An Exploration in Using Cognitive Coherence Theory to Automate BDI Agents' Communicational Behavior

Philippe Pasquier, Andrillon Nicolas and Brahim Chaib-draa
Université Laval - DAMAS, Pavillon Pouliot
Département d'Informatique et de Génie Logiciel
Sainte-Foy G1K 7P4 (Québec, Canada)
{pasquier, andrillon, chaib}@iad.ift.ulaval.ca

ABSTRACT

The cognitive coherence theory for agent communication pragmatics allows modelling a great number of agent communication aspects while being computational. This paper describes our exploration in applying the cognitive coherence pragmatic theory for BDI agents communication. The presented practical framework rely on our dialogue games based agent communication language (DIAGAL) and our dialogue game simulator toolbox (DGS). It provides the necessary theoretical and practical elements for implementing the theory as a new layer over classical BDI agents. In doing so, it brought a general scheme for automatizing agents' communicational behavior. Finally, we give an example of the resulting system execution.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multi-agent systems*

Keywords

pragmatic, coherence, dissonance, dialogue utility

1. INTRODUCTION

Agents and multi-agents technologies allow the conception and development of complex applications. In the current distributed data processing paradigm, the fundamental characteristic of these systems is the agents skill in communicating with each other in a useful way regarding to their individual and collective goals. If numerous works have aimed to define agents communication languages, few have concentrated on their dynamic and automatic use by agents. This last task is left to the system designers, who analyse and specify manually the agent communicational behavior, usually by means of rules or by designing ad hoc protocols and static procedures to use them. In this paper, we introduce our investigation toward a theoretical and practical framework for the pragmatic of agent communication, i.e. the automation of agents' communicational behaviors.

In this paper, we first summarize our approach for agent communications pragmatic, the cognitive coherence theory (section 2). This conceptual framework is based on a unification of the cognitive dissonance theory which is one of main motivational theories in social psychology and Thagard's philosophy of mind theory: the coherence theory. After detailing our dialogue games based agent communication language (DIAGAL) (section 3), we briefly present our dialogue game simulator (DGS) (section 4), a practical framework to experience dialogue games. We indicate then, how our coherence pragmatic approach was implemented to automate conversations using DIAGAL games, among BDI agents (section 5). Finally, we give an example of automatic conversation between agents to illustrate our "complete" automatic communication framework (section 6).

2. DIALOGUE PRAGMATICS

2.1 The cognitive coherence framework

In cognitive sciences, cognitions gather all cognitive elements: perceptions, propositional attitudes such as beliefs, desires and intentions, feelings and emotional constituents as well as social commitments. From the set of all cognitions result attitudes which are positive or negative psychological dispositions towards a concrete or abstract object or behavior. All attitudes theories, also called cognitive coherence theories appeal to the concept of homeostasis, i.e. the human faculty to maintain or restore some physiological or psychological constants despite the outside environment variations. All these theories share as a premise the *coherence principle* which puts coherence as the main organizing mechanism: *the individual is more satisfied with coherence than with incoherence*. The individual forms an opened system whose purpose is to maintain coherence as much as possible (one also speaks about balance or about equilibrium). Attitude changes result from this principle in incoherence cases.

Our pragmatic theory follows from those principles by unifying and extending the cognitive dissonance theory, initially presented in 1957 by Festinger [11] with the coherence theory of the computational philosopher Thagard [29]. This last theory allows us to directly link the cognitive dissonance theory with notions, common in AI and MAS, of elements and constraints.

In our theory, elements are both private and public agent's cognitions: beliefs, desires, intentions and social commitments. Elements are divided in two sets: the set A of accepted elements (which are interpreted as true, activated or valid according to the elements type) and the set R of rejected elements (which are interpreted as false, inactivated or not valid according to the type of elements). Every non-explicitly accepted element is rejected. Two types of non-ordered binary constraints on these elements are inferred from the pre-existing relations that hold between them in the agent's cognitive model:

- *Positive constraints*: positive constraints are inferred from positive relations which can be: explanation relations, deduction relations, facilitation relations and all other positive associations considered.
- *Negative constraints*: negative constraints are inferred from negative relations: mutual exclusion, incompatibility, inconsistency and all the negative relations considered.

For each of these constraints a weight reflecting the importance and validity degree for the underlying relation is attributed. These constraints can be satisfied or not: a positive constraint is satisfied if and only if the two elements that it binds are both accepted or both rejected. On the contrary, a negative constraint is satisfied if and only if one of the two elements that it binds is accepted and the other one rejected. So, two elements are said to be *coherent* if they are connected by a relation to which a satisfied constraint corresponds. And conversely, two elements are said to be *incoherent* if and only if they are connected by a relation to which a non-satisfied constraint corresponds. Given a partition of elements among A and R, one can measure the *coherence degree* of a non-empty set of elements by adding the weights of constraints connected to this set (the constraints of which at least a pole is an element of the considered set) which are satisfied divided by the total number of concerned constraints. Symmetrically, the incoherence of a set of cognitions can be measured by adding the weights of non-satisfied constraints concerned with this set and dividing by the total number of concerned constraints.

In this frame, the basic hypothesis of the cognitive dissonance theory is that incoherence (what Festinger names dissonance [11]) produces for the agent a tension which incites him to change. The more intense the incoherence, the stronger are the insatisfaction and the motivation to reduce it. A cognition incoherence degree can be reduced by: (1) abolishing or reducing the importance of incoherent cognitions (2) adding or increasing the importance of coherent cognitions.

Festinger's second hypothesis is that in case of incoherence, the individual is not only going to change his cognitions or to try to change others's ones to try to reduce it, he is also going to avoid all the situations which risk increasing it. Those two hypotheses were verified by a large amount of cognitive and social psychology studies and experiences [34].

One of the major interests of the cognitive dissonance theory captured by our formulation is to supply incoherence measures, i.e. a metric for cognitive coherence. These measures match exactly the dissonance intensity measures first defined by Festinger.

