Artificial Intelligence and Law (2006)
DOI 10.1007/s10506-006-9008-y © Springer 2006

Collective decision-making process to compose divergent interests and perspectives

MAXIME MORGE
Laboratoire d'Informatique Fondamentale de Lille, Batiment, M3 - F-59655, Villeneuve D'ascq, Cedex, France
E-mail: morge@lifl.fr

Abstract. We propose in this paper DIAL, a framework for inter-agents dialogue, which formalize a collective decision-making process to compose divergent interests and perspectives. This framework bounds a dialectics system in which argumentative agents play and arbitrate to reach an agreement. For this purpose, we propose an argumentation-based reasoning to manage the conflicts between arguments having different strengths for different agents. Moreover, we propose a model of argumentative agents which justify the hypothesis to which they commit and take into account the commitments of their interlocutors according to their reputations. In the scope of our dialectics system, a third agent is responsible of the final decision outcome which is taken by resolving the conflict between two players according to their competences and the advanced arguments.

Key words: Multi-agents system, dialogue

1. Introduction

In public decision such as land planning, the success of the decision outcome depends on the extent to which people believe it has been reached fairly. The actors must have a role in forming even if they disagree with the final outcome. The decisions must be collective and argued (McBurney and Parsons, 2001). This observation changes the way we appreciate the democratic procedures. We can distinguish two different modalities of formation of political will: the representative democracy and the *dialogical democracy*. On one hand, the representative democracy is a process in which individual preferences are aggregated to obtain outcome. The electors and the lay public delegate their power to the elect and the experts. On the other hand, the dialogical democracy is a participative fair effect process to compose the interests and perspectives (Callon et al., 2001). The civil society debate and deliberate. In this paper, we aim at formalizing such a process with a framework for inter-agent communication.

Most of the existing formal framework for inter-agent interaction are based on speech acts theory (Searle, 1969). 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 (Singh, 2000). (3) However isolated communicative acts do not suffice to achieve a common goal, the existing protocols are too rigid for the debate (Dignum et al., 2001).

By contrast, recent works (Prakken, 2000; Dung, 1995; Amgoud and Parsons, 2001; Bench-Capon, 2002; Amgoud and Cagrol, 2002; Schweimeier, 2002; McBurney and Parsons, 2002; Bench-Capon, 2003; Labrie et al., 2003) are inspired by the dialectics (Perelman and Olbrechts-Tyteca, 1958; Hamblin, 1970; Walton and Krabbe, 1995; Maudet, 2001). Not being a first attempt, this paper combines and reconciles these argumentative techniques in a coherent framework for the formalization of a dialogical democracy. We present in this paper a dialectics system, *i.e.* a formal framework in which agents communicate to reach a collective decision. (1) The argumentation-based reasoning mechanism manages the interaction between conflicting arguments. (2) Since the communication language has a public semantics, every agents confer the same meaning to the messages and any third agent is able to draw similar inferences. (3) The dialogue is a flexible and refined process to reach an agreement.

Paper overview. In Section 2, we show how the existing argumentative techniques can be combined to formalize a dialogical democracy. Section 3 presents the argumentation framework, *i.e.* the reasoning mechanism of agents. In accordance with this background, we describe in Section 4 our model of argumentative agents. In Section 5, we define the formal framework in which the agents collaborate. The Section 6 presents the dialogue protocol used to reach an agreement.

2. Motivation

In this section, we show why the argumentation framework, and the framework for inter-agent dialogues proposed here are different to the existing ones.

An argumentation framework offers way to compare arguments with a contradiction relation to determine their acceptances. As displayed in the Table I, this framework (Dung, 1995) has been extended. When the framework is built around an underlying logic language, arguments are not abstract entities but relations of consequence between a premise and a conclusion. Since we want to explicitly represent the knowledge of agents, we need an argumentation logic. When the framework considers one (or many) priority relation(s), arguments have one (or many) strength(s). Since each agent has her own priorities, we need a value-based argumentation framework. We propose in the Section 3 a value-based argumentation logic.

Table I. Existing argumentation frameworks

Priority/Language	No	Yes
None	Argumentation system (Dung, 1995)	Argumentation logic (Schweimeier, 2002)
One	Preference-based argumentation system (Amgoud and Cagrol, 2002)	Preference-based argumentation logic (Amgoud and Cagrol, 2002; Kakas and Moratis, 2002)
Many	Value-based argumentation system (Bench-Capon, 2002; Bench-Capon, 2003)	*Value-based argumentation logic* (Morge, 2005)

A framework for inter-agent dialogues is a framework for inter-agent communication inspired by the dialectics (Hamblin, 1970; Walton and Krabbe, 1995; Maudet, 2001). As displayed in the Table II, we distinguish the existing frameworks bounding a formal area where dialogues take place, and the ones which focus on the model of agents. Since we want to build a multi-agents system, we need a model of agents (cf Section 4). Since we want to warrant that an agreement is reached at the end of the dialogues, we need to bound a formal area (cf Section 5). DIAL is a coherent framework which combines these two aspects and reconciles the underlying concepts coming from the literature.

We present in the next section the underlying argumentation framework on which DIAL is built.

