$).

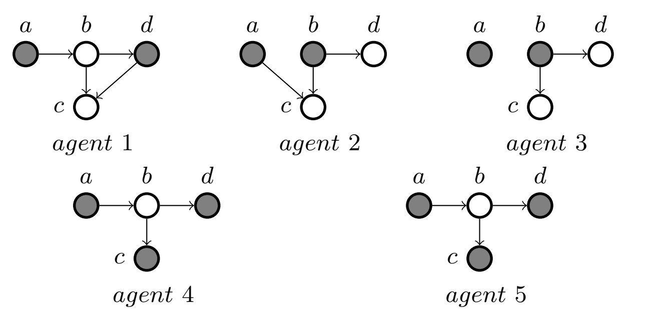
Observe that the notion of destructive control intuitively says that a group has the control to overweight any possible attempt to establish a given relation. This of course vacuously holds when no agent from the other group supports the relation at all, in which case the relation is not even playable.

Definition 7. *Let N be a set of agents and $\langle AS_1 \dots AS_n\rangle$ be the collection of their argumentation systems. We will say that $(a, b) \in \cup_{1\dots n} R(i)$ is* **playable** *by a $X \in \{PRO, CON\}$, denoted by $X_\bullet(a, b)$, iff there is an $a_i \in X$ such that $(a, b) \in R(i)$.*

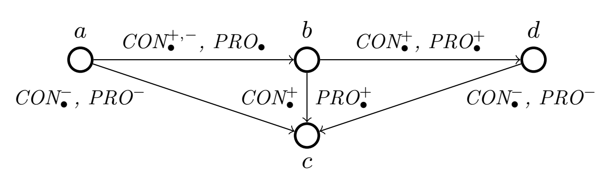
For the sake of readability, we will only specify the information about playability when it is relevant.

Definition 8. *Let N be a set of agents and $\langle AS_1 \dots AS_n\rangle$ be the collection of their argumentation systems. The* **global arguments-control graph** *is $ACG_N = \langle A, L\rangle$ is constructed as follow: (1) $L = \cup_{i \in 1\dots n} R(i)$ (2) Label each $(a, b) \in L$ by the information about control and playability for each group $X \in \{PRO, CON\}$.*

Example 3. *Five agents have the following argumentation systems:*



The issue of the dialogue is the argument c. We have $CON = \{a_1, a_2, a_3\}$, $PRO = \{a_4, a_5\}$. The global arguments-control graph is the following:



4. PROPERTIES

We now discuss three distinct properties that we wish to analyze on the basis of the ACG: (i) who wins the debate?, (ii) does the outcome of the debate coincide with that of the merged system?, and (iii) is it useful to allow moves that reinforce previous moves?

4.1 Who wins the debate?

The first question that we address is whether an omniscient observer would know *a priori* which group of the debate could possibly or necessarily win the debate, in particular whether some debates are "open" (*i.e.* not pre-determined [8]).

DEFINITION 9. *We will say that the issue of the debate is a* **possible outcome** *for a group X if this group has a possibility to set the acceptability status of this argument to coincide in the debate and in their individual system. The issue is a* **necessary outcome** *for X iff this issue is not a possible outcome for $\overline{X}$.*

DEFINITION 10. *A* **path for d controlled by** *CON is an odd-length path from x to d such that (i) CON has constructive control on all the attack edges for d, and (ii) CON has destructive control on all the defense edges for d attacking x.*

Note that condition (ii) covers in particular the case where the first node x is not attacked. Controlling a path is not enough since alternative defenses may exist. By extension, we then define the notion of a tree controlled by CON.

DEFINITION 11. *A* **tree for d controlled by** *CON is a tree such that (i) d is the root, (ii) all the paths from the leaves to d are controlled by CON, and (iii) for any attack edge yRx of the tree, it contains all the defense edges zRy such that $PRO^+(z, y)$.*

This gives us a condition guaranteeing that a favourable outcome can be attained by CON.