2.2 Dialogue as coherence seeking

As we argue elsewhere [21, 22], using coherence as a motivational motor allows us to model a great number of expected features for dialogue pragmatic. In particular, it allows us to answer the following questions:

1. *Why agents should dialogue ?* Agents dialogue in order to reduce incoherences they cannot reduce alone. We distinguish internal (or personal) incoherence from external (or collective) incoherence depending on whose elements are involved in the incoherence[1].
2. *When should an agent take a dialogue initiative, on which subject and with whom ?* An agent engages in a dialogue when an incoherence magnitude exceeds a fixed level[2] and he cannot reduce it alone. Whether because it is an external incoherence and he cannot accept or reject external cognitions on his own, or because it is an internal incoherence he fails to reduce alone. The subject of this dialogue should thus focus on the elements which constitute the incoherence. The dialogue partners are the other agents involved in the incoherence if it is an external one or an agent he thinks could help him in the case of a merely internal incoherence.
3. *By which type of dialogue ?* Even if we gave a general mapping of incoherence types toward dialogue types [22], the theory is generic enough for being applied to any conventional communicational framework. Hereafter (section 5), we gave the procedural scheme for this choice using DIAGAL dialogue games as primitive dialogue types.
4. *How to define and measure the utility of a conversation ?* As we state in [21], following the coherence principle and the classical definition of utility functions, the utility of a dialogue is the difference between the incoherence before and after this dialogue. Furthermore, we define the expected utility of a dialogue as the incoherence reduction in case of success of the dialogue, i.e. the expected dialogue results are reached. As dialogues are attempts to reduce incoherence, expected utility is used to choose between different competing dialogues types (dialogue games in our case).
5. *When to stop dialogue or else, how to pursue it ?* The dialogue stops when the incoherence is reduced or else either it continues with a structuration according to the incoherence reductions chain or it stops because things cannot be re-discussed anymore (this case where incoherence persists often leads to attitude change as described in section 5).
6. *What are the impacts of the dialogue on agents' private cognitions ?* In cases where dialogue, considered as an attempt to reduce an incoherence by working on the external world, definitively fails, the agent reduces the incoherence by changing his attitudes in order to recover coherence (this is the attitude change process described in section 5).
7. *Which intensity to give to illocutionary forces of dialogue acts ?* Evidently, the intensities of the illocutionary forces of dialogue/speech acts generated are influenced[3] by the incoherence magnitude. The more important the incoherence magnitude is, the more intense the illocutionary forces are.
8. *What are the impacts of the dialogue on agents' mood ?* The general scheme is that: following the coherence principle, coherence is a source of satisfaction and incoherence is a source of dissatisfaction. We decline emotional attitudes

[1] In the presented system, external elements are social commitments.

[2] This level or a "Should I dialogue ?" function allows us to model different strategies of dialogue initiative.

[3] Actually, this is not the only factor, as we exemplify elsewhere, other factors could also matter: social role, hierarchical positions,...

from internal coherence dynamic (happiness arises from successful reduction, sadness from failed attempt of reduction, fear from a future important reduction attempt, stress and anxiety from an incoherence persistence,...).

9. *What are the consequences of the dialogue on social relations between agents ?* Since agents can compute and store dialogue utility, they can build and modify their relations with others agents in regard to their past dialogues. For example, they can strengthen relations with agents with whom past dialogues were efficient and useful, according to their utility measures, ...

All those dimensions of our theory - except 7, 8 and 9 - will be exemplified in section 6. But before implementing our pragmatic theory we need an agent communication language.

3. A DIALOGUE GAME LANGUAGE BASED ON COMMITMENTS: DIAGAL

DIAGAL[DIAlogue Games Agent Language] is our commitment-based agent language in which we define semantics of the communicative acts in terms of public notions, e.g. social commitments [6]. The use of those public cognitions allows us to overcome classical difficulties of "intentional" agent communication approach: the sincerity hypothesis does not hold anymore and the semantic verification problem is solved (see [23] for explanations).

3.1 Social commitments

As our approach is based on commitments, we start with some details about the notion of commitment. The notion of commitment is a social one, and should not be confused with the notion of individual commitment used to emphasize individual intention persistance. Conceptually, social commitments model the obligations agents contract toward one another. Crucially, commitments are oriented responsibilities contracted towards a partner or a group. In the line of [33], we distinguish action commitments from propositional commitments.

Commitments are expressed as predicates with an arity of 6. An accepted action commitment thus take the form:

$$C(x, y, \alpha, t, s_x, s_y)$$

meaning that x is committed towards y to α at time t, under the sanctions s_x and s_y. The first sanction specifies conditions under which x reneges its commitment, and the second specifies conditions under which y can withdraw from the considered commitment. Those sanctions[4] could be social sanctions (trust, reputation,...) as well as material sanctions (economical sanctions, repairing actions, ...). An accepted propositional commitment would be have propositional content p instead α. Rejected commitments take the form $\neg C(x, y, \alpha, t, s_x, s_y)$ meaning that x is not committed toward y to α

This notation for commitments is inspired from [27], and allows us to compose the actions or propositions involved in the commitments: $\alpha_1|\alpha_2$ classically stands for the choice, and $\alpha_1 \Rightarrow \alpha_2$ for the conditional statement that α_2 will occur in case of the occurrence of the event α_1. Finally, the operations on the commitments are just creation and cancellation.

Now, we need to describe the mechanism by which the commitments are discussed and created during the dialogue. This mechanism is precisely modelled within our game structure. To account for the fact that some commitments are established within the contexts of some games *and only make sense within this context* [16, 19], we make explicit the fact that those *dialogical commitments* are particular to game g (by indicating g as a subscript). This will typically be the case of the dialogue rules involved in the games, as we will see below.

[4]Since we did not investigate a whole agent architecture in this paper, we leave sanctions as a realistic conceptual abstraction.

3.2 Game Structure

We share with others [7, 12, 19] the view of dialogue games as structures regulating the mechanism under which some commitments are discussed through the dialogue. Unlike [7, 19] however, we adopt a strict commitment-based approach within game structure and express the dialogue rules in terms of dialogical commitments. Unlike [12] on the other hand, we consider different ways to combine the structures of the games.

In our approach, games are considered as bilateral structures defined by:

- *entry conditions*, (E): conditions which must be fulfilled at the beginning of the game, possibly by some accommodation mechanism;
- *success conditions*, (S): conditions defining the goal of the initiator participant when engaged in the game;
- *failure conditions*, (F): conditions under which the initiator can consider that the game reached a state of failure;
- *dialogue rules*, (R): rules specifying what the conversing agents are "dialogically" committed to do.

As previously explained, all these notions, even dialogue rules, are defined in terms of (possibly conditional, possibly dialogical) commitments. Within games, conversational actions are time-stamped as "turns" (t_0 being the first turn of dialogue within this game, t_f the last).

3.3 Grounding and composing the games

The specific question of how games are grounded through the dialogue is certainly one of the most delicate [17]. Following [25], we assume that the agents can use some meta-acts of dialogue to handle games structure and thus propose to enter in a game, propose to quit the game, and so on. Games can have different status: they can be *open*, *closed*, or simply *proposed*. How this status is discussed in practice is described in a *contextualization* game which regulates this meta-level communication. Figure 1 indicates the current contextualisation moves and their effects in terms of commitments. For example, when a proposition to enter a game j ($prop.in(x, y, j)$) is played by the agent x, y is committed to accept ($acc.in$), to refuse ($acc.in$) or to propose entering another game j' ($prop.in(y, x, j')$), which would lead to a presequencing type of dialogue games structuration.