3. Argumentation framework

We present in this section a value-based argumentation logic, *i.e.* an argumentation framework built around an underlying logic language, and in which the agents individually evaluate the strength of arguments.

A multi-agents system is a set of social and autonomous agents (denoted $\mho_A = \{ag_1, \ldots, ag_n\}$). They share knowledge, *i.e.* a set of sentences in a common knowledge language, denoted $\mathcal{L}_\mho$. Moreover, the agents use the

Table II. Existing frameworks for inter-agent dialogues

Formal area/ Agent model	No	Yes
No	DIAGAL (Labrie et al., 2003)	AMP (Amgoud and Parsons, 2001)
Yes	Discussion (Prakken, 2000)	*DIAL* (Morge, 2005)

same classical inference, denoted $\vdash_{\mho}$. As in (Kakas and Moratis, 2002), the formulae of the background logic are defined as rules as $r : L_0 \leftarrow L_1, \ldots, L_n$ where $L_0 \ldots L_n$ are positive or explicit negative ground literals[1].

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

DEFINITION 1. (Value-based argumentative theory)
Let $\mho_A = \{ag_1, \ldots, ag_n\}$ be a set of agents. The *value-based argumentative theory* $AT_{\mho_A} = \langle \mathcal{T}, V, \text{promote} \rangle$ 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$(r) \in V$.

Since agents are individuated by their value hierarchies (Perelman and Olbrechts-Tyteca, 1958), the values have different priorities for different agents:

DEFINITION 2. (Value-based argumentative theory of an agent)
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, $\ll_i$ where:

- $AT_{\mho_A} = \langle \mathcal{T}, V, \text{promote} \rangle$ 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.

A priority relation is a transitive, irreflexive, asymmetric, and complete relation on V. It stratifies the theory into finite non-overlapping sets as in (Amgoud and Cagrol, 2002). On one hand, a priority relation captures the value hierarchy of a particular agent. On the other hand, the theory gathers the knowledge shared by the agents. The arguments are built on this knowledge.

An argument is a relation of consequence between a premise and a conclusion:

DEFINITION 3. (Argument) 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:

1. P is consistent and minimal (for set inclusion);

2. $P \vdash_{\mho} c$.
   P is the premise of A, written P = premise(A). c is the conclusion of A, denoted c = conclusion(A).

In other words, the premise is a set of rules from which the conclusion can be inferred.

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.

DEFINITION 4. (Attacks) 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 iff:
$\exists P_1 \subseteq P, P_2 \subseteq P'$ such as $P_1 \vdash_{\mho} L$ and $P_2 \vdash_{\mho} \neg L$.
The relation of attack is symmetric and objective, i.e. independent of the considered agent.

Because each agent is associated with a particular priority relation, the agents individually evaluate the strength of arguments:

DEFINITION 5. (Strength of an argument)
Let $\text{AT}_i = \langle T, V, \text{promote}, \ll_i\rangle$ be the value-based argumentative theory of the agent ag_i *and* $A = \langle P, c\rangle \in \mathcal{A}(\mathcal{T})$ an argument. According to AT_i, the *strength of A* (written $\text{strength}_i(A)$) is the least important value promoted by one rule in the premise.
In other words, the strength of arguments depends on the priority relation.

Since the agents individually evaluate the strength of arguments, an agent can ignore the attack of an argument over another argument. 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 6. (Defeats for an agent) Let $\text{AT}_i = \langle T, V, \text{promote}, \ll_i\rangle$ 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 $\text{defeats}_i(A, B)$) 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. $\neg(\text{level}_i(P_1) \ll_i \text{level}_i(P_2))$.
   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 and subjective. Considering the individuated viewpoint of each agent, we focus on the following notion of acceptance:

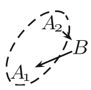
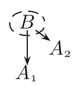
DEFINITION 7. (Subjective acceptance)
Let $\mathrm{AT}_i = \langle T, V, \text{promote}, \ll_i \rangle$ be the value-based argumentative theory of the agent ag_i. Let $A \in \mathcal{A}(\mathcal{T})$ be an argument and $S \subseteq \mathcal{A}(\mathcal{T})$ a set of arguments.
A is subjectively acceptable by AT_i with respect to S iff $\forall B \in \mathcal{A}(\mathcal{T})$ $\text{defeats}_i(B, A) \Rightarrow \text{defeats}_i(S, B)$.

The set of subjectively acceptable arguments consists of a consistent position, also called preferred extension, which is a maximal set of admissible arguments (Dung, 1995). In other words, this set defends itself from all attacks, and cannot be extended without introducing a conflict. Since the priority relation is an ordering relation, the preferred extension is unique and non-empty (Bench-Capon, 2002). The following example illustrate this argumentation framework.

