CORE
Metadata, citation and similar papers at core.ac.uk
Provided by HAL-EMSE

# A dialectics system in which argumentative agents play and arbitrate to reach an agreement

Maxime Morge
ENS Mines
158, cours Fauriel
F-42023 Saint Etienne Cedex 2, France
morge@emse.fr

## ABSTRACT

We propose in this paper a formal framework in which agents arbitrate and play to reach an agreement. The argumentation-based reasoning manages the conflicts between arguments having different strengths for different agents. The argumentative agents justify the hypothesis to which they commit and take into account the commitments of their interlocutors. A third agent is responsible of the final decision outcome which is taken by resolving the conflict between two players according to their competence.

## 1. INTRODUCTION

Most real-world decisions must compose divergent actor interests and perspectives. Because the success of a decision depends on the extent to which people believe it has been reached fairly, they must have a role in forming even if they disagree with the final outcome. This observation changes the way we appreciate the democratic procedures. A decision must be collective and argued [12]. We can distinguish two different modalities of formation of political will : the representative democracy and the dialogical democracy. The representative democracy is a process of political-will in which the individual preferences are aggregated to obtain outcome. The electors an the lay public delegate their power to the elect and the experts. The dialogical democracy is a participative fair effect process to compose the interests and perspectives [7]. The civil society debate and deliberate. We claim that such a process can be formalized by a Multi-Agent System (MAS) where autonomous and social agents argue.

Most of the existing formal framework for inter-agent interaction are based on speech acts theory [15]. For example, FIPA-ACL define communicative acts by pre/post conditions bearing on the mental attitudes of agents. This mentalistic approach is not suitable for our objective. (1) The mental concepts are not adapted to manage the conflicts. (2) The communication has no public semantics to be judged in an objective perspective. (3) However isolated communicative acts do not suffice to achieve a common goal, the existing protocols are too rigid for the debate [9].

By contrast, recent works [4, 13, 14] are inspired by the dialectics [6, 10, 16]. Not being a first attempt, this paper focus on the technical exploration of these argumentative techniques for the formalization of a dialogical democracy. We present in this paper a dialectics system. This is a formal framework in which agents communicate to reach a collective decision. (1) The argumentation-based reasoning mechanism manages the interaction between arguments for and against some conclusions. (2) Since the communication language has a social semantics, every agents confer the same meaning to the messages and any third agent is able to draw similar inferences from the conflicts. (3) The dialogue is a flexible and refined process to reach an agreement.

**Paper overview.** Section 2 presents the reasoning mechanism of agents to manage the conflicts between arguments having different strengths for different agents. In accordance with this background, we describe in section 3 a system of argumentative agents where each agent justifies her position and takes into account the positions of her interlocutors. In section 4, we define the formal framework in which the agents collaborate to reach an agreement. The section 5 illustrates this framework with a persuasion protocol.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
*Argumentation in AI and Law* '05 Bolognia, Italy
Copyright 2005 ACM X-XXXXX-XX-X/XX/XX ...$5.00.

## 2. ARGUMENTATION FRAMEWORK

When a set of autonomous agents argue, they share knowledge on which they have their own priorities. That is the reason why the argumentation framework which is proposed in this section is equivalent to a value-based argumentation framework [3, 4]. Moreover, this framework is based upon a logic language [11, 1].

A multi-agents system is a set of social and autonomous agents (written $\mho_A$). Each of them has an identifier $\mathrm{ag}_i \in \mho_A$. They share knowledge, *i.e.* a set of sentences in a common **knowledge language**, denoted $\mathcal{L}_\mho$. Moreover, the agents use the same classical inference, denoted $\vdash_\mho$. As in [11], the formulae of the background logic are defined as rules as $r : L_0 \leftarrow L_1, ..., L_n$ where $L_0 \dots L_n$ are positive or explicit negative ground literals.

The agents share an argumentative theory, i.e. a set of rules promoting values:

DEFINITION 1. *Let $\mho_A = \{ag_1, \dots, ag_n\}$ be a set of agents. The **value-based argumentative theory** $AT_{\mho_A} = \langle \mathcal{T}, V, promote \rangle$ is defined by a triple where:*

- *$\mathcal{T}$ is a theory, i.e. a finite set of rules in $\mathcal{L}_\mho$;*

- *$V$ is a non-empty finite set of values $\{v_1, \ldots, v_t\}$;*
- *$promote : \mathcal{T} \rightarrow V$ maps from the rules to the values.*

*We say that the rule $r$ relates to the value $v$ if $r$ promotes $v$. For every $r \in \mathcal{T}$, $promote(r) \in V$.*

The values have different priorities for different agents [6]. Therefore, agents are individuated by their priorities between values:

DEFINITION 2. *Let $ag_i \in \mho_A$ be an agent. The **value-based argumentative theory of the agent $ag_i$** is a 4-tuple $AT_i = \langle \mathcal{T}, V, promote, \ll_i \rangle$ where :*

- *$AT_{\mho_A} = \langle \mathcal{T}, V, promote \rangle$ is a value-based argumentative theory as previously defined;*
- *$\ll_i$ is the priority relation of the agent $ag_i$, i.e. a strict complete ordering relation on $V$.*

The priority relation of the agent $\mathrm{ag}_i$ is a transitive, irreflexive, asymmetric, and complete relation on $V$. This static relation makes it possible for agents to stratify the theory into finite non-overlapping sets as in [1]. According to the agent $\mathrm{ag}_i$, the priority level of a non-empty theory $T \subseteq \mathcal{T}$ (written $\mathrm{level}_i(T)$) is the least important value promoted by one element in $T$.