Move	Operations
$prop.in(x, y, j)$	$create(y, C_j(y, x, acc.in(y, x, j)$ $\|ref.in(y, x, j)\|prop.in(y, x, j')))$
$prop.out(x, y, j)$	$create(y, C_j(y, x, acc.out(y, x, j)$ $\|ref.out(y, x, j)))$
$acc.in(x, y, j)$	create dialogical commitments for game j
$acc.out(x, y, j)$	suppress dialogical commitments for game j
$ref.in(x, y, j)$	no effect on the public layer
$ref.out(x, y, j)$	no effect on the public layer

Figure 1: DIAGAL contextualisation game.

Concerning the possibility to combine the games, the seminal work of [33] and the follow-up formalisation of [25] have focused on the classical notions of *embedding* and *sequencing*. Even if, recent works, including ours, extend this to other combinations [19, 6], in our present simulation framework, we only consider the three games' compositions allowed by the previous contextualisation game.

- *Sequencing* noted $g_1; g_2$, which means that g_2 is proposed after the termination of g_1.
- *Pre-sequencing* noted $g_2 \rightsquigarrow g_1$, which means that g_2 is opened while g_1 is proposed. Pre-sequencing is used to establish, to enable some of g_1 entry conditions or to explicite some information prior to the entrance in g_1.
- *Embedding* noted $g_1 < g_2$, which means that g_1 is opened while g_2 was already opened.

A game stack captures that commitments of the embedded games are considered as having priority over those of the embedding game.

3.4 Basic games

Up to now we have introduced four basic building dialogue games, which are exactly those which lead (in case of success) to the four types of commitments which can hold between two agents X and Y, namely:

1. for an attempt to have an action commitment from Y toward X accepted, agent X can use a "request" game (rg);
2. for an attempt to have an action commitment from X toward Y accepted, agent X can use an "offer" game (og);
3. for an attempt to have a propositional commitment from X toward Y accepted, agent X can use an "inform" game (ig);
4. for an attempt to have a propositional commitment from Y toward X accepted, agent X can use an "ask" game (ag).

Next subsections detail those four games. Sanctions were omitted in our games specifications just for better readability.

3.4.1 Request game (rg)

This game captures the idea that the initiator (x) "request" the partner (y) for an action α and this latter can "accept" or "reject". The conditions and rules are:

$$
\begin{array}{l|l}
E_{rg} & \neg C(y,x,\alpha,t_0) \text{ and } \neg C(y,x,\neg\alpha,t_0) \\
S_{rg} & C(y,x,\alpha,t_f) \\
F_{rg} & \neg C(y,x,\alpha,t_f) \\
R_{rg} & C_g(x,y,\mathit{request}(x,y,\alpha),t_0) \\
 & C_g(y,x,\mathit{request}(x,y,\alpha) \Rightarrow \\
 & C_g(y,x,\mathit{accept}(y,x,\alpha)|\mathit{refuse}(y,x,\alpha),t_1),t_0) \\
 & C_g(y,x,\mathit{accept}(y,x,\alpha) \Rightarrow C(y,x,\alpha,t_2),t_0) \\
 & C_g(y,x,\mathit{refuse}(y,x,\alpha) \Rightarrow \neg C(y,x,\alpha,t_2),t_0)
\end{array}
$$

Figure 2: Conditions and rules for the request game.

3.4.2 Offer game (og)

An offer is a promise that is conditional upon the partner's acceptance. To make an offer is to put something forward for another's choice (of acceptance or refusal). To offer then, is to perform a conditional commissive. Precisely, to offer α is to perform a commissive under the condition that the partner accept α. Conditions and rules are in this case:

$$
\begin{array}{l|l}
E_{og} & \neg C(x,y,\alpha,t_0) \text{ and } \neg C(x,y,\neg\alpha,t_0) \\
S_{og} & C(x,y,\alpha,t_f) \\
F_{og} & \neg C(x,y,\alpha,t_f) \\
R_{og} & C_g(x,y,\mathit{offer}(x,y,\alpha),t_0) \\
 & C_g(y,x,\mathit{offer}(x,y,\alpha) \Rightarrow \\
 & C_g(y,x,\mathit{accept}(y,x,\alpha)|\mathit{refuse}(y,x,\alpha),t_1),t_0) \\
 & C_g(x,y,\mathit{accept}(y,x,\alpha) \Rightarrow C(x,y,\alpha,t_2),t_0) \\
 & C_g(x,y,\mathit{refuse}(y,x,\alpha) \Rightarrow \neg C(x,y,\alpha,t_2),t_0)
\end{array}
$$

Figure 3: Conditions and Rules for the offer game.

3.4.3 Inform game (ig)

Notice that a human partner can be disposed to be in accord or agreement with someone without uttering any word. He can also agree by doing an explicit speech act. In this case - required for agents since they do not support implicit communication - the partner can agree or disagree. The conditions and rules for this couple is the following:

$$
\begin{array}{l|l}
E_{ig} & \neg C(x,y,p,t_0) \text{ and } \neg C(x,y,\neg p,t_0) \\
S_{ig} & C(x,y,p,t_f) \\
F_{ig} & \neg C(x,y,p,t_f) \\
R_{ig} & C_g(x,y,\mathit{assert}(x,y,p),t_0) \\
 & C_g(y,x,\mathit{assert}(x,y,p) \Rightarrow \\
 & C_g(y,x,\mathit{agree}(y,x,p)|\mathit{disagree}(y,x,p),t_1),t_0) \\
 & C_g(x,y,\mathit{agree}(y,x,p) \Rightarrow C(x,y,p,t_1),t_0) \\
 & C_g(y,x,\mathit{disagree}(y,x,p) \Rightarrow \neg C(x,y,p,t_2),t_0)
\end{array}
$$

Figure 4: Conditions and rules for the inform game.

3.4.4 Ask game (ag)

We use "ask" in the sense of asking a closed question, which consists of requesting the partner to agree or disagree with a proposition p. According to these remarks, we propose the following structure for the *ask* game:

$$
\begin{array}{l|l}
E_{ag} & \neg C(y,x,p,t_f) \text{ and } \neg C(y,x,\neg p,t_f) \\
S_{ag} & C(y,x,p,t_f) \\
F_{ag} & \neg C(y,x,p,t_f) \\
R_{ag} & C_g(x,y,\mathit{question}(x,y,p),t_0) \\
 & C_g(y,x,\mathit{question}(x,y,p) \Rightarrow \\
 & C_g(y,\mathit{agree}(y,x,p)|\mathit{disagree}(y,x,p),t_1),t_0) \\
 & C_g(y,x,\mathit{agree}(y,x,p) \Rightarrow C(y,x,p,t_2),t_0) \\
 & C_g(y,x,\mathit{disagree}(y,x,p) \Rightarrow \neg C(y,x,p,t_2),t_0)
\end{array}
$$

Figure 5: Conditions and rules for the ask game.