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 Table III. 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:

- $A_1 = (\{r_6, r_3(bush)\}, pres(bush) \leftarrow)$;
- $A_2 = (\{r_5, r_{22}(kerry), r_{11}\}, pres(bush) \leftarrow)$;

Table III. The value-based argumentative theory of the agents

$\ll_1$	V	$\mathcal{T}$		$\ll_2$	V	$\mathcal{T}$	
↑	v_1	$r_{11} : pres(bush) \leftarrow \neg(pres(kerry))$		↑	v_1	$r_{11} : pres(bush) \leftarrow \neg(pres(kerry))$	
		$r_{21} : pres(kerry) \leftarrow \neg(pres(bush))$				$r_{21} : pres(kerry) \leftarrow \neg(pres(bush))$	
	v_2	$r_{12}(x) : weak(x) \leftarrow silly(x)$			v_2	$r_{12}(x) : weak(x) \leftarrow silly(x)$	
		$r_{22}(x) : \neg\ pres(x) \leftarrow weak(x)$				$r_{22}(x) : \neg\ pres(x) \leftarrow weak(x)$	
	v_6	$r_6 : current\_pres(bush) \leftarrow$			v_3	$r_3(x) : pres(x) \leftarrow current\_pres(x)$	
	v_5	$r_5 : weak(kerry) \leftarrow$			v_4	$r_4 : silly(bush) \leftarrow$	
	v_4	$r_4 : silly(bush) \leftarrow$			v_5	$r_5 : weak(kerry) \leftarrow$	
	v_3	$r_3(x) : pres(x) \leftarrow current\_pres(x)$			v_6	$r_6 : current\_pres(bush) \leftarrow$	

– $A'_2 = (\{r_5, r_{22}(kerry)\}, \neg pres(kerry) \leftarrow)$;
– $B = (\{r_4, r_{12}(bush), r_{22}(bush), r_{21}\}, pres(kerry) \leftarrow)$;
– $B' = (\{r_4, r_{12}(bush), r_{22}(bush)\}, \neg pres(bush) \leftarrow)$..

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. The strength of A_1 is v_3 and the strength of B' is v_4. Therefore, 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. Therefore, A_2 defeats B but B do not defeat A_2. The set $\{A_1 A_2\}$ is subjectively acceptable with respect to $\mathcal{A}(\mathcal{T})$.

Let us consider the value-based argumentative theory of the agent ag_2. The strength of A_1 is v_6 and the strength of B' is v_4. Therefore, B defeats A_1. The strength of A_2 is v_5 and the strength of B' is v_4. Therefore, B defeats A_2. The set $\{B\}$ is subjectively acceptable with respect to $\mathcal{A}(\mathcal{T})$.

We have defined here the reasonning mechanism of agents. In the next section, we present a model of agents which take into account arguments coming from their interlocutors.

4. Model of agents

Since the beliefs of agents can be common, complementary or contradictory, the agents exchange hypothesis and argue. For this purpose, the model of agents which is proposed in this section is similar to the AMP[2] model (Amgoud and Parsons, 2001; McBurney and Parsons, 2002). Contrary to the AMP model, our agents individually valuate the perceived commitments with respect to the estimated reputation of the agents from whom the information is obtained.

The argumentative agents, which have their own beliefs, record the commitments of their interlocutors. Moreover, the argumentative agents individually valuate the reputation of their interlocutors. Therefore, an argumentative agent is in conformance with the following definition:

DEFINITION 8. (Argumentative Agent)
The *argumentative agent* $\text{ag}_i \in \mho_A$ is defined by a 6-tuple
$\text{ag}_i = \langle \mathcal{T}_i, V_i, \ll_i, \text{promote}_i, \cup_{j \neq i} \text{CS}^i_j, \prec_i \rangle$ 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;
– $\text{promote}_i : \mathcal{T}_i \rightarrow V_i$ 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 $\mathcal{L}_\mho$.$\text{CS}^i_j(t)$ 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: $\mathcal{T}_{\Omega_A} \subseteq \bigcap_{\text{ag}_i \in \mho_A} \mathcal{T}_i$. Similary, we call *common values* the values explicitly shared by the agents: $V_{\Omega_A} \subseteq \bigcap_{\text{ag}_i \in \mho_A} V_i$. The common rules relate to the common values. For every $r \in \mathcal{T}_{\Omega_A}$, $\text{promote}_{\Omega_A}(r) = v \in V_{\Omega_A}$. The personal theories can be complementary or contradictory. We call *joint theory* the set of rules distributed in the system: $\mathcal{T}_{\mho_A} = \bigcup_{\text{ag}_i \in \mho_A} \mathcal{T}_i$. The agent own rules relate to the agent own values. For every $r \in \mathcal{T}_i - \mathcal{T}_{\Omega_A}$, $\text{promote}_i(r) = v \in V_i - V_{\Omega_A}$.

We can distinguish two ways for an agent to valuate the commitments of her interlocutors: either in accordance with a global social order as in (Amgoud and Parsons, 2001), or in accordance with a local perception of the interlocutor: the reputation. Obviously, this way is more flexible. Reputation is a social concept that links an agent to her interlocutors. It is also a leveled relation (Castelfranchi and Falcone, 1998). The individuated reputation relations, which are transitive, irreflexive, asymmetric, and complete relations on $\mho_A$, preserve these properties. $\text{ag}_j \prec_i \text{ag}_k$ denotes that an agent ag_i trusts an agent ag_k more than another agent ag_j.