The priority relation captures the value hierarchy of an agent on one hand and the theory gathers the knowledge on the other hand. The arguments are built on this knowledge. An argument is composed of a conclusion and a premise, *i.e.* a set of rules from which the conclusion can be inferred:

DEFINITION 3. *Let $T$ be a theory in $\mathcal{L}_\mho$.*
*An **argument** is couple $A = \langle P, c \rangle$ where $c$ is a rule and $P \subseteq T$ is a non-empty set of rules such as : i) $P$ is consistent and minimal for set inclusion ii) $P \vdash_\mho c$. $P$ is the premise of $A$, written $P = premise(A)$. $c$ is the conclusion of $A$, denoted $c = conclusion(A)$.*

$A'$ is a sub-argument of $A$ if the premise of $A'$ is included in the premise of $A$. $A'$ is a trivial argument if the premise of $A'$ is a singleton. Since the theory $T$ can be inconsistent, the set of arguments (denoted $\mathcal{A}(T)$) will conflict.

The relation of attack between arguments in a theory is based on the possible conflicts between a literal and its explicit negation.

DEFINITION 4. *Let $T$ be a theory in $\mathcal{L}_\mho$ and $A = \langle P, c \rangle, B = \langle P', c' \rangle \in \mathcal{A}(T)$ two arguments. $A$ **attacks** $B$ (written $attacks(A, B)$) iff :*
*$\exists P_1 \subseteq P, P_2 \subseteq P'$ such as $P_1 \vdash_\mho L$ and $P_2 \vdash_\mho \neg L$.*
*Similarly, we say that a set $S$ of arguments attacks $B$ if $B$ is attacked by an argument in $S$.*

Because each agent is associated to a particular priority relation, the attack of an argument over another argument can be ignored. The strength of an argument ($\mathrm{strength}_i(A) = \mathrm{level}_i(\mathrm{premise}(A))$ depends on the priority relation $\ll_i$. According to an agent, an argument defeats another argument if they attack each other and the second argument is not stronger than the first one:

DEFINITION 5. *Let $AT_i = \langle \mathcal{T}, V, promote, \ll_i \rangle$ be the value-based argumentative theory of the agent $ag_i$ and $A = \langle P, c \rangle$, $B = \langle P', c' \rangle \in \mathcal{A}(\mathcal{T})$ two arguments.*
*$A$ **defeats** $B$ **for $AT_i$** (written $defeats_i(A, B)$) iff*
*$\exists P_1 \subseteq P, P_2 \subseteq P'$ such as :*

1. *$P_1 \vdash_\mho L$ and $P_2 \vdash_\mho \neg L$;*
2. *$\neg(level_i(P_1) \ll_i level_i(P_2))$.*

*Similarly, we say that a set $S$ of arguments defeats $B$ if $B$ is defeated by an argument in $S$.*

Contrary to the relation of attack, the relation of defeat is asymmetric. Moreover, the attack is an objective relation and the defeat relation is a subjective one. Because we are interested in the individuated viewpoint of each agent, we focus on the following notion of acceptance:

DEFINITION 6. *Let $AT_i = \langle \mathcal{T}, V, promote, \ll_i \rangle$ be the value-based argumentative theory of the agent $ag_i$. Let $A \in \mathcal{A}(\mathcal{T})$ be an argument and $S \subseteq \mathcal{A}(\mathcal{T})$ a set of arguments. $A$ **is subjectively acceptable by $AT_i$ wrt** $S$ iff*
*$\forall B \in \mathcal{A}(\mathcal{T})$ $defeats_i(B, A) \Rightarrow defeats_i(S, B)$;*

EXAMPLE 1. *Let us consider two American citizen arguing about the new president. The value-based argumentative theory of the agent $ag_1$ (resp. $ag_2$) is represented at left (resp. at right) of the figure 1.*
*The two agents share a theory, i.e. a set of rules ($r_{11}, \ldots, r_6$) and a set of values ($v_1, \ldots, v_6$). The rules corresponding to the goal relate to the value $v_1$. The common sense rules relate to the value $v_2$. The other rules, which specify particular opinions, relate to the values $v_3, \ldots, v_6$. According to an agent, a value above another one has priority over it. Because the theory is inconsistent, the five following arguments conflict :*
$A_1 = (\{r_6, r_3(bush)\}, pres(bush) \leftarrow)$
$A_2 = (\{r_5, r_{22}(kerry), r_{11}\}, pres(bush) \leftarrow)$
$A'_2 = (\{r_5, r_{22}(kerry)\}, \neg pres(kerry) \leftarrow)$
$B = (\{r_4, r_{12}(bush), r_{22}(bush), r_{21}\}, pres(kerry) \leftarrow)$
$B' = (\{r_4, r_{12}(bush), r_{22}(bush)\}, \neg pres(bush) \leftarrow)$
*$A'_2$ is a sub-argument of $A_2$ and $B'$ is a sub-argument of $B$.*
*Let us consider the value-based argumentative theory of the agent $ag_1$ (cf figure 1). The strength of $A_1$ is $v_3$ and the*