Notice that in those games, the included speech acts are labelled with a relative integer (not shown on the Figures) indicating the illocutionary force intensity degree relatively to the default basic illocutionary force degree. For example, in the request game the request stand for the directive category for action which is mapped to: suggest: -2, direct: -1, request: 0, demand: 1, order: 2. Allowing agents to use the appropriate illocutionary forces intensity degree for each dialogue/speech act leads to many variations of those basic games.

4. THE DIALOGUE GAME SIMULATOR

We have developed a toolbox, the dialogue game simulator, in order to simulate and visualize games-based dialogue as presented in the previous section while allowing the integration of some future concepts. The dialogue games simulator (DGS) aims to be an effective tool for games testing and validation as well as a mean of exploring different agent architectures concerning dialogue pragmatics. DGS main interface allows managing connected agents, loading dialogue games and visualizing synthetic dialogue diagrams. DGS was developed in JAVA using JACK^{TM}agent technology [13]. In this section, we briefly present the various components of DGS.

4.1 Game files

As mentioned previously, a game is composed of entry conditions, success conditions, failure conditions and rules. In DGS, each of these game components is defined in its own file, adding to the possible information re-use while facilitating the maintainability of the files. All those files are written in XML. Using XML has the advantage of being easily manageable in liaison with JAVA while offering a good way of describing information. The DTD (Document Type Definition), associated with XML files, describes the precise way in which the game designer must create his files. That gives designers and users a mean of knowing if a game conforms to the specifications and if it is manageable by the simulator.

The games are loaded when the simulator starts and are placed in a list where agents can charge them when connecting.

4.2 Agenda and dialogue manager

The *agenda* and *dialogue manager* are the principal tools provided by DGS. Those tools should be included/embedded in all agents who aim to use loaded DIAGAL Dialogue Games. The agenda is a kind of individual "commitment store" where commitments are classified according to the time they were contracted. This structure contains commitments in action and propositional commitments that hold as well as dialogical commitments in action deduced from the current dialogue game(s) rules. Each agent has his own agenda which does not contain commitments of all agents which are connected to the simulator, but only those for which he is debtor or creditor.

The agenda is managed by the agent's dialogue manager module which adds or removes commitments according to current dialogue games rules and external events. A commitment in action is fulfilled when an action (perceived as an external event) that corresponds exactly to its description occurs. The dialogue manager also checks that every agent's operations conforms to the current contextualisation and opened dialogue games.

4.3 Action Board and Game Stack

The *action board* stores the actions which were played during simulation. It is modelled as an UML sequence diagram. Each workspace has its own action board where users can observe the exchanges of messages between agents as well as the time which is attached to these actions. It is represented as a history of the actions carried out relating to each initiated dialogue. The action board aims to help the simulator user understand and analyze what occurred in a dialogue between two agents.

The *game stack* is a common structure used by dialogue managers of conversing agents to keep track of the embedded games during a conversation. Each time a new game is opened, it is placed on the top of the stack inside the related workspace and it becomes the current game of this workspace. The stack makes it possible to know which game will become active when the top one will be closed and withdrawn from the stack. This stack is also used to manage the priority between the games: the top element having more priority over the bottom element.

4.4 Dialogue Workspace

The *dialogue workspace* is an environment which contains all the data which are specific to a dialogue between two agents: games stack, actions board and some information about hierarchical relations between conversing agents.

In Figure 6, we present a simplified overview of the DGS framework. This figure presents two agents interacting trough a dialogue workspace. They communicate by sending each other messages (communicative actions) and as such messages are produced, the simulator places them into the actions board. In accordance with the current game on the game stack, the dialogue managers of the sender and receiver agents deduce the appropriate commitments from the game files and places them into their agendas.

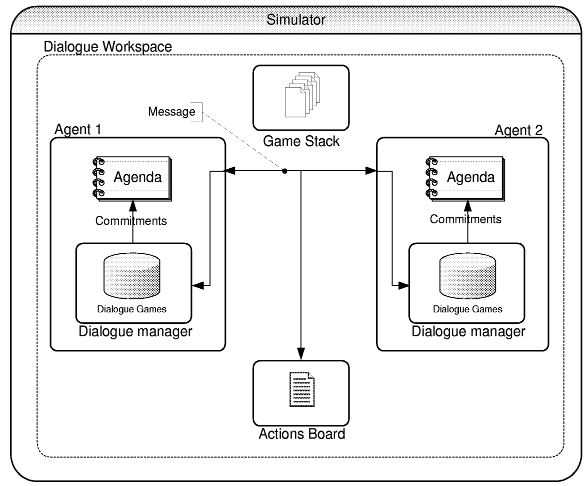

Figure 6: Simulator overview.

In its current form, DGS allows simulating conversations between pairs of software agents (three agents resulting in three pairs). The next section focuses on our first attempt to implement the coherence theory for automatizing dialogues between BDI agents. Those dialogues would take place in the DGS framework using precisely DIAGAL dialogue games presented in the previous sections.

5. INTEGRATING COHERENCE THEORY TO BDI AGENTS

5.1 Linking private and social cognitions

In this section, we describe our first attempt to the complex task of integrating the concepts of the coherence theory in BDI agents practical reasoning. More precisely, we implemented our coherence pragmatics as a new layer above the existing BDI architectures. Since we do not propose a whole coherentist approach for agent modelling, we will have to extend the classical BDI framework so that it can fit with our approach. In particular, traditional BDI frameworks do not involve social commitments treatments.

Choosing a conventional approach for agent communication leads us to extend the intentional paradigm for agent practical reasoning issued from rational interaction theories: *cognitive agent should not reason solely about his and others intentions, he should also reason about potential and already existing social commitments* (coming from held dialogues or system's conventions). In order to use our pragmatic theory to automatize the communication level of the traditional BDI abstract architecture, we need to connect private cognitions (mental states) with public ones (social commitments).

Prior to those links, we assume that our intentional layer is filtered from the BDI agent's whole intentions set. We assume that the intention we receive are either *social individual intentions* or failed individual intentions[5]. Social individual intentions are intentions concerning goals which require social aspects to be worked on. For example, an employee which has an intention about something his boss would be responsible for would have to make some social commitments socially accepted before achieving it. More generally, any intention that is embedded in a somewhat collective activity would have to be a social individual intention except if it is part of an already socially accepted collective plan. Those social intentions are intentions about a (even indirectly) collective state of affairs indicating that those intentions will be part of an external incoherence. Finally, individual intentions concerning goals which do not match any individual plan or which associated plan failed could be included in this layer (this matches the case where the agent faces an internal incoherence he cannot reduce alone). This phase of identifying intentions which could have a social impact appears to be crucial for integrating conventional approaches to existing cognitive agent architectures.