PROPOSITION 4.1. *If there exists a tree for d controlled by CON, then the issue d is a possible outcome for CON.*

PROOF. (Sketch.) Observe that because we have a tree no edge can be played both as an attack and a defense edge. Then CON can certainly win by making sure that all the attack moves of the tree are placed, since it can respond to any possible defense edge of the tree on which PRO has constructive control, and it can certainly remove any other defense edge which could be played by PRO (because it must hold the destructive control on these edges). ■

A couple of remarks are in order here. If the ACG itself happens to be a tree, then the above condition is necessary and sufficient to guarantee that the outcome is necessary for CON. However, in general, things turn out to be much more involved. First this condition is not necessary for the outcome to be possible for CON: this group of agents may win in absence of such a tree (in fact, even in absence of a single path controlled by himself). This may look counter-intuitive, but the reason lies on the fact that the control of an edge may be gained during the debate in specific circumstances. We do not elaborate on this point here but a related observation is developed in Section 4.3. Secondly, this condition is not strong enough to guarantee that the issue is a necessary outcome for CON. Indeed, in absence of coordination, agents of CON may not play the moves of the tree only. And there are cases where this may make d a possible outcome for PRO. To see why this may be the case, recall that an edge may be both an attack and a defense edge for the same issue d, as it may appear on several distinct paths. When that happens, this edge may be used as an attack edge, preventing the deployment of the path controlled by CON. The following notions of *switch* captures this.

DEFINITION 12. *An edge (x, y) on a path P is a* **switch** *for d if (i) it is a defense for d on P, (ii) it is playable by CON, (iii) there exists an even-length path from y to d such that all the attack edges are playable by CON and all the defense edges are playable by PRO. So it is also a potential attack for d via a different path.*

Essentially, what this definition says is that there is a possibility that this edge (x, y) may be played as an attack by CON. As mentioned before this may harm CON own line of attack. Following this, we say that there exists a *switch for path P for d controlled by CON* if there exists a defense edge for d attacking x (the first node of P) that is a switch. Each path in which a switch for P is an attack edge is called a *switch path of P*.

We are now in a position to informally state some conditions under which d may not be a necessary outcome for CON despite the existence of a tree controlled by himself. In fact it is the case when there exists a set of switches $\mathcal{S}$ such that: (i) for any tree t for d controlled by CON, there exists a switch belonging to $\mathcal{S}$ for a path (of t) for d controlled by CON; (ii) there must exist a sequence of moves such that (1) all the switches in $\mathcal{S}$ are actually played, and (2) PRO has the destructive control over an attack edge of each resulting switch path; (iii) there must exist a sequence of moves such that all the switches in $\mathcal{S}$ are maintained.

It may not be immediately clear to the reader why the mere existence of switches —Cond (i)— does not imply the fact that they can be played —Cond. (ii.1): after all, the definition requires a path of playable moves reaching the switch to exist. The subtlety lies on the fact that these paths may interact when they share some arguments. In this case, the existence of a path may preclude other paths to be played. Intuitively, Cond (iii) caters for the fact that a switch may be "patched" by CON if he manages to append an odd-length path right behind the switch.

The next question is whether these conditions can be simply expressed on the basis of the ACG. For (i) and (ii.2) this is obvious. For (ii.1) and (iii) this is more challenging because the definition refers to possible sequences of moves. We will rely instead on sufficient conditions:

PROPOSITION 4.2. *The issue d may not be a necessary outcome for CON if there exists a set of switches $\mathcal{S}$ such that: (i) for any tree t for d controlled by CON there exists a switch belonging to $\mathcal{S}$ for a path (of t) for d controlled by CON; (ii) there exists a set of switch paths $\mathcal{P}$ for $\mathcal{S}$ such that these paths do not share any arguments (except d), and PRO has destructive control over an attack edge of each of these paths; (iii) for all switches $(x, y) \in \mathcal{S}$, there does not exist any even-length path P reaching d meeting x, such that $[x, d]$ constitutes an odd-path length, and such that CON has* *the constructive control on all attack edges, and PRO has the constructive control on all the defense edges.*