| $\ll_1$ | $V$ | premise($A_1$) | premise($A_2$) | premise($A'_2$) | premise($B$) | premise($B'$) |
|---|---|---|---|---|---|---|
| ↑ | $v_1$ | | $r_{11}$ | | $r_{21}$ | |
| | $v_2$ | | $r_{22}$(kerry) | $r_{22}$(bush) | $r_{12}$(bush) $r_{22}$(bush) | $r_{12}$(bush) $r_{22}$(bush) |
| | $v_6$ | $r_6$ | | | | |
| | $v_5$ | | $r_5$ | $r_5$ | | |
| | $v_4$ | | | | $r_4$ | $r_4$ |
| | $v_3$ | $r_3$(bush) | | | | |

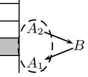



**Figure 2:** The value-based argumentative theory of the first agent

*strength of $B'$ is $v_4$. $B$ defeats $A_1$ but $A_1$ does not defeat $B$. The strength of $A'_2$ is $v_5$ and the strength of $B$ is $v_4$. $A_2$ defeats $B$ but $B$ do not defeat $A_2$. The set $\{A_1 A_2\}$ is subjectively acceptable wrt $\mathcal{A}(\mathcal{T})$.*
*Let us consider the value-based argumentative theory of the agent $ag_2$ (cf figure 1).*
*The strength of $A_1$ is $v_6$ and the strength of $B'$ is $v_4$. $B$ defeats $A_1$ but $A_1$ do not defeat $B$. The strength of $A_2$ is $v_5$ and the strength of $B'$ is $v_4$. $B$ defeats $A_2$ but $A_2$ do not defeat $B$. The set $\{B\}$ is subjectively acceptable wrt $\mathcal{A}(\mathcal{T})$.*

In such an argumentation framework, the knowledge and the values are common but the values have different priorities for different agents. At the opposite, the knowledge is distributed in a system of argumentative agents.

| $\ll_1$ | $V$ | $\mathcal{T}$ |
|---|---|---|
| ↑ | $v_1$ | $r_{11}$ : pres(bush) $\leftarrow \neg$(pres(kerry)) |
| | | $r_{21}$ : pres(kerry) $\leftarrow \neg$(pres(bush)) |
| | $v_2$ | $r_{12}(x)$ : weak($x$) $\leftarrow$ silly($x$) |
| | | $r_{22}(x)$ : $\neg$pres($x$) $\leftarrow$ weak($x$) |
| | $v_6$ | $r_6$ : current_pres(bush) $\leftarrow$ |
| | $v_5$ | $r_5$ : weak(kerry) $\leftarrow$ |
| | $v_4$ | $r_4$ : silly(bush) $\leftarrow$ |
| | $v_3$ | $r_3(x)$ : pres($x$) $\leftarrow$ current_pres($x$) |

| $\ll_2$ | $V$ | $\mathcal{T}$ |
|---|---|---|
| ↑ | $v_1$ | $r_{11}$ : pres(bush) $\leftarrow \neg$(pres(kerry)) |
| | | $r_{21}$ : pres(kerry) $\leftarrow \neg$(pres(bush)) |
| | $v_2$ | $r_{12}(x)$ : weak($x$) $\leftarrow$ silly($x$) |
| | | $r_{22}(x)$ : $\neg$pres($x$) $\leftarrow$ weak($x$) |
| | $v_3$ | $r_3(x)$ : pres($x$) $\leftarrow$ current_pres($x$) |
| | $v_4$ | $r_4$ : silly(bush) $\leftarrow$ |
| | $v_5$ | $r_5$ : weak(kerry) $\leftarrow$ |
| | $v_6$ | $r_6$ : current_pres(bush) $\leftarrow$ |

**Figure 1:** The value-based argumentative theory of the agents

| $\ll_2$ | $V$ | premise($A_1$) | premise($A_2$) | premise($A'_2$) | premise($B$) | premise($B'$) |
|---|---|---|---|---|---|---|
| ↑ | $v_1$ | | $r_{11}$ | | $r_{21}$ | |
| | $v_2$ | | $r_{22}$(kerry) | $r_{22}$(bush) | $r_{12}$(bush) $r_{22}$(bush) | $r_{12}$(bush) $r_{22}$(bush) |
| | $v_3$ | $r_3$(bush) | | | | |
| | $v_4$ | | | | $r_4$ | $r_4$ |
| | $v_5$ | | $r_5$ | $r_5$ | | |
| | $v_6$ | $r_6$ | | | | |

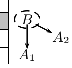



**Figure 3:** The value-based argumentative theory of the second agent

## 3. SYSTEM OF ARGUMENTATIVE AGENTS

Because the beliefs of social agents can be common, complementary or contradictory, agents argue to exchange hypothesis and reason together. For this purpose, the system of argumentative agents which is proposed in this section is similar to the AMP framework [2, 13]. Moreover, agents valuate the perceived commitments with respect to the estimated competence of the agents from whom the information is obtained.