In this context, we can return to the general question: what are the links between social commitments and private mental states ? As a first answer, we propose linking private and public cognitions as follows[6]:

- According to the classic practical reasoning scheme, private cognitions finally end in intentions through deliberation and we make the usual distinction between *intention to* (do something or make someone doing something) and *intention that* (a proposition holds) [4];
- Regarding public cognitions, we distinguish *commitments in action* from *propositional commitments* [33];
- An accepted commitment is the socially accepted counterpart of an intention, commitments in action are the counterparts of "intentions to" and propositional commitments are the counterparts of "intentions that".

Those relations are not completely new since many authors have already considered individual intentions as a special kind of individual commitment [4, 32]. Our links extend this to reach the social level in the appropriate cases (social individual intentions or failed individual intentions). Constraints between the intentional private layer and the social commitments layer would be inferred from those links as well as any other logical links between intentions and social commitments.

5.2 BDI formulation of the attitude change process

In our model, *any agent tries to maximize his coherence*, i.e. tries to reduce his incoherences beginning with the most intense one. To reduce an incoherence, the agent has to accept or reject cognitions to better satisfy the constraints which connect them. These cognitions can be private or public. To be able to integrate communication into our model, it is now necessary to introduce the fundamental link which exists between our formulation of the cognitive dissonance theory and the notion of resistance to change.

All the cognitions are not equally modifiable. This is what Festinger names the resistance to change of cognitions. The resistance to change of a cognition is a function of the number and the importance of the elements with which it is coherent, also depending on its type, age, as well as the way by which it was acquired: perception, reasoning or communication. Social commitments are particular cognitions which are not individually modifiable but must be socially established and dialogue games are tools for attempting to establish collectively accepted commitments. That is, in order to get a social commitment accepted, an agent has to have a dialogue. Dialogues are the only means for agents to try to establish social commitments coherent with their private cognitions. However, after those dialogues, some commitments can remain incoherent with private intentions.

After any dialogue game, the discussed commitment is either accepted or rejected. As we saw before, an accepted commitment is not modifiable anymore without facing the associated sanctions. And we assume that a discussed commitment which is still rejected will gain in resistance to change. The point here is that an agent could not make attempts to have the desired commitment accepted indefinitely.

This resistance to change and associated sanctions would partially forbid the agent to gain coherence by changing the commitment acceptance state. We could simplify by saying that the discussed commitments usually stand for social obligations and fix one of the poles of the constraints which are connected to them. To reduce possible incoherence while conforming to discussed commitments, agents should then change their private cognitions to restore the coherence. This is the spring of the *attitude change* in our system and it formalizes the vision of the psychologists Brehm and Cohen on this subject [5], supported by a great number of experiments.

In the present simplified framework, the only private cognitions we consider are the intentions, but we assume that the underlying BDI layer would spread the attitude change among all the private cognitions. An example of this attitude change mechanism is supplied in section 6.

In MAS, knowing when an agent should try to modify the environment (the public social commitments layer, among others) to satisfy his intentions, and when the agent has to modify his mental states to be coherent with his environment is a crucial question. In practical reasoning, this question take the form: when an agent should reconsider his intention and deliberate again and when should he persist in acting in the previous deliberated way ? As we have just seen, within our approach, agents face the same problem and different strategies toward the modification of already discussed commitments (including reasoning about sanctions and resistance to change in order to know if the agent should persist or not) would lead to different individual commitment types in a way analogous with that of Rao and Georgeff [24]. The main difference is that this choice, like others, would be dynamically based on expected utility, i.e. expected coherence gain.

In Figure 7, we sum up (hiding the quantitative level of calculus) the means by which we link intentions, social commitments and

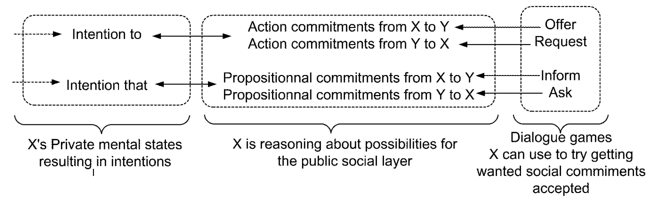

Figure 7: Links between private cognitions, public cognitions and DIAGAL dialogue games.

[5]With the "individual" qualifier in both, we mean that we do not refer to notions of we-intention or collective intentions such as those developed by Searle [26] or Tuomela [31]. Here, intentions are classical private intentions.

[6]Although, we give a first account here, much more work should be done on this point.

DIAGAL dialogue games. From the left to right we have two types of intentions linked with the four possible corresponding commitments types (the four ones seen in section 3.4). Notice that until they have been really discussed, those commitments are only potential commitments generated by the agent to reason with. To cohere with one of its accepted intentions, an agent will usually (according to the expected utility calculus) consider trying to get the corresponding commitment accepted. To make such an attempt, the agent will choose a DIAGAL dialogue game whose success condition unify with the wanted commitment.

5.3 The expected utility function

As we have seen it in section 2.1, the whole agent cognitive coherence is expressed as the sum of weights of satisfied constraints divided by the sum of weights of all constraints[7]. At each step of his reasoning, an agent will search for a cognition acceptance state change which maximizes the coherence increase, taking into account the resistance to change of that cognition (technically a 1-optimal move). If this attitude is a commitment, the agent will attempt to change it through dialogue and if it is an intention, it will be changed through attitude change. In that last case, we call the underlying architecture of the agents to spread the attitude change and re-deliberate.

In our implementation, an agent determines which is the most useful cognition acceptance state change by exploring all states reachable from its current state and select the cognition which can *in case of a successful change* be the most useful to change. A state is said to be reachable if it can be obtained from the current state by modifying only one cognition. Since all cognition cannot be equally modified, we introduced a notion of cost to take into account resistance to change or sanctions associated to cognitions. All explored states are so evaluated through an *expected utility function* expressed as below:

$$\begin{aligned} g(exploredState) = \quad & coherence(exploredState) \\ & -cost(cognitionChanged) \end{aligned}$$

where $exploredState$ is the evaluated state, $cognitionChanged$ is the cognition we are examining the change, and $cost$ is a cost function expressed as:

1. if $cognitionChanged$ is an intention, its cost of change equals its resistance to change;
2. if $cognitionChanged$ is an rejected commitment, its cost of change equals its resistance to change (which is initially low but which is increased at each unfruitful attempt to establish it);
3. if $cognitionChanged$ is an accepted commitment, its cost of change is provided by its associate sanction.