PROOF. (Sketch.) Cond (i) ensures that at least an attack path of each tree controlled by CON can be potentially switched. Cond (ii) suffices to guarantee that the switch paths are independent, so all the switches can actually be played, and the switch paths subsequently cut. As for (iii), observe that there are two ways to render a switch (xRy) ineffective. Either by simply removing it, but this necessarily requires a path (leading to the issue) to meet the node y of the switch (so that playing xRy would be relevant). This is impossible. Or by appending an odd-length path (a "patch") on x of the switch, such that a move meeting the node y of the switch could now be played. This can only happen when there exists an even-length path reaching d meeting x, such that $[x, d]$ constitutes an odd-path length, and such that CON has the constructive control on all attack edges, and PRO has the constructive control on all the defense edges; otherwise there is necessarily a possibility that this path does not reach x. ■

Note that for (ii) it may be the case that switch paths, even interacting, allow the switch to be played. For (iii) it may be the case, on the other hand, that even if such paths leading to patches do exist, they could not be played because they are interacting in a way that makes them mutually exclusive. What this discussion suggests is that obtaining a full characterization of outcomes is certainly very challenging in the general case. It provides however a simple way to construct examples of debates that are indeed open:

EXAMPLE 3 CONT. *We easily see that the issue c is a* **possible outcome** *for the agents in CON: CON can attack c with b. Then, the only possible move for PRO is to defend d with aRb. However, CON can remove this attack, and PRO has no other move.*

But c is also a **possible outcome** *for PRO: CON can start with dRc, which is playable by a_1. Then, a_5 will defend with bRd, and a_1 counter-attack with aRb. If the next move of PRO is to remove dRc, then CON has no other move left: it cannot add the attack bRc, as it is defended by a; and it cannot remove the edge (a, b) as it does not drop c from the extension. In this case, (a, b) is a switch and the merged outcome is then (only) reachable.*

4.2 Does it coincide with the MAS?

The next step here is to characterize the convergence and/or the reachability of the merged outcome. We have already seen that the merged outcome is not always reachable, but is it possible to find some case for which it is? To answer this question, we first need the following lemma.

LEMMA 1. *Let N be a set of agents, $ACG_N = \langle A, L\rangle$ be the global arguments-control graph and $MAS_N = \langle A, M\rangle$ be the merged argumentation system. If there is no edge $(a, b) \in L$ such that a group $X \in \{CON, PRO\}$ has the constructive and destructive control of (a, b), then all the edges controlled constructively in the ACG belong to the MAS, whereas all the edges controlled destructively in the ACG do not belong to the MAS.*

PROOF. By remark (4), we know that either $X^+(a, b)$ and $\overline{X}^+(a, b)$, or $X^-(a, b)$ and $\overline{X}^-(a, b)$. Take the case of constructive control: we have $|add_{(a,b)} \cap X| > |rem_{(a,b)} \cap \overline{X}|$ and $|rem_{(a,b)} \cap X| < |add_{(a,b)} \cap \overline{X}|$. As $X \cap \overline{X} = \{\}$, we have $|add_{(a,b)}| > |rem_{(a,b)}|$. Then, by definition of the merged argumentation system, we know that $(a, b) \in M$ (that is (a, b) is an edge of the MAS). The case of destructive control is similar. ■

This lemma leads to the following proposition.

PROPOSITION 4.3. *Let N be a set of agents, and $ACG_N = \langle A, L\rangle$. If there is no edge $(a, b) \in L$ such that $X^{+,-}(a, b)$, then the merged outcome is reachable.*

PROOF. We know from Lemma 1 that all the edges controlled constructively in the ACG belong to the MAS, whereas all the edges controlled destructively in the ACG do not belong to the MAS. Let d be the issue of the debate.