The argumentative agents, which have their own beliefs and their own values, record the commitments of their interlocutors. Moreover, each argumentative agent valuates the reputation of her interlocutors. Therefore, an argumentative agent is in conformance with the following definition:

DEFINITION 7. *The* ***argumentative agent*** $ag_i \in \mho_A$ *is defined by a 6-tuple* $ag_i = \langle \mathcal{T}_i, V_i, \ll_i, promote_i, \cup_{j\neq i} CS^i_j, \prec_i \rangle$ *where:*

- $\mathcal{T}_i$ *is a personal theory, i.e. a set of personal rules in* $\mathcal{L}_\mho$*;*
- $V_i$ *is a set of personal values;*
- $promote_i : \mathcal{T}_i \to V_i$ *maps from the personal rules to the personal values;*
- $\ll_i$ *is the priority relation, i.e. a strict complete ordering relation on* $V_i$*;*
- $CS^i_j$ *is a commitment store, i.e. a set of rules in* $\mathcal{L}_\mho$*.* $CS^i_j(t)$ *contains commitments taken before or at time* $t$*, where agent* $ag_j$ *is the debtor and agent* $ag_i$ *the creditor;*
- $\prec_i$ *is the reputation relation, i.e. a strict complete ordering relation on* $\mho_A$*.*

The personal theories are not necessarily disjoint. We call common theory the set of rules explicitly shared by the agents: $\mathcal{T}_{\Omega_A} \subseteq \cap_{\mathrm{ag}_i \in \mho_A} \mathcal{T}_i$. The personal theories can be complementary or contradictory. We call joint theory the set of rules distributed in the system: $\mathcal{T}_{\mho_A} = \cup_{\mathrm{ag}_i \in \mho_A} \mathcal{T}_i$. Similarly, the sets of personal values are not necessarily disjoint. We call common values the values explicitly shared by the agents: $V_{\Omega_A} \subseteq \cap_{\mathrm{ag}_i \in \mho_A} V_i$. The personal rules relate to the personal values. For every $r \in \mathcal{T}_i, \mathrm{promote}_i(r) = v \in V_i$. On the one hand, the common rules relate to the common values. For every $r \in \mathcal{T}_{\Omega_A}$, $\mathrm{promote}_{\Omega_A}(r) = v \in V_{\Omega_A}$. On the other hand, the agent own rules relate to the agent own values. For every $r \in \mathcal{T}_i - \mathcal{T}_{\Omega_A}, \mathrm{promote}_i(r) = v \in V_i - V_{\Omega_A}$.

Reputation is a social concept that links an agent to her interlocutors. It is also a leveled relation [8]. The individuated reputation relations, which are transitive, irreflexive, asymmetric, and complete relations on $\mho_A$, preserve these properties. $\mathrm{ag}_j \prec_i \mathrm{ag}_k$ denotes that an agent $\mathrm{ag}_i$ trusts an agent $\mathrm{ag}_k$ more than another agent $\mathrm{ag}_j$.

In order to take into account the rules notified in the commitment stores, the agents are associated to the following extended theory:

DEFINITION 8. *The* ***extended theory of the argumentative agent*** $ag_i$ *is the value-based argumentative theory* $AT^*_i = \langle \mathcal{T}^*_i, V^*_i, promote^*_i, \ll^*_i \rangle$ *where:*

- $\mathcal{T}^*_i = \mathcal{T}_i \cup [\bigcup_{j\neq i} CS^i_j]$ *is the extended personal theory of the agent composed of the personal theory of this agent and the set of perceived commitments;*
- $V^*_i = V_i \cup [\bigcup_{j\neq i} \{v^i_j\}]$ *is the extended set of personal values of the agent composed of the set of personal values of this agent and the reputation values of her interlocutors;*
- $promote^*_i : \mathcal{T}^*_i \to V^*_i$ *is the extension of the function* $promote_i$ *which maps from the rules in the extended personal theory to the extended set of personal values. On the one hand, the personal rules relate to the personal values. On the other hand, the rules in the commitment store* $CS^i_j$ *relate to the reputation value* $v^i_j$*;*
- $\ll^*_i$ *is the extended priority relation of the agent, i.e. a strict complete ordering relation on* $V^*_i$*.*

When the values are common, they are universal as true. When the values are agent's own, they are in conformance with her particular aspiration [6]. That is the reason why the common values have priority over the other values. Since the agent argues, the personal values will have priority over the reputation values. Henceforth, the extended priority relation of the agent is constrained as follows:

$$\forall \mathrm{ag}_j \in \mho_A \ \forall v_\omega \in V_{\Omega_A} \forall v \in V_i - V_{\Omega_A} \ (v^i_j \ll^*_i v \ll^*_i v_\omega)$$

An argument is acceptable by the argumentative agent $\mathrm{ag}_i$ if it is subjectively acceptable by $\mathrm{AT}^*_i$ wrt the extended set of arguments $\mathcal{A}(\mathcal{T}_i^*)$. $\mathcal{S}_i^*$ denotes the set of acceptable arguments for the argumentative agent $\text{ag}_i$. This agent is convinced by a rule $r$ if it is the conclusion of an acceptable argument: $\exists A \in \mathcal{S}_i^*$ conclusion$(A) = r$.