In order to take into account the rules notified in the commitment stores, each agent is associated with the following extended theory:

DEFINITION 9. (Extended theory of an agent)
The *extended theory of the argumentative agent ag $_i$* is the value-based argumentative theory $\text{AT}_i^* = \langle T_i^*, V_i^*, \text{promote}_i^*, \ll_i^* \rangle$ where:

- $\mathcal{T}_i^* = \mathcal{T}_i \cup [\bigcup_{j \neq i} \text{CS}_j^i]$ is the extended personal theory of the agent composed of the personal theory and the set of perceived commitments;
- $V_i^* = V_i \cup [\bigcup_{j \neq i} \{v_j^i\}]$ is the extended set of personal values of the agent composed of the set of personal values and the reputation values associated with her interlocutors;
- $\text{promote}_i^* : \mathcal{T}_i^* \rightarrow V_i^*$ 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 one hand, the personal rules relate to the personal values. On the other hand, the rules in the commitment store CS_j^i relate to the reputation value v_j^i;
- $\ll_i$* is the extended priority relation of the agent, i.e. an ordering relation on V_i*.

Since the debate is a collaborative process, the agents share common rules (goal, common sense, ...) of prime importance. That is the reason why the common values have priority over the other values (Perelman and Olbrechts-Tyteca, 1958). Since the agents argue, they estimate themselves more competent than their interlocutors. That is the reason why the personal values have priority over the reputation values. In other words, the extended priority relation of the agent is constrained as follows:

$$\forall \text{ag}_j \in \mho_A \forall_{v_\omega} \in V_{\Omega_A} \forall v \in V_i - V_{\Omega_A} (v_j^i \ll_i^* v \ll_i^* v_\omega)$$

We can easily demonstrate that the extended priority relation is a strict complete ordering relation. An argument is acceptable by the argumentative agent ag_i if it is subjectively acceptable by AT_i* with respect to the extended set of arguments $\mathcal{A}(\mathcal{T}_i^*)$. $\mathcal{S}_i^*$ denotes the set of acceptable arguments for the argumentative agent 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 common *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 = speaker(M_k)) and addressed to a hearer (H_k = hearer(M_k)), i.e. one agent in the audience of the message. A_k = 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 can have an argumentative and public semantics (Bentahar et al., 2004). Because a commitment enrich the extended theory of the creditor, the speech acts can have a public semantics. Because a commitment could be justified by the extended theory of the debtor, the speech acts can have an argumentative semantics.

For example, the Figure 1 shows the 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.

In a similar way, the Figure 2 shows the 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.

The other speech acts (question(h), challenge(h), unknow(h), and withdraw(h)) are used to manage the sequence of moves (cf Section 6). They have no particular effects on the commitments stores, neither particular rational conditions of utterance. Since the speech act withdraw(h) has no effect on the

- 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$(A) = h$
- PUBLIC SEMANTICS: for any agent ag_k in the audience $\text{CS}_i^k(t) = \text{CS}_i^k(t-1) \cup \{h\}$

Figure 1. Semantics for assert an hypothesis h at time t.

- MESSAGE: $M_l = \langle \text{ag}_i, \text{ag}_j, \text{concede}(h) \rangle$
- ARGUMENTATIVE SEMANTICS: $\exists A \in \mathcal{A}(\mathcal{T}_i^\star)$ conclusion$(A) = h$ with (premise$(A) \neq \{h\} \wedge$ premise$(A) \not\subseteq \cup_{j \neq i} \text{CS}_j^i$)
- PUBLIC SEMANTICS: for any agent ag_k in the audience $\text{CS}_i^k(t) = \text{CS}_i^k(t-1) \cup \{h\}$

Figure 2. Semantics for concede an hypothesis h at time t.

commitments stores, we considered that the commitments stores are cumulative (Hamblin, 1970; Walton and Krabbe, 1995).

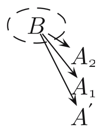
We can note that the rational conditions for the assertion of a hypothesis and for its explicit negation are not necessary mutually excluded. These nondeterministic situations make it possible for agents to choice. That is the reason why Parsons *et al.* (McBurney and Parsons, 2002) define a set of argumentative tactics. If an agent is *thoughtful.* she can assert any hypothesis for which she has an acceptable argument. Otherwise, she is *confident*. If an agent is *skeptical*, she can concede any hypothesis for which she has an acceptable argument. Otherwise, she is *credulous*. Because arguing agents exchange their convictions, they are thoughtful and credulous.

The hypothesis which are received must be valuated. For this purpose, the commitments will be individually considered in accordance with the estimated reputation of the agents from whom the information is obtained. The following example illustrates this principle.

EXAMPLE 2. If the argumentative agent ag_1 utters the following message: $M_1 = \langle \text{ag}_1, \text{ag}_2, \text{assert}(\, pres(bush) \leftarrow) \rangle$, then the extended theory of the argumentative agent ag_2 is as represented in the Table IV. The extended personal theory is composed of the personal theory 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:

Table IV. The extended theory of the second argumentative agent

$\ll_2^*$	V_2^*	$\mathcal{T}_2^*$
↑	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	$r_6: current\_pres(bush) \leftarrow$
	v_1^2	$\{pres(bush) \leftarrow\} = \text{CS}^2{}_1$



$V_{\Omega_A} = \{v_1, v_2\}$. 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 with respect to this argument but with respect to 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)* $\leftarrow$. Since she is thoughtful, she only asserts *pres(bush)* $\leftarrow$.