(1) Let us assume that $d \in \mathcal{E}(MAS)$. Thus, for all $x \in A$ such that xMd, there is a even-length path $P = (x_1, x_2, \ldots, x, d)$ which defends d. As all these edges belong to the MAS, we know from Lemma 1 that they belong to the ACG, and that they are controlled constructively by PRO and by CON. Thus, CON can play all the attack edges of P, whereas PRO can defend d by adding all the defense edges of P. As x_1 is not attacked in the MAS, there are two possibilities in the ACG:

  - Either x_1 is not attacked in the ACG. In this case, CON can not attack x_1, and then has no possibility to drop d from $\mathcal{E}(AS(GB))$.
  - Or there is an attack edge (y, x_1) in the ACG. As this edge is not in the MAS, we know that $PRO^-(y, x_1)$. So, PRO can remove this edge and then ensure that $d \in \mathcal{E}(AS(GB))$

  As this reasonment holds for all defense path in the MAS, and is playable with our protocol, d is reachable.

(2) Let us assume now that $d \notin \mathcal{E}(MAS)$. So, there is an odd-length path $P = (x_1, x_2, \ldots, x, d)$ in the MAS which attacks d. As all these edges belong to the MAS, we know from Lemma 1 that they belong to the ACG, and that they are controlled constructively by PRO and by CON. Thus, CON can play all the attack edges of P, whereas PRO can defend d by adding all the defense edges of P. As x_1 is not attacked in the MAS, there are two possibilities in the ACG:

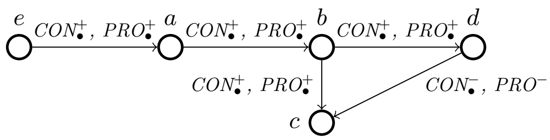
  - Either x_1 is not attacked in the ACG. In this case, PRO can not attack x_1, and then has no possibility to put d in $\mathcal{E}(AS(GB))$.
  - Or there is an attack edge (y, x_1) in the ACG. As this edge is not in the MAS, we know that $CON^-(y, x_1)$. So, CON can remove this edge and then ensure that $d \notin \mathcal{E}(AS(GB))$

  As this path is playable with our protocol, we know that d is reachable.

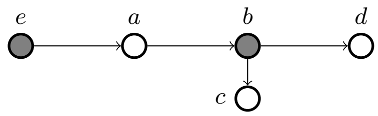
■

Note that we can only ensure the reachability. The following example shows that we do not have guaranteed convergence.

Example 4. *Consider the following global arguments-control graph and merged argumentation systems, where c is the issue.*



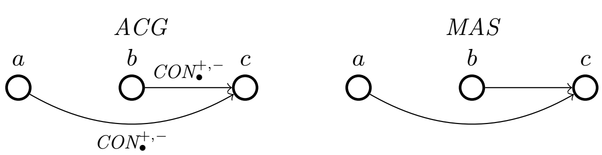
From Lemma 1, we know that the graph of the merged argumentation system is the following:



Thus, $c \notin \mathcal{E}(MAS)$. However, if we suppose that the edge (d, c) is playable for CON, c is a possible outcome for PRO: CON can start by adding dRc. Then, PRO will defend with bRd, and CON counter-attack with aRb. If the next move of PRO is to remove the attack dRc, then CON has no other move left: it cannot add the attack bRc, as it is defended by a; and it cannot remove the edge aRb as it does not drop c from the extension. But c is also a possible outcome for CON: the merged outcome is (only) reachable.

Another important remark is that the converse of Proposition 4.3 is false: as shown by the following example, it is possible for a group to have constructive and destructive control of an edge of the global arguments-control graph, and to ensure the reachibility of the merged outcome.

Example 5. *Consider the following global arguments-control graph, where c is the issue.*



In this graph, CON has the constructive and destructive control over two edges, and the merged outcome is reachable: the outcome is necessary for CON, and $c \notin \mathcal{E}(MAS)$.