The agents utter messages to exchange their beliefs. The syntax of messages is in conformance with the **communication language**, $\mathcal{CL}_\mho$. A message $M_k = \langle S_k, H_k, A_k \rangle \in \mathcal{CL}_\mho$ has an identifier $M_k$. It is uttered by a speaker ($S_k = \text{speaker}(M_k)$) and addressed to a hearer ($H_k = \text{hearer}(M_k)$), *i.e.* one agent in the audience. $A_k = \text{act}(M_k)$ is the speech act of the message. It is composed of a locution and a content. The locution is one of the following: question, assert, unknow, concede, challenge, withdraw. The content, also called **hypothesis**, is a rule or a set of rules in $\mathcal{L}_\mho$.

The speech acts have an argumentative and social semantics [5]. Because a commitment enrich the extended theory of the creditor, the speech acts have a social semantics. Because a commitment could be justified by the extended theory of the debtor, the speech acts have an argumentative semantics.

For example, the figure 4 shows the axiomatic semantics associated with the assertion of an hypothesis. An agent can assert a hypothesis if she has an argument for it. The corresponding commitments stores are updated.

- Message:
  $M_l = \langle \text{ag}_i, \text{ag}_j, \text{assert}(h) \rangle$
- Argumentative semantics:
  $\exists A \in \mathcal{A}(\mathcal{T}_i^*)$ conclusion$(A) = h$
- Social semantics:
  for any agent $\text{ag}_k$ in the audience
  $\text{CS}_i^k(t) = \text{CS}_i^k(t-1) \cup \{h\}$

**Figure 4:** Axiomatic semantics for assert an hypothesis $h$ at time $t$

In a similar way, the figure 5 shows the axiomatic semantics associated with the concession of an hypothesis. The rational condition for the assertion and the rational condition for the concession of the same hypothesis by the same agent distinguish themselves. Agents can assert hypothesis whether they are supported by a trivial argument or not. By contrast, agents do not concede all the hypothesis she hears in spite of they are all supported by a trivial argument.

- Message:
  $M_l = \langle \text{ag}_i, \text{ag}_j, \text{concede}(h) \rangle$
- Argumentative semantics:
  $\exists A \in \mathcal{A}(\mathcal{T}_i^*)$ conclusion$(A) = h$ with
  $(\text{premise}(A) \neq \{h\} \wedge \text{premise}(A) \not\subseteq \cup_{j \neq i} \text{CS}_j^i)$
- Social semantics:
  for any agent $\text{ag}_k$ in the audience
  $\text{CS}_i^k(t) = \text{CS}_i^k(t-1) \cup \{h\}$

**Figure 5:** Axiomatic semantics for concede an hypothesis $h$ at time $t$

Because the other speech acts (question$(h)$, challenge$(h)$, unknow$(h)$, and withdraw$(h)$) have no particular effects on the commitments stores, they have neither particular rational conditions of utterance. In other words, we considere that the commitments stores are cumulative [10, 16]. We can note that the rational conditions for the assertion of a hypothesis and for its explicit negation are not necessary mutually excluded. Since each argumentative agent has different priorities, they will be thoughtful[2]. In other words, an agent can assert any hypothesis for which she has an acceptable argument.

The hypothesis which are received must be valuated. Since the agents are more or less authoritative regarding the domain, the commitments will be considered in accordance with the estimated competence of the agents from whom the information is obtained.

Example 2. *If the argumentative agent $ag_1$ utters the following message:*

$$M_1 = \langle ag_1, ag_2, assert(pres(bush) \leftarrow) \rangle$$

*the extended theory of the argumentative agent $ag_2$ is represented in the figure 6. The extended personal theory is com-*

| $\ll_2^*$ | $V_2^*$ | $\mathcal{T}_2^*$ | $\mathcal{T}_{\Omega_A}$ |
|---|---|---|---|
| ↑ | $v_1$ | $r_{11}$ : pres(bush) $\leftarrow \neg$(pres(kerry))<br>$r_{21}$ : pres(kerry) $\leftarrow \neg$(pres(bush)) | |
| | $v_2$ | $r_{12}(x)$ : weak$(x) \leftarrow$ silly$(x)$<br>$r_{22}(x)$ : $\neg$pres$(x) \leftarrow$ weak$(x)$ | |
| | $v_3$ | $r_3(x)$ : pres$(x) \leftarrow$ current_pres$(x)$ | |
| | $v_4$ | $r_4$ : silly(bush) $\leftarrow$ | |
| | $v_5$ | $r_5$ : weak(kerry) $\leftarrow$ | |
| | $v_6$ | $r_6$ : current_pres(bush) $\leftarrow$ | |
| | $v_1^2$ | $\text{CS}_1^2 \supseteq r_7$ : pres(bush) $\leftarrow$ | |

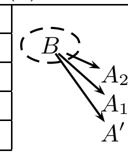



**Figure 6:** The extended theory of the second argumentative agent

*posed of the personal theory of this agent and the hypothesis advanced by the agent $ag_1$: $pres(bush) \leftarrow$. The extended set of personal values is composed of the set of personal values and the reputation value of the agent $ag_2$. The rules, which correspond to the goal or the common sense, consist of the common theory which relate to the common values: $V_{\Omega_A} = \{v_1, v_2\}$.*

*The arguments $A_1$, $A_2$, and $A'$ support bush. The argument $B$ support kerry. $A'$ is a trivial argument based on the commitment stores which supports bush. Therefore, the hypothesis $pres(bush) \leftarrow$ cannot be conceded by the agent $ag_2$ wrt this argument but wrt $A_1$ or $A_2$. Because the agent $ag_2$ has conflicting arguments she can assert the two hypothesis $pres(kerry) \leftarrow$ and $pres(bush) \leftarrow$. If she is thoughtful, she only assert $pres(bush) \leftarrow$.*

In such a system of argumentative agents, the knowledge is exchanged but the agents have different convictions. At the opposite, a dialectics system makes it possible to reach an agreement.