We have presented here a model of agents who exchange hypothesis and argue. In the next section, we bound a formal area where dialogues take place.

5. 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. Since we want to warrant that an agreement is reached at the end of the dialogues, we need to bound a formal area, called dialectics system, in which agents play and arbitrate. The dialectics system which is proposed in this section is similar to (Prakken, 2000). Moreover, we add a third agent who arbitrates in accordance with the estimated competence of the players and the advanced arguments.

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 common *moves language*. A move $\mathrm{move}_k = \langle M_k, R_k, P_k \rangle \in \mathcal{ML}_{\mho}$ has an identifier move_k. It contains a message M_k as defined before. $R_k = \mathrm{reply}(\mathrm{move}_k)$ is the identifier of the move to which move_k responds. A move (move_k) is either an initial move ($\mathrm{reply}(\mathrm{move}_k) = \mathrm{nil}$) or a replying move ($\mathrm{reply}(\mathrm{move}_k) \neq \mathrm{nil}$). $P_k = \mathrm{protocol}(\mathrm{move}_k)$ is the name of the protocol which is used.

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

DEFINITION 10. (Dialectics system)
Let $\mathrm{AT}_{\Omega_A} = \langle \mathcal{T}_{\Omega_A}, V_{\Omega_A}, \mathrm{promote}_{\Omega_A} \rangle$ be a common 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 $DS_{\Omega_M}(r_0, \mathrm{AT}_{\Omega_A}) = \langle N, \mathrm{wit}, H, T, \mathrm{convention}, Z, (u_p)_{p \in N} \rangle$ where:

- $N = \{\mathrm{init}, \mathrm{part}\} \subset \mho_A$ is a set of two argumentative agents called players: the initiator and the partner;
- $\mathrm{wit} \in \mho_A$ is a third argumentative agent, called witness, with a personal theory restricted to the common theory ($\mathcal{T}_{\mathrm{wit}} = \mathcal{T}_{\Omega_A}$);
- $\Omega_M \subseteq \mathcal{ML}_{\mho}$ 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 $T(h) = \text{init}$ else $T(h) = \text{part}$;
- 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;
- $u_{\text{init}}, u_{\text{part}} : Z \rightarrow \{-1, 1\}$ are the utility functions determining if a player is a winner.

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. We call *dialogue line* 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.

The witness attends to the debate, she is in the audience of the messages. The witness computes the final agreement. At the history h, the witness is associated to the extended theory $\text{AT}^*_{\text{wit}}(h) = \langle \mathcal{T}^*_{\text{wit}}(h), V^*_{\text{wit}}$, promote $^*_{\text{wit}}, \ll^*_{\text{wit}} \rangle$ where:

- the extended personal theory is composed of the common 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 common values and the reputation values of the two players: $V^*_{\text{wit}} = V_{\Omega_A} \cup \{v^{\text{wit}}_{\text{init}}, v^{\text{wit}}_{\text{part}}\}$.

$\mathcal{S}^*_{\text{wit}}(h)$ denotes the set of acceptable arguments for the witness which depends on the history and the reputation of players. The reputation relation of the witness corresponds to the global social order. 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 (or not) attribute the victory at the end of the dialogue to the player p which is (or not) convincing: $u_p(h) = 1$ (or $u_p(h) = -1$). In this way, the arbitrage of the witness depends on the arguments exchanged and the estimated competence of the players.

We can remark that the witness do not necessary know all the arguments of the players at the end of the dialogue. That is the reason why we introduce a fictive agent: the omniscient agent. She computes the theoretical possible agreement in the dialectics system. Whatever the history h is, the omniscient agent is associated to the theory $\mathrm{AT}_{\mathrm{omni}} = \langle \mathcal{T}_{\mathrm{omni}}, V_{\mathrm{omni}}, \mathrm{promote}_{\mathrm{omni}}, \ll_{\mathrm{omni}} \rangle$ where:

- the personal theory is composed of the personal theories of players: $\mathcal{T}_{\mathrm{omni}} = \mathcal{T}_{\mathrm{init}} \cup \mathcal{T}_{\mathrm{part}}$;
- the set of values is composed of the set of personal values and the reputation values of the two players: $V_{\mathrm{omni}} = V_{\Omega_A} \cup V^{\mathrm{omni}}_{\mathrm{init}} \cup V^{\mathrm{omni}}_{\mathrm{part}}$.

  $\mathcal{S}^*_{\mathrm{omni}}$ denotes the set of acceptable arguments for the omniscient agent which only depends on the reputation of the players. Her reputation relation is the same as the witness one. We said that r_0 is a *consensus* (written consensus(r_0)) if the witness is convinced by r_0.