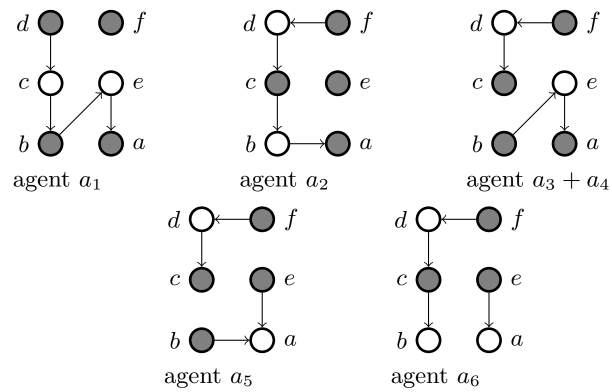
4.3 Is it useful to allow reinforcement?

A natural extension is to consider that a move may also be relevant as long as it *reinforce* (or symmetrically *weakens*) an edge which, if deleted and all other things being equal, would change the status of the issue. Essentially, besides the relevant moves as defined in the previous section, this would allow agents to augment the weight of an existing attack, and we refer to this as a *reinforcement move*. Symetrically, agents may weaken an attack even if its does not directly delete it, and we refer to this as a *weakening move*. This extended protocol would allow any number of such relevant moves during a group's turn, but (as before) would only switch to the other side after a change of the current status of the issue. However, the following proposition tells us that it is not beneficial for an agent to play reinforcement moves. Worse, and rather counter-intuitively, it can actually be damaging for agents to do so.

Proposition 4.4. *Let D_1 be sequences where no agent plays reinforcement or weakening moves, and D_2 be sequences such that X only may play reinforcement moves (but $\overline{X}$ may play weakening moves).*

If $X = PRO$ (resp. CON), then (i) for any $\sigma_2 \in D_2$ with $d \in \mathcal{E}(AS(GB^{\sigma_2}_{t\to\infty}))$ (resp. $d \notin \mathcal{E}(AS(GB^{\sigma_2}_{t\to\infty}))$), there exists $\sigma_1 \in D_1$ such that $AS(GB^{\sigma_1}_{t\to\infty}) = AS(GB^{\sigma_2}_{t\to\infty})$. Further (ii) there exists $\sigma_2 \in D_2$ with $d \notin \mathcal{E}(AS(GB^{\sigma_2}_{t\to\infty}))$ (resp. $d \in \mathcal{E}(AS(GB^{\sigma_2}_{t\to\infty}))$), such that for any $\sigma_1 \in D_1$ it is not the case that $AS(GB^{\sigma_1}_{t\to\infty}) = AS(GB^{\sigma_2}_{t\to\infty})$.

Proof. (Sketch.) We show (ii) by constructing an example where an agent *loses* some destructive control by using reinforcement. It involves 6 agents.



The issue of the debate is a. There are four agents PRO and two agents CON. The key to the analysis is to see that PRO agents initially hold constructive and destructive control on (c, b). Now compare the following sequences of moves. In the first one, a_5 plays bRa. Then a_1 plays cRb and a_2 reinforces this move. At this point, PRO loses its destructive control on (c, b). Assume a CON agent plays dRc. a_1 can remove bRa. Then a_5 can play eRa, a_3 can defend with bRe. Now, with a CON agent playing fRd, the debate is doomed with the issue *out*. In the alternative case where no reinforcement is played, we are in the case discussed in the first protocol: PRO can remove the attack cRb, and win the debate. ■

This result tells us that in the absence of coordination, agents are better off employing moves that are directly relevant, hence adopting a "wait and see" approach. Still, using reinforcement moves may prove useful in practice, in contexts where the debate is limited: for instance, agents may be impressed by seemingly large majorities and avoid these issues to concentrate on some other ones.