5.4 The treatment algorithm

Our agents behavior is guided by their coherence and their social commitments. At each step of the simulation, our agents consult their agendas and behave in order to fulfill the commitments which have been deduced from previous actions of agents and rules of dialogue games. When agents must determine the actions they have to produce, they apply the following algorithm:

```
1:  Procedure CommunicationPragmatics()
2:  List commitments=agenda.getCommitments();
3:  List dialogCommitments=
4:  agenda.getDialogCommitments();
5:  treatCommitments();
6:  if dialogCommitments.isEmpty() then
7:     initiateDialogue();
8:  else
9:     treatDialogCommitments();
10: end if
```

As we have seen in section 3.1, we distinguish two types of commitments: the dialogical ones and the extra-dialogical ones. The procedure for treating the extra-dialogical commitments (line 5) consists in updating the cognitive model of the agent by browsing extra-dialogical commitments in the agenda and operate as follows. (1) Each time an accepted commitment is encountered, the corresponding commitment in the agent's cognitive model is marked as accepted. If the corresponding intention in the cognitive model of the agent is rejected, then the agent call the underlying BDI architecture for a possible attitude change process. (2) Each time a rejected commitment is encountered, the resistance to change of the corresponding potential commitment in his cognitive model is increased, so that after eventually several unsuccessful attempts, this commitment will be so expensive to establish that it will not constitute an useful change of cognition[8]. This last case would lead to attitude change. This operation is performed before treating the dialogical commitments in order that as soon as a commitment is established, it is taken into account in the rest of the dialogue.

The procedure of initiating a dialogue (line 7) consists in searching the most useful cognition to change[9]. If it is a commitment, the agent initiates a dialogue with the appropriate dialogue game, or begins an attitude change process if it is an intention. The choice of the appropriate dialogue game is made by unifying the commitment the agent wants to establish with the conditions of success of the games loaded in the simulator.

Treating dialogical commitments (line 9) consists in exploring all the possible actions that are determined by dialogue games and selecting the one which has the best consequences on coherence. If the extra-dialogical commitment which is concerned by the current game is not the most useful change for the agent, it will embed a game by proposing the entrance in a new, subjectively more appropriate, dialogue game.

Notice that coordination of dialogue turns is ensured by the dialogue games rules and the resulting extra-dialogical commitments order in the agents' agendas. Finally, this algorithm is called each time:

- the underlying BDI architecture finish a deliberation process (or a re-deliberation process after a positive reconsider() call initiated by our algorithm as a possible attitude change process). We assume that the produced intentions are either social individual intentions or individual intentions that the agent could not realize alone.
- the agent have something in his agenda. This ensures, that the agent re-execute this algorithm until all dialogs are closed

[7] Notice that the general coherence problem: give the elements partition between A and R that maximize coherence is NP-complete. A formal demonstration could be found in [30].

[8] Notice that following Rao and Georgeff vocabulary [24] the amount of the increase in resistance to change will lead to the different commitment strategies: if this increase in the resistance to change is null the agent will be blindly committed in trying to get this social commitment accepted, if the increase is drastically important this individual commitment will be an open-minded one and in between, we would get a wild range of single minded commitment strategies. Notice that those commitment strategies could dynamically depend on: the incoherence magnitude, the dialogue topic, the partner,…

[9] There could be none, for example if the coherence is already maximal.

and that the agent will treat dialogue initiated by others. For example, when the agent receive a $prop.in$ message for entering a particular dialogue game, the corresponding dialogical commitment given by the contextualisation game his added to his agenda. Notice that, we assume as a first simplification that the agent is dialogically cooperative and that he systematically accept entering the game (in the treatDialogCommitments() procedure).

Finally, we implement JACK™ BDI[10] agents using this pragmatic framework to manipulate DIAGAL dialogue games within the DGS.

6. EXAMPLE

Let's assume that we have two agents, Paul and Peter, who have agreed on a common plan to go to the concert of their favorite band and split the whole bill. A subtask of this plan is to go to buy the tickets at the store. Paul has been assigned this task and is now about to deliberate about the way he will go to the store. He has to choose between two mutually exclusive intentions: the one of taking a cab and the one of going by foot. We assume that Paul's underlying BDI architecture has accepted the first one and rejected the second one (perhaps in order to save time). As they will split the bill (and that taking a cab costs money), Peter would rather that Paul went by foot. Thus, he has the rejected intention that Paul takes a cab and the accepted one that Paul goes by foot.

Both intentions may be associated with two corresponding potential commitments (according to links established in section 5.1): the social commitment from Paul toward Peter to take a cab and the social commitment from Paul toward Peter to go by foot. In addition, the commitment to take a cab and the intention of taking a walk are incompatible, as well as the commitment of taking a walk and the intention of taking a cab. From this initial state, according to our model, a positive constraint between intention and pending commitment is induced from the correspondance relation and negative constraints are induced from the the mutually exclusive relation and the incompatibility relations. Figure 8 presents the network of intentions of both Paul (on the left side) and Peter (on the right) as well as the pending rejected commitments. Notice that the commitments represented are potential commitments used by agents to reason. At this stage, they are not real social commitments since they have not been established by dialogue. In this example, a weight of 1 has been affected to all constraints as a simplification[11].

In DGS, we can decide which agent has the acting initiative, thus determining on whom incoherence dialogue will be taken. We will assume that Paul has the initiative. Initially, as shown by Figure 8, Paul has three satisfied constraints (number 1, 3 and 4) in an amount of five constraints so it has a coherence of 0.6. Paul will therefore try to increase it by localizing the most useful cognition to change. The Figure 9 shows the different states that can be reached by Paul from its initial situation. Below each is indicated the coherence c obtained in this state as well as the value of the expected utility function g. According to those results, Paul will make an attempt to get the commitment $C(Paul, Peter, take\_a\_Cab)$ accepted. Since it is a social commitment, Paul will use one of the dialogue games which are tools to attempt establishing commitments. Since this commitment is a commitment toward Peter, Peter will be the dialogue partner. Paul will then choose between the available dialogue games which success condition unify with the desired commitment. The only DIAGAL dialogue game which have a success condition of the form $C(initiator, partner, action)$ is the offer game.

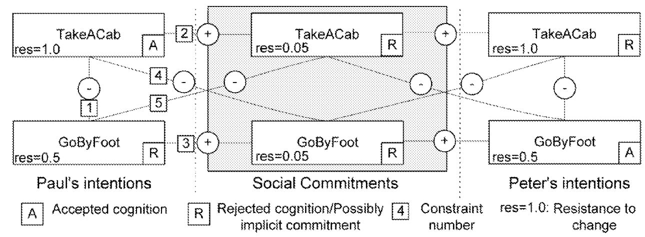

Figure 8: Cognitive models of Paul and Peter.