## 4. DIALECTICS SYSTEM

When a set of social and autonomous agents argue, they reply to each other in order to reach the goal of the interaction. In this section we present the formal framework in which agents play and arbitrate to reach an agreement. The dialectics system which is proposed in this section is similar to [14]. Moreover, a third agent arbitrates in accordance with the estimated competence of the players.

During exchanges, the speech acts are not isolated but they respond each other. The moves are messages with some attributes to control the sequence. The syntax of moves is in conformance with the **moves language**. A move $\text{move}_k = \langle M_k, R_k, P_k \rangle \in \mathcal{ML}_{\mho}$ has an identifier $\text{move}_k$. It contains a message $M_k$ as defined before. $R_k = \text{reply}(\text{move}_k)$ is the identifier of the move to which $\text{move}_k$ responds. A move ($\text{move}_k$) is either an initial move ($\text{reply}(\text{move}_k) = \text{nil}$) or a replying move ($\text{reply}(\text{move}_k) \neq \text{nil}$). $\text{P}_k = \text{protocol}(\text{move}_k)$ is the name of the protocol which is used.

A dialectics system is composed of three agents. In this formal framework, two players play moves in front of a third agent to check the validity of an initial hypothesis.

DEFINITION 9. *Let $AT_{\Omega_A} = \langle \mathcal{T}_{\Omega_A}, V_{\Omega_A}, promote_{\Omega_A} \rangle$ be a value-based argumentative theory and $r_0$ a rule in $\mathcal{L}_{\mho}$. The **dialectics system** on the topic $r_0$ is a 7-tuple $DS_{\Omega_M}(r_0, AT_{\Omega_A}) = \langle N, wit, H, T, convention, Z, (u_p)_{p \in N} \rangle$ where :*

- *$N = \{init, part\} \subset \mho_A$ is a set of two argumentative agents called players: the initiator and the partner;*
- *$wit \in \mho_A$ is a third argumentative agent, called witness, with a personal theory restricted to the common theory ($\mathcal{T}_{wit} = \mathcal{T}_{\Omega_A}$);*
- *$\Omega_M \subseteq \mathcal{ML}_{\mho}$ is a set of well-formed moves;*
- *$H$ is the set of histories, i.e. the sequences of well-formed moves s.t. the speaker of a move is determined at each stage by a turn-taking function and the moves agree with a convention;*
- *$T : H \rightarrow N$ is the turn-taking function determining the speaker of a move. If the length of the history is null or even then $T(h) = init$ else $T(h) = part$;*
- *$convention : H \rightarrow \Omega_M$ is the function determining the moves which are allowed or not to expand an history;*
- *$Z$ is the set of dialogues, i.e. the terminal histories which consist of maximally long histories;*
- *$u_{init}, u_{part} : H \rightarrow \{-1, 1\}$ are the partial utility functions determining if a player is a winner at a history.*

In order to be well-formed, the initial move is a question about the topic from the initiator to the partner and a replying move from a player references an earlier move uttered by the other player. In this way, the backtracks are allowed. The dialogue line is the sub-sequence of moves where all the backtracks are ignored. In order to avoid loops, the redundancy of hypothesis is forbidden in the assertions of the same dialogue line. Obviously, all the moves should contain the same value for the protocol parameter.

At the history $h$, the witness is associated to the extended theory
$\text{AT}^*_{\text{wit}}(h) = \langle \mathcal{T}^*_{\text{wit}}(h), V^*_{\text{wit}}, \text{promote}^*_{\text{wit}}, \ll^*_{\text{wit}} \rangle$ where:

- the extended personal theory is composed of the personal theory and the commitments of players: $\mathcal{T}^*_{\text{wit}}(h) = \mathcal{T}_{\Omega_A} \cup \text{CS}^{\text{wit}}_{\text{init}}(h) \cup \text{CS}^{\text{wit}}_{\text{part}}(h)$;
- the extended set of values is composed of the set of personal values and the reputation values of the two players: $V^*_{\text{wit}} = V_{\Omega_A} \cup \{v^{\text{wit}}_{\text{init}}, v^{\text{wit}}_{\text{part}}\}$.

$\mathcal{S}^*_{\text{wit}}(h)$ (resp. $\mathcal{S}^*_p(h)$) denotes the set of acceptable arguments for the witness (resp. one player).

The witness is responsible of the stable agreement. The acceptability of arguments by the witness depends on the reputation of the players. In other words, the arbitrage of the witness depends on the arguments exchanged and the estimated competence of the players. We said that $r_0$ is **provable at the history** $h$ (written $\text{provable}^h(r_0)$) if the witness is convinced by $r_0$ at the history $h$.