Contrary to the omniscient agent, the witness does not necessary know all the arguments in the dialectics system. We can easily demonstrate the set of arguments are constrained as follows: $\mathcal{A}(\mathcal{T}^*_{\mathrm{wit}}(h)) \subseteq (\mathcal{A}(\mathcal{T}_{\mathrm{init}}) \cup \mathcal{A}(\mathcal{T}_{\mathrm{part}}) \subseteq \mathcal{A}(\mathcal{T}_{\mathrm{omni}})$ Therefore, the witness and the omniscient agent can have different convictions.

Since the goal is an agreement, the dialogue is sound if either the witness is convinced by the topic or she is convinced by its explicit negation. Since some arguments can be passed in silence or not known by any player, a dialogue will be complete if the omniscient agent and the witness have the same convictions.

DEFINITION 11. (Soundness and completeness of dialogue)
Let $h \in Z$ be a dialogue. Even if $\neg r_0$ replaces r_0:

h is *sound* iff provable $^h(r_0)$;
h is *complete* iff consensus(r_0) $\Rightarrow$ provable $^h(r_0)$.

We have bound here the area in which the dialogues take place. The quality of dialogues (soundness and completeness) depends on the dialogue protocol. In order to play persuasions, we formalize in the next section a dialogue protocol.

6. Persuasion

When a set of social and autonomous agents debate, they collaborate to confront their convictions. In this section we illustrate our dialectics system with the protocol of persuasion proposed in (Amgoud et al., 2002) where agents resolve their conflict by verbal means (Walton and Krabbe, 1995).

Contrary to (Amgoud et al., 2002), we formalize this protocol to demonstrate that persuasions are finites and leads to an agreement.

The persuasion protocol is an unique-respond protocol where players can reply just once to the other player's moves. The convention consists of the sequence rules represented in the Table V. Each rule specifies the authorized replying moves. For example, the rule of "Assertion/Refutation" (written $\mathrm{sr}_{A/R}$) specifies the authorized moves replying to the previous assertion (assert(H)). The speech acts resist or surrender to the previous one. Contrary to the resisting acts, the surrendering acts close the dialogue. A concession (concede(H)) surrenders to the previous assertion. A challenge (challenge(h)) and a refutation (assert($\neg h$)) resist to the previous assertion.

As previously said, the speech acts question(h), challenge(h), unknow(h), and withdraw(h) are used to manage the sequence of moves. On the one side, a question initiates the dialogue. On the other side, a plea of ignorance and a withdrawal close the dialogue. A challenge is a request for an argument.

We can note that the rational conditions of utterances for the allowed replying moves are not necessary mutually excluded. That is the reason why Parsons *et al.* (McBurney and Parsons, 2002) define a set of conventional tactics. If an agent is *open-minded*, she try to give her conviction, *i.e* a concession (concede(H)) or a refutation (assert($\neg h$)), in replying to the previous assertion. Otherwise, she is *argumentative*. In the same way, we tell that an agent is *cooperative* if she give her convictions, i.e. a confirmation (assert(h)) or an invalidation (assert($\neg h$)), in replying to a question. Otherwise, she is egoist. Since the debate is a collaborative process, agents are open-minded and cooperative.

The Figure 3 shows a persuasion game in the extensive form game representation where nodes are game situations and edges are associated with moves. For example, 2.3^{init} denotes a game situation where the exponent indicates that the initiator is the speaker of the next move. The exponent of game-over situations are boxes (*e.g.* $2.1^{\Box}$, $3.2^{\Box}$, and $4.2^{\Box}$). For evident clarity reasons, the game which follows the situation 2.2^{init}, 4.4^{init}, and 7.2^{part} are not

Table V. Set of speech acts and the potential answers

Sequences rules	Speech acts	Resisting replies	Surrendering replies
$\mathrm{sr}_{Q/A}$	Question(h)	Assert(h) assert($\neg h$)	Unknow(h)
$\mathrm{sr}_{A/R}$	Assert(H)	Challenge(h), $h \in H$ assert($\neg h$), $h \in H$	Concede(H)
$\mathrm{sr}_{C/A}$	Challenge(h)	Assert(H), $H \vdash_{\mho} h$	Withdraw(h)
sr_{T}	Unknow(H)	$\emptyset$	$\emptyset$
	Concede(H)	$\emptyset$	$\emptyset$
	Withdraw(H)	$\emptyset$	$\emptyset$

represented in the figure. In order to confront her conviction with the partner, an agent initiates a persuasion. If the partner has no arguments against/for the topic, she pleads ignorance and closes the dialogue (cf game situation $2.1^{\square}$). If the players have the same convictions, the witness is convinced and the dialogue closes (cf game situation $3.2^{\square}$). Otherwise, the goal of the dialogue is the resolution of the conflict by verbal means.

The termination of the persuasion can be warranted, whether the initial convictions of the players are.

THEOREM 1 *The persuasion* $(h \in Z)$ *is finite.*

Proof The game situations 2.2^{init} and 2.3^{init} are equivalent by symmetry on the content of the previous assertion. The game situation 2.3^{part} is equivalent to the game situation 4.4^{init} by recursion on the content of the previous assertion. The game situations 3.1^{part} and 5.1^{part} are equivalent by symmetry on the propositional content of the previous assertion. Moreover, the redundancy of information is forbidden in the assertions of the same dialogue line. Then, no loop will happen in dialogues. Because the variable domains are finite, the personal theories are finite. Therefore, $\mathcal{T}^*_{\text{part}}$ and $\mathcal{T}^*_{\text{init}}$ are also finite. Consequently, the recursion is finite and the dialogue closes.
Contrary to the termination, the soundness of the persuasion requires some particular conditions.