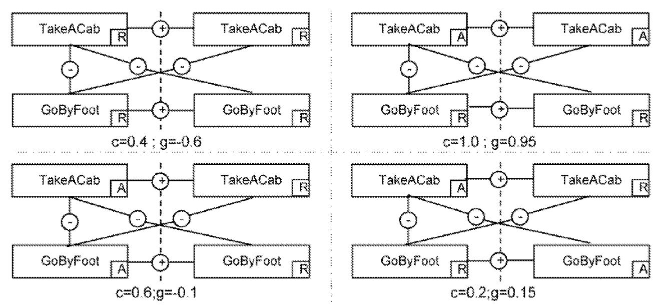

Figure 9: Sates explored by Paul

Paul will thus propose to Peter to play this game, we suppose that Peter is dialogically cooperative and would accept to play the game. Then, according to the request game rules, Paul will produce a directive speech act with an appropriate illocutionary force intensity degree[12].

Before replying, Peter will check if he does not have a higher incoherence to reduce by searching its own most useful change of cognition and locate the commitment from Paul toward him to go by foot, as shown on figure 10.

Thus, Peter will embed a DIAGAL request dialogue game concerning this commitment. Paul will answer Peter according to its coherence (which would decrease in case of acceptance) and deny the proposition and the resistance to change of the still rejected commitment will increase. The embedded request game is then closed. To illustrate the attitude change, we have drastically increased the resistance of change of the commitment of taking a cab in order that Peter's expected utility function will select the intention that Paul went by foot as the most potentially useful change. At the end of this embedded dialogue game, Peter's treatCommitments() procedure will recall the underlying BDI architecture for

[10]JACK is a commercial JAVA agent framework due to Agent Oriented Systems (AOS) which implements PRS (Procedural Reasoning System) and dMars (Distributed Multi Agent Reasoning System) concepts [13].

[11]Considerations about the hybrid symbolic connextionnist knowledge representation techniques would get us out of the scope of this article. We refer the interested reader to Sun's work [28].

[12]We illustrate our example with the use of basic illocutionary forces intensity degree for the speech/dialogue acts (here the "offer"), but DIAGAL allows us to choose a specific strength degree for each speech act. Thus, the strength degree could have been linked to : (1) Paul's current incoherence magnitude, (2) Paul's expected increase of coherence, that is the expected utility and (3) social positions of Peter and Paul, . . .

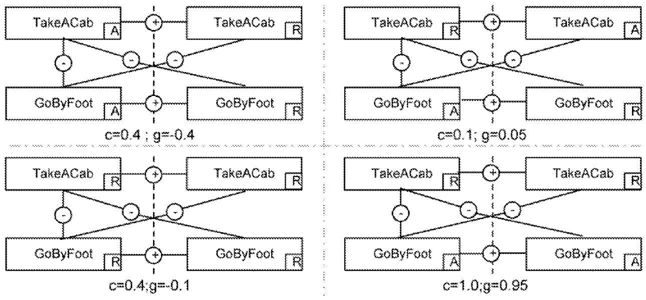

Figure 10: States explored by Peter before replying.

a re-deliberation which would at least include the rejection of the "intention to" that Paul went by foot.

Propagating attitude change and re-deliberation (which would normally be processed by the underlying architecture) is simulated in our present system by systematically revising as many intentions as possible as long as it increases whole coherence. The new cognitive models of the agents after this dialogue are those of Figure 11. Paul's intentions remains unchanged since no social commitment conflicts with its intentions while Peter's ones have been reevaluated.

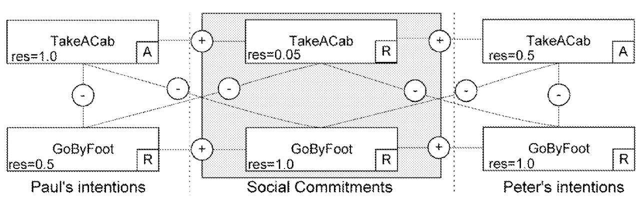

Figure 11: Cognitive models of Paul and Peter.

Peter, according to his new set of intentions will then accept Paul's offer to take a cab and they will finally quit the embedding dialogue offer game. After this dialogue, both agents will have all their constraints satisfied (i.e. a coherence of 1).

Resulting dialogues

The diagram of sequence shown on Figure 12 illustrates the messages exchanged between Paul and Peter as detailed above. This diagram is actually part of the action board which DGS fills during the execution so that the user can see what the agents are doing.

The two dialogue games initiated by Paul and Peter are presented as well as speech-acts used by both agents. Notice that all those steps were held automatically by the agents implementing our coherence theory for communication pragmatics in the way described earlier.

In the case where Peter is given the initiative at the beginning, the symmetrical dialogue would have happened, Peter trying to establish the commitment of going by foot, Paul imbricating a game on the commitment of taking a cab, denied by Peter and both finally agreeing on Paul going by foot. In that case the dialog result in the opposite situation. This is normal since we consider that the commitments socially rejected by dialogue gain a very high resistance to change. It results in a non-persistance of intentions in case of refusal (i.e. an open-minded commitment strategy). In that particular case (chosen in order to simplify the exemple), dialogue initiative play a crucial role.

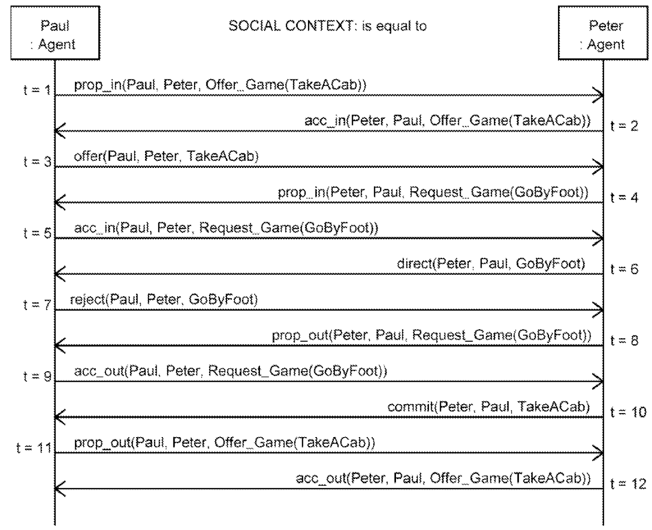

Figure 12: Dialogues between Paul and Peter

7. CONCLUSION AND PROSPECTS

The cognitive coherence theory for agent communication pragmatics allows modelling a great number of agent communication dimensions while being computational. This paper describes our exploration in applying the cognitive coherence pragmatic theory for BDI agents communication. The presented practical framework relies on our dialogue games based agent communication language (DIAGAL) and our dialogue game simulator toolbox (DGS). It provides the necessary theoretical and practical elements for implementing the theory as a new layer over classical BDI agents. In doing so, it brought a general scheme for automatizing agents communicational behavior.