Since the witness arbitrates, she will attribute the victory to one, two or none of the players in accordance with their initial convictions. The witness decides if a player is a winner or not at a history:

- if $\text{provable}^h(r_0)$ then $(u_p(h) = 1 \Leftrightarrow \exists A \in \mathcal{S}^*_p(h_0)\ \text{conclusion}(A) = r_0)$;
- if $\text{provable}^h(\neg r_0)$ then $(u_p(h) = 0 \Leftrightarrow \exists A \in \mathcal{S}^*_p(h_0)\ \text{conclusion}(A) = r_0)$.

Because the definition of the dialectics system make no assumptions about the initial convictions of players, we can consider any dialogue which the final outcome is an agreement: persuasion, information-seeking, inquiry or any combination of these basic dialogues [16].

## 5. PERSUASION

When a set of social and autonomous agents argue, they collaborate to confront their convictions. In this section we illustrate our dialectics system with the protocol of persuasion proposed in [2] where agents resolve their conflict by verbal means.

The persuasion protocol (denoted PP) is an unique-respond protocol where players can reply just once to the other player's moves. In other words, the backtracks are forbidden and all the histories are dialogue lines. Therefore, the convention is restricted to the set of sequence rules represented in the table 5. Each rule specifies the authorized replying moves. For example, the rule of "Assertion/Refutation" (written $\text{sr}_{A/R}$) specifies the authorized moves replying to the previous assertion ($\text{assert}(H)$). The speech acts resist or surrender to the previous one. The surrendering acts close the dialogue. A concession ($\text{concede}(H)$) surrenders to the previous assertion. A challenge ($\text{challenge}(h)$) and the assertion of its explicit negation ($\text{assert}(\neg h)$) resist to the previous assertion.

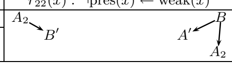
| Sequences rules | Speech acts | Resisting replies | Surrendering replies |
|---|---|---|---|
| $\text{sr}_{Q/A}$ | $\text{question}(\{r_0\})$ | $\text{assert}(\{r_0\})$<br>$\text{assert}(\{\neg r_0\})$ | $\text{unknow}(\{r_0\})$ |
| $\text{sr}_{A/R}$ | $\text{assert}(H)$ | $\text{challenge}(h)$, $h \in H$<br>$\text{assert}(\neg h)$, $h \in H$ | $\text{concede}(H)$ |
| $\text{sr}_{C/A}$ | $\text{challenge}(h)$ | $\text{assert}(H)$, $H \vdash_{\mho} h$ | $\text{withdraw}(h)$ |
| $\text{sr}_T$ | $\text{unknow}(H)$ | $\emptyset$ | $\emptyset$ |
| | $\text{concede}(H)$ | $\emptyset$ | $\emptyset$ |
| | $\text{withdraw}(H)$ | $\emptyset$ | $\emptyset$ |

**Table 1: Set of speech acts and the potential answers.**

We can note that the rational conditions of utterances for the allowed replying moves are not necessary mutually excluded. Since the players are *open-minded*, they will try to give their opinions, *i.e* a concession ($\text{concede}(H)$) or a refutation ($\text{assert}(\neg h)$), in replying to the assertions [2]. Since the players are *cooperative*, they will give their opinions, *i.e.* a confirmation ($\text{assert}(\{r_0\})$) or an invalidation (assert($\{\neg r_0\}$)), in replying to a question. The players resist in the debate.

An agent initiates a persuasion when she wants confront her viewpoint with the partner. If the partner has no arguments against/for the topic, she pleads ignorance and closes the dialogue. If the players have the same convictions, the witness is convinced and the dialogue closes. Otherwise, the goal of the dialogue is the resolution of the conflict by verbal means.

Example 3. *Let us consider the persuasion about the candidate kerry. The extended argumentative theories of the players are represented in the figure 3. The initiator is open-minded and the partner is cooperative. The commitments stores result from the sequence of moves. The arbitrage of*

| $\mathcal{T}_{\mho_A}$ | | | | |
|---|---|---|---|---|
| $V^*_{\text{init}}$ | $\mathcal{T}^*_{\text{init}}$ | $\mathcal{T}_{\Omega_A}$ | $\mathcal{T}^*_{\text{part}}$ | $V^*_{\text{part}}$ |
| $v_1$ | | $r_{11}$ : pres(bush) $\leftarrow \neg$(pres(kerry))<br>$r_{21}$ : pres(kerry) $\leftarrow \neg$(pres(bush)) | | $v_1$ |
| $v_2$ | | $r_{12}(x)$ : weak($x$) $\leftarrow$ silly($x$)<br>$r_{22}(x)$ : $\neg$pres($x$) $\leftarrow$ weak($x$) | | $v_2$ |
| $v_5$ | $r_5(x)$ : weak(kerry) $\leftarrow$ |  | $r_4$ : silly(bush) $\leftarrow$ | $v_4$ |
| $v^{\text{init}}_{\text{part}}$ | pres(kerry) $\leftarrow$ | | $\neg$pres(kerry) $\leftarrow$<br>weak(kerry) $\leftarrow$ | $v^{\text{part}}_{\text{init}}$ |