THEOREM 2 *If the initial convictions of players are (even if it inverts) such as*:

– *the initiator is convinced of* $\neg r_0 : \exists A \in \mathcal{S}^*_{\text{init}}(h_0)$ *conclusion*$(A) = \neg r_0$;
– *the partner is convinced of* $r_0 : \exists A \in \mathcal{S}^*_{\text{part}}(h_0)$ *conclusion*$(A) = r_0$.

Then the persuasion $(h \in Z)$ *is sound.*

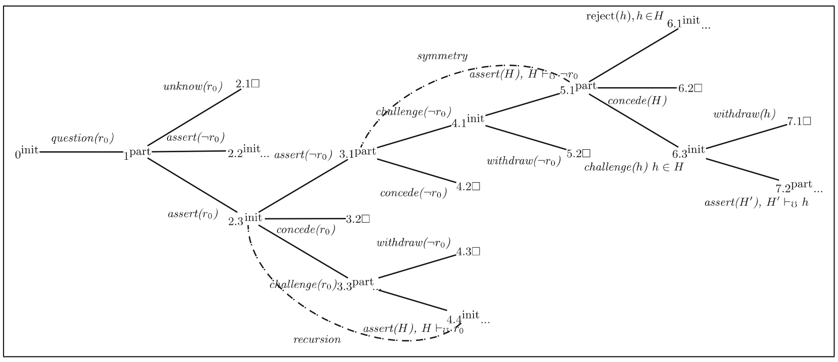


Figure 3. Persuasion in an extensive form game representation.

Proof The partner is convinced by r_0, cooperative and thoughtful. Therefore, the game situation 2.3^{init} is reached. The initiator is convinced by $\neg r_0$, thoughtful and open-minded. Therefore, the game situation 3.1^{part} is reached. Because the redundancy of information is forbidden, the game situation 4.1^{init} is reached. In the game-over situation $5.2^{\square}$, the witness has a trivial argument for r_0 and a trivial argument for $\neg r_0$: either init $\prec_{\text{wit}}$ part and the witness is convinced by r_0, or part $\prec_{\text{wit}}$ init and the witness is convinced by $\neg r_0$. In the game-over situation $6.2^{\square}$, the witness is convinced by $\neg r_0$. The other game-over situations are equivalent by symmetry on the content of the previous moves. Finally, the witness is either convinced by r_0 or $\neg r_0$. In other words, the persuasion is sound.
The following example illustrates such a dialogue protocol.

EXAMPLE 3. Let us consider the persuasion about the candidate *kerry*. The extended argumentative theories of the players are represented in the Figure 4. The commitments stores result from the sequence of moves. The arbitrage of the witness depends on the estimated competence of the players. Since the initiator is considered as more competent than the partner, the witness is convinced by *pres*(*bush*) $\leftarrow$ and the initiator wins.

Because the personal theories of players can be complementary, the omniscient agent can have an argument for/against the topic which neither the initiator nor the partner have. That is the reason why a persuasion between players with different convictions is not necessarily complete, whatever or not the protocol is unique-respond. Obviously, the persuasion is complete if the initiator is ignorant ($\mathcal{T}_{\text{init}} = \mathcal{T}_{\Omega_A}$).

We have formalize here a dialogue protocol to warrant that an agreement is reached.

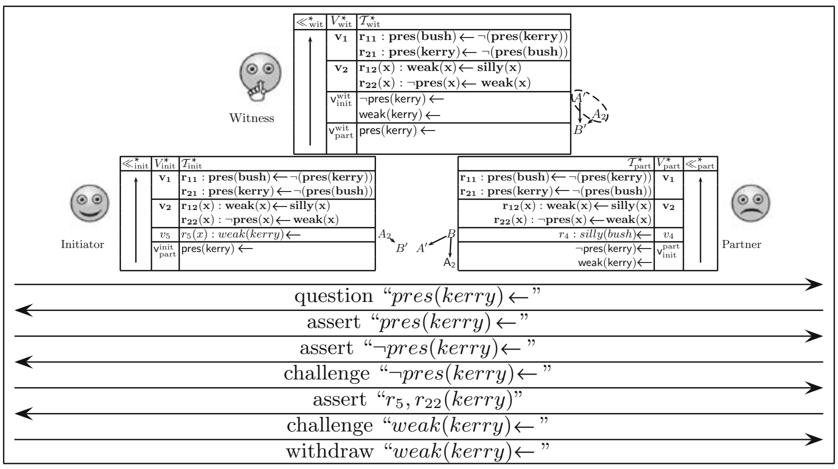


Figure 4. An example of persuasion.