Classically, practical reasoning equals deliberation plus means-ends reasoning. Deliberation is about deciding what states of affairs the agent wants to achieve whereas means-ends reasoning is about deciding how to achieve these states of affairs. Within our model, coherence gain evaluation trough the expected utility function is part of the deliberation process whereas selecting a dialogue games by unifying its success conditions with the wanted social result is part of the mean-end reasoning. We also insist on the dialogue effect on agent's private mental states trough the attitude change process. This process is activated by a kind of reconsider() function (see [24]) which has been modelled and integrated in our expected utility function and which results depends on the chosen individual commitment strategy.

Although the architecture presented in this paper is efficient, much more work remains to be done. In particular we want to : (1) work more profoundly on the links between private and public cognitions (2) provide a well-founded theory for sanction and social relations dynamic management[13] (3) extend the current framework with argumentation seen as constraints propagation allowing agents to reason about others' cognitive constraints and thus taking them into account, introducing cooperation.

In this article we choose to apply our theory as a new layer above the existing BDI architectures. But, a long term work would be to propose a pure coherentist approach for the whole cognitive

[13]Memorizing dialogue utility measures defined in our coherence theory could be of great help for this purpose.

agents architecture. This would permit to take more advantage of the power of coherentist approachs [29], using the powerful hybrid symbolic-connexionist formalisms attached with them.

8. REFERENCES

[1] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the 4th Conference on Multi-Agent Systems (ICMAS)*, Boston, 2000.

[2] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In *Proceedings of the European Conference on Artificial Intelligence (ECAI)*, Berlin, 2000.

[3] T. Bench-Capon, P. E. S. Dunne, and P. H. Leng. Interacting with knowledge-based systems through dialogue games. In *11th International Conference on Expert Systems and Applications Avignon*, pages 123–140, 1991.

[4] M. E. Bratman. What is intention? In P. R. Cohen, J. L. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 15–32. The MIT Press: Cambridge, MA, 1990.

[5] J. Brehm and A. Cohen. *Explorations in Cognitive Dissonance*. John Wiley and Sons, inc, 1962.

[6] B. Chaib-draa, N. Maudet, and M.A. Labrie. Diagal, a tool for analysing and modelling commitment-based dialogues between agents. In *Proceedings of Canadian AI 2003*, 2003.

[7] M. Dastani, J. Hulstijn, and L. V. der Torre. Negotiation protocols and dialogue games. In *Proceedings of the BNAIC*, 2000.

[8] F. Dignum, B. Dunin-Keplicz, and R. Vebrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Intelligent Agent VII: Proceedings of the Seventh International Workshop on Agent Theories, Architectures and Languages (ATAL 2000)*, pages 150–166, LNAI, 1986, Berlin, Germany, Springer, 2000.

[9] F. Dignum and M. Greaves, editors. *Issues in agent communication*. LNAI N0 1916, Springer-Verlag, Berlin, 2000.

[10] C. Excelente-Toledo, R. A. Bourne, and N. R. Jennings. Reasoning about commitments and penalties for coordination between autonomous agents. In *Proceedings of Autonomous Agents*, 2001.

[11] L. Festinger. *A Theory of Cognitive Dissonance*. CA : Stanford University Press, 1957.

[12] R. F. Flores and R. C. Kremer. A formal theory for agent conversations for actions. *Computational intelligence*, 18(2), 2002.

[13] N. Howden, R. Rnnquist, A. Hodgson, and A. Lucas. Jack intelligent agents summary of an agent infrastructure. In *5th International Conference on Autonomous Agents*, 2001.

[14] J. Hulstijn. *Dialogue models for inquiry and transaction*. PhD thesis, University of Twente, The Netherlands, 2000.

[15] J. Levin and J. Moore. Dialogue-games: meta-communication structure for natural language interaction. *Cognitive science*, 1(4):395–420, 1978.

[16] N. Maudet. *Modéliser les conventions des interactions langagières: la contribution des jeux de dialogue*. PhD thesis, Université Paul Sabatier, Toulouse, 2001.

[17] N. Maudet. Negociating games — a research note. *Journal of autonoumous agents and multi-agent systems*, 2002. (Submitted).

[18] N. Maudet and B. Chaib-draa. Commitment-based and dialogue-game based protocols–new trends in agent communication language. *The Knowledge Engineering Review*, 17(2):157–179, 2002.

[19] P. McBurney, S. Parsons, and M. Wooldridge. Desiderata for agent argumentation protocols. In *Procceedings of the First International Conference on Autonomous Agents and Multi-Agents*, 2002.

[20] D. Moore. *Dialogue game theory for intelligent tutoring systems*. PhD thesis, Leeds Metropolitan University, England, 1993.

[21] P. Pasquier and B. Chaib-draa. The cognitive coherence approach for agent communication pragmatic. In *Proceedings of AAMAS'03*. To appear, 2003.

[22] P. Pasquier and B. Chaib-draa. Cohérence et conversations entre agents : vers un modèle basé sur la consonance cognitive. In J.P. Müller and P. Mathieu, editors, *Systèmes multi-agents et systèmes complexes, actes des JFIADSMA'02*, pages 188–203, Hermes Science Publication, Paris, 2002.

[23] P. Pasquier and B. Chaib-draa. Engagements, intentions et jeux de dialogue. In *Modèles Formels de l'Interaction (MFI'03)*, 2003. To appear.

[24] A. S. Rao and M. P. Georgeff. An abstract architecture for rational agents. In C. Rich, W. Swartout, and B. Nebel, editors, *Proceedings of Knowledge Representation and Reasoning (KR&R-92)*, pages 439–449, 1992.

[25] C. Reed. Dialogue frames in agent communication. In *Proceedings of the Third International Conference on MultiAgent Systems (ICMAS)*, 1998.

[26] J. R. Searle. Collective intentions and actions. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 401–416. The MIT Press: Cambridge, MA, 1990.

[27] M. P. Singh. A social semantics for agent communication language. In *[9]*, pages 31–45. 2000.

[28] R. Sun. *Connectionist-Symbolic Integration*, chapter An introduction to hybrid connectionist-symbolic models. Lawrence Erlbaum Associates., 1997.

[29] P. Thagard. *Coherence in Thought and Action*. The MIT Press, 2000.

[30] P. Thagard and K. Verbeurgt. Coherence as constraint satisfaction. *Cognitive Science*, 22:1–24, 1998.

[31] R. Tuomela and K. Miller. We-intentions. *Philosophical Studies*, 53:367–389, 1988.

[32] G.H. von Wright. *Freedom and determination*. North Holland Publishing Co., 1980.

[33] D. Walton and E. Krabbe. *Commitment in dialogue*. State University of New York Press, 1995.

[34] R. Wickland and J. Brehm. *Perspectives on Cognitive Dissonance*. NY: Halsted Press, 1976.