5. RELATED WORKS AND CONCLUSION

As already mentioned, our work is close in spirit to the work of Rahwan and Larsson [12]. An important difference with our approach though is that agents control the arguments they can advance in the debate, but that no disagreement takes place regarding the attack relations between these arguments. Another recent proposal of great interest is that of Caminada and Pigozzi [2]. The authors propose different procedures to aggregate different labellings for a given argumentation system into a collective one. The property they want to ensure is that the obtained collective outcome is in some sense compatible with the individual ones. A related contribution of Rahwan and Tohmé [14], which investigates the same question and derives general conclusions on the possibility (or impossibility) to perform such an aggregation, under classical assumptions. As mentioned already, these approaches assume that agents agree on the underlying argumentation system, even though they may have different views on the preferred labelling. Finally, a multiparty protocol for agents equipped with defeasible logic reasoning abilities is investigated in [9]. Each agent initially puts forward an initial claim and the protocol lets iteratively each agent defend his claim or attack the claim of opposing agents (by relying on a sophisticated technique to identify the most effective counter-arguments).

In our proposal, a multiagent protocol regulates the exchange of arguments among focused agents on the basis of the relevance of the moves as proposed by [10]. Although all the agents share the same set of arguments, they may have different views on the attack relations among these arguments. In case of discrepancy on a relation we have opted for a majoritarian approach: the side supported by the highest number of agents wins (more sophisticated approaches are discussed in [15]). Furthermore, even though agents exchange arguments on a common gameboard, it is important to note that no central authority gets to know the whole argumentation system of each agent. We have investigated some formal properties of this protocol. In particular, we have shown that there are cases where the outcome is not entirely pre-determined from the initial situation, and given non-trivial sufficient conditions to identify such debates (based on different notions of control of attacks by a group of agents). We have also given conditions under which the merged outcome can be reached, and discussed a natural extension of the protocol where moves can be reinforced (but showed that agents can only be worse off by using these extended set of moves).

A natural follow-up of this work would be to provide some insights regarding how often the debate is indeed open or how often coincidence with merged outcome is observed. Experiments could prove instructive in this respect. As for possible extensions of this work, it is clear that any relaxations of these assumptions brings about some complexity. If agents do not focus on a single issue, among other things, we may not simply distinguish two groups PRO and CON, and it becomes necessary to specify complex preferences over combinations of issues. If we relax the assumption of the set of arguments being shared, we then need to deal (see [3]) with the complex problem of how agents would react in the presence of arguments they were not aware of before. In the perspective of modeling practical debate platforms as mentioned in the introduction, all these aspects will require of course careful study.

Acknowledgements.

We thank Dionysios Kontarinis, Sébastien Konieczny, and the anonymous reviewers, for their numerous and detailed comments.

6. REFERENCES

[1] T. Bench-Capon. Value-based argumentation frameworks. In *Proc. of the Int. Workshop on Non-Monotonic Reasoning (NMR'02)*, pages 443–454, 2002.

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

[3] 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.

[4] 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.

[5] 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.

[6] P. Dunne, A. Hunter, P. McBurney, S. Parsons, and M. Wooldridge. Inconsistency tolerance in weighted argument systems. In *Proc. of the 8th Int. Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'09)*, 2009.

[7] R. Loui. Process and policy: Resource-bounded nondemonstrative reasoning. *Computational Intelligence*, 14(1):1–38, 2002.

[8] 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.

[9] D. H. Pham, G. Governatori, and S. Thakur. Extended defeasible reasoning for common goals in n-person argumentation games. *Journal of Universal Computer Science*, 15(13):2653–2675, 2009.

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

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

[12] I. Rahwan and K. Larson. Pareto optimality in abstract argumentation. In *Proc. of the 23rd Conference on Artificial Intelligence (AAAI'08)*, pages 150–155, 2008.

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

[14] I. Rahwan and F. A. Tohmé. Collective argument evaluation as judgement aggregation. In *Proc. of the 10th Int. Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'10)*, pages 417–424, 2010.

[15] F. A. Tohmé, G. A. Bodanza, and G. R. Simari. Aggregation of attack relations: A social-choice theoretical analysis of defeasibility criteria. In *Proc. of FoIKS*, volume 4932 of *LNCS*, pages 8–23. Springer, 2008.

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