7. Conclusions

We have proposed in this paper DIAL, a framework for inter-agents dialogue, which formalize a collective decision-making process to compose divergent interests and perspectives. This framework bounds a dialectics system in which argumentative agents play and arbitrate to reach an agreement. For this purpose, we have proposed an argumentation-based reasoning to manage the conflicts between arguments having different strengths for different agents. Moreover, we have proposed a model of argumentative agents which justify the hypothesis to which they commit and take into account the commitments of their interlocutors according to their reputations. In the scope of our dialectics system, a third agent is responsible of the final decision outcome which is taken by resolving the conflict between two players according to their competences and the advanced arguments.

In (Morge, 2004; Morge, 2005), we have proposed a multi-agents system based upon DIAL to support public decision in land planning. Each stakeholders is assisted by an argumentative agent representing him in automated dialogues. This intelligent system provide tools for the collaborative development of the argumentation schemes and to check the consistency or the inconsistency among users preferences allowing the conflicts and the consensus seeking. An implementation and an empirical assessment must come to valid the adequacy and the significance of our approach.

Notes

[1] For evident computational reasons, each variable has a finite domain value set.
[2] Due to the name of the authors: Amgoud, Maudet and Parsons

References

Amgoud, L. and Cayrol, C. (2002). A Reasoning Model based on the Production of Acceptable Arguments, Annals of Maths and AI 34(1–3): 197–215.

Amgoud, L., Maudet, N., and Parsons, S. (2002). An Argumentation-based Semantics for Agent Communication Languages. In *Proceedings of the 15th European Conference on Artificial Intelligence*, 38–42. IOS Press, Amsterdan.

Amgoud, L. and Parsons, S. (2001). Agent dialogues with conflicting preferences. In *Proceedings of the International Workshop on Agent Theories, Architectures and Languages*.

Bench-Capon, T.J.M. (2002). Value based Argumentation Frameworks. In *Proceedings of Non Monotonic Reasoning*, 444–453.

Bench-Capon, T.J.M. (2003). Persuasion in Practical Argument using Value based Argument Frameworks, Journal of Logic and Computation 13(3): 429–448.

Bentahar, J., Moulin, B., and Chaib-draa, B. (2004) Specifying and Implementing a Persuasion Dialogue Game using Commitments and Arguments. In *Proceedings of the workshop ArgMAS in AAMAS*, 130–148.

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

Callon, M. Lascoumes, P. and Barthe, Y. (2001). *Agir dans un monde in- certain*. Seuil.

Castelfranchi, C. and Falcone, R. (1998). Principles of Trust in Mas: Cognitive Anatomy, Social Importance, and Quantification. In *Proceedings of ICMAS'98*, 72–79.

Dignum, F., Dunin-Keplicz, B., and Verbrugge, R. (2001). Agent Theory for Team Formation by Dialogue. In *Agent Theories Architectures and Languages*, number LNAI 1986 in Intelligent Agents VII, 150–166, Springer-Verlag.

Dung, P.M. (1995). On the Acceptability of Arguments and its Fundamental Role in Non-monotonic Reasoning, Logic Programming and *n*-person Games, Artifical Intelligence 77(2): 321–357.

Hamblin, C.L. (1970). *Fallacies*. Methuen.

Kakas, A.C. and Moratis, P. (2002). Argumentative Agent Deliberation, Roles and Context. In *Electronic Notes in Theoretical Computer Science*, vol. 70. Elsevier.

Labrie, M., Chaib-draa, B., and Maudet, N. (2003). Diagal: A tool for analyzing and modelling commitment-based dialogues between agents. In *Proceedings of the 16th Canadian Conference on Artificial Intelligence*, vol. 2671 of *LNAI*, Halifax, June 2003, 353–369, Springer-Verlag.

Maudet, N. (2001). *Modéliser les Conventions des Interactions langagiéres: la Contribution des jeux de dialogues*. PhD thesis, Université Paul Sabatier, Toulouse, Mai.

McBurney, P. and Parsons, S. (2001). Intelligent Systems to Support Deliberative Democracy in Environmental Regulation, Information and Communications Technology Law 10(1): 33–43.

McBurney, P. and Parsons, S. (2002). Games that Agents Play: A Formal Frame-work for Dialogues between Autonomous Agents, Journal of Logic, Language and Information 11(3): 315–334.

Morge, M. (2004). Computer Supported Collaborative Argumentation. In *Proceedings of 4th workshop on Computational Models of Natural Argument (CNMA04)*.

Morge, M. (2005). Système dialectique multi-agents pour l'aide la concertation. PhD thesis, ENS Mines, Saint Etienne, Juin 2005.

Prakken, H. (2000). On Dialogue Systems with Speech Acts, Arguments, and Counterarguments. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'2000)*, number 1919 in Lecture Notes in AI, 224–238, Springer Verlag.

Schweimeier, M.S.R. (2002). Notions of Attack and Justified Arguments for Extended Logic Programs. In *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI02)*, 536–540, Amsterdam, IOS Press.

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

Singh, M. (2000). A Social Semantics for Agent Communication Languages. In *Proceedings of the IJCAI Workshop on Agent Communication Languages*. Springer-Verlag.

Walton, D. and Krabbe, E. (1995). *Commitment in Dialogue*. SUNY Press.