$$\text{move}_1 = \langle \text{init}, \text{part}, \emptyset, \text{PP}, \text{question}(\text{pres}(\text{kerry}) \leftarrow) \rangle$$
$$\text{move}_2 = \langle \text{part}, \text{init}, \text{move}_1, \text{PP}, \text{assert}(\text{pres}(\text{kerry}) \leftarrow) \rangle$$
$$\text{move}_3 = \langle \text{init}, \text{part}, \text{move}_2, \text{PP}, \text{assert}(\neg\text{pres}(\text{kerry}) \leftarrow) \rangle$$
$$\text{move}_4 = \langle \text{part}, \text{init}, \text{move}_3, \text{PP}, \text{challenge}(\neg\text{pres}(\text{kerry}) \leftarrow) \rangle$$
$$\text{move}_5 = \langle \text{init}, \text{part}, \text{move}_4, \text{PP}, \text{assert}(\{r_5, r_{22}(\text{kerry})\}) \rangle$$
$$\text{move}_6 = \langle \text{part}, \text{init}, \text{move}_5, \text{PP}, \text{challenge}(\text{weak}(\text{kerry})) \rangle$$
$$\text{move}_7 = \langle \text{init}, \text{part}, \text{move}_6, \text{PP}, \text{withdraw}(\text{weak}(\text{kerry})) \rangle$$

**Figure 7:** An example of persuasive dialogue

*the witness depends on the estimated competence of the players. If the partner is considered as more competent than the initiator, pres(kerry) $\leftarrow$ is provable when the dialogue closes. Otherwise, $\neg$pres(kerry) $\leftarrow$ is provable.*

## 6. CONCLUSIONS

This paper has been a technical exploration of the argumentative techniques for the formalization of a collective decision-making with a justified outcome. For this purpose, we have proposed a formal framework in which agents play and arbitrate to reach an agreement. The argumentation-based reasoning which has been proposed manage the conflicts between arguments having different strengths for different agents. The argumentative agents justify the hypothesis to which they commit and take into account the commitments of their interlocutors. A third agent is responsible of the final decision outcome and resolves the conflict between two players according to their competence. As the dialectics framework is well-defined, we are able to investigate the proprieties of dialogues such as the correctness and the completeness.

The dialectics process proposed in this paper compose the interests and perspectives. However, we cannot deny the importance of collective learnings to enrich the political debate. A rational citizen would determine his preferences after learning the preferences of others. For this purpose, the priorities must be dynamic and determined in the course of the dialogue.

## 7. REFERENCES


[1] Leila Amgoud and Claudette Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Maths and AI*, 34(1-3):197–215, 2002.

[2] Leila Amgoud, Nicolas Maudet, and Simon Parsons. An argumentation-based semantics for agent communication languages. In F. van Harmelen, editor, *Proc. of the 15th European Conference on Artificial Intelligence*, pages 38–42. IOS Press, Amsterdan, 2002.

[3] T.J.M Bench-Capon. Value based argumentation frameworks. In *Proceedings of Non Monotonic Reasoning*, pages 444–453, 2002.

[4] T.J.M Bench-Capon. Persuasion in practical argument using value based argument frameworks. *Journal of Logic and Computation*, 13(3):429–448, 2003.

[5] Jamal Bentahar, Bernard Moulin, and Brahim Chaib-draa. Specifying and implementing a persuasion dialogue game using commitments and arguments. In *Proc. of the workshop ArgMAS in AAMAS*, pages 130–148.

[6] Perelman C. and Olbrechts-Tyteca L. *Traité de l'Argumentation - La Nouvelle Rhétorique*. Presses Universitaires de France, 1958.

[7] Michel Callon, Pierre Lascoumes, and Yannick Barthe. *Agir dans un monde incertain*. Seuil, 2001.

[8] C. Castelfranchi and R. Falcone. Principles of trust in mas: Cognitive anatomy, social importance, and quantification. In *Proceedings of ICMAS'98*, pages 72–79, 1998.

[9] Frank Dignum, Barbara Dunin-Keplicz, and Rineke Verbrugge. Agent theory for team formation by dialogue. In Cristiano Castelfranchi and Yves Lesperance, editors, *Agent Theories Architectures and Languages*, number LNAI 1986 in Intelligent Agents VII, pages 150–166. Springer-Verlag, 2001.

[10] Charles L. Hamblin. *Fallacies*. Methuen, 1970.

[11] Antonis C. Kakas and Pavlos Moratis. Argumentative agent deliberation, roles and context. In Jurgen Dix, Joo Alexandre Leite, and Ken Satoh, editors, *Electronic Notes in Theoretical Computer Science*, volume 70. Elsevier, 2002.

[12] P. McBurney and S. Parsons. Intelligent systems to support deliberative democracy in environmental regulation. *Information and Communications Technology Law*, 10(1):33–43, 2001.

[13] Peter McBurney and Simon Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11(3):315–334, 2002.

[14] Henry Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In M. Ojeda-Aciego, I.P.d. Guzman, G. Brewka, and L.M. Pereira, editors, *Proc. of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'2000)*, number 1919 in Lecture Notes in AI, pages 224–238. Springer Verlag, 2000.

[15] J.R. Searle. *Speech Acts : An Essay in the Philosophy of Language*. Cambridge University Press, 1969.

[16] D. Walton and E. Krabbe. *Commitment in Dialogue*. SUNY Press, 1995.