Available online at www.sciencedirect.com
SCIENCE DIRECT®
Cognitive Systems Research 6 (2005) 364–395
Cognitive Systems RESEARCH
www.elsevier.com/locate/cogsys

Agent communication pragmatics: the cognitive coherence approach

Action editor: Vasant Honavar

Philippe Pasquier *, Brahim Chaib-draa

DAMAS Laboratory, Computer Science and Software Engineering Department of Laval University, Sainte-Foy, Que., Canada G1K 7P4

Received 19 February 2005; accepted 5 March 2005
Available online 22 June 2005

Abstract

Different approaches have investigated the syntax and semantics of agent communication languages. However, these approaches have not indicated how agents should dynamically use communications. Instead of filling this pragmatics gap, most approaches have mainly focused on the 'structure' of dialogues even though developers are more interested in agents' capabilities of having 'useful' automated conversations with respect to their goals rather than in their abilities to structure dialogues. This led us to work on a theory of the use of conversations between agents. In this paper, we propose a pragmatics theory which extends and adapts the cognitive dissonance theory (a major theory of social psychology) to multi-agent systems by unifying it with the theory of coherence in thought and action that issues from computational philosophy of mind. Precisely, we show how this theory allows us to provide generic conceptual tools for the automation of both agent communicational behavior and attitude change processes. This new motivational model is formulated in terms of constraints and elements of cognition and allows us to define cognitive incoherences and dialogue utility measures. We show how these measures could be used to solve common problems and answer some critical questions concerning agent communication frameworks use. Finally, our exploration in applying the cognitive coherence pragmatics theory as a new communication layer over classical BDI agents is presented. It relies on our dialogue games based agent communication language (DIAGAL) and our dialogue games simulator toolbox (DGS). The resulting framework provides the necessary theoretical and practical elements for implementing our theory. In doing so, it brings in a general scheme for automatizing agents' communicational behavior as it is exemplified in this article.

* Corresponding author.
E-mail addresses: pasquier@iad.ift.ulaval.ca (P. Pasquier), chai@iad.ift.ulaval.ca (B. Chaib-draa).
URLs: http://www.damas.ift.ulaval.ca/~pasquier/ (P. Pasquier), http://www.damas.ift.ulaval.ca/~chaib/ (B. Chaib-draa).

1389-0417/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2005.03.002

Keywords: Multi-agents systems; Agent communication language; Pragmatics; Cognitive coherence; Dialogue games; Social commitments

1. Introduction

Agents and multi-agents techniques allow the conception and development of complex software applications. In the current distributed data processing paradigm, the fundamental characteristic of these systems is the agents' skill in communicating with each other about their individual and collective goals in a useful way. While numerous works have aimed to define agents' communication languages (hereafter noted ACL for agent communication language), few have concentrated on their dynamic and automatic use by agents. This last task is left to the system designers, who usually specify manually, by means of rules, the agent communicational behavior. In this paper, we introduce a theoretical framework, along with its first implementation, for the automation of this behavior as a first step to fill this gap.

After explaining our motivations in detail (Section 2), we present a new approach for agent communication pragmatics: the cognitive coherence theory. This conceptual framework is based on a unification of the cognitive dissonance theory (Festinger, 1957), which is one of main motivational theories in social psychology with Thagard's theory (Thagard, 2000) of coherence in thought and action (Section 3). Then, we indicate how this conceptual framework supplied generic answers in order to automate different aspects of conversations in multi-agents systems (Section 4). In particular, we indicate how this theory answers (even partially) what we consider to be the main questions of pragmatics (in the sense of a theory of the dynamic use of language) and which are loosely treated in the artificial intelligence (AI) and multi-agent systems (MAS) literature:

- When should an agent take a dialogue initiative, on which subject, with whom and why (Section 4.2)?
- Which type of dialogue type (dialogue unit) is chosen by this agent to do so (Section 4.3)?
- When to stop dialogue or if not, how to pursue it (Section 4.6.2)?
- How to define and measure the utility of a conversation (Section 4.6)?
- What are the impacts of the dialogue on agents' attitudes (Section 3.2)?
- Which intensity to give to illocutionary forces of dialogue acts (Section 4.7)?
- What are the impacts of the dialogue on agents' mood (Section 4.7)?
- What are the consequences of the dialogue on social relations between agents (Section 4.6.3)?

Finally, we show how the proposed conceptual approach has been validated in the context of BDI [beliefs, desire and intention] agents. To do so, we define BDI agents that use our dialogue game based agent language (DIAGAL; Section 5 introduces this language) in our dialogue game simulator (DGS, Section 6 presents this tool). Refinements and algorithms (Section 7) used for this validation as well as an example of the system execution and the resulting dialogue illustrate our computational theory (Section 8) of agent communication pragmatics. We then discuss some prospects of this ongoing research work (Section 9) before concluding (Section 10).

2. Motivations

2.1. Structural versus cognitive coherence

In communication theory, one generally distinguishes between *cognitive theories* and *interactional theories* (Littlejohn, 1992). Interactional theories articulate around the notion of *structural coherence* (often called conversational coherence; Craig, 1983) and deal with the shape of communication, answering questions such as: what are

the dialogue units, what are the structural regularities of conversations or what are the conventional aspects constraining the dialogue shape/structure.

In contrast, cognitive theories deal with message production as well as message perception and reception. Central questions are then: what to communicate, when to communicate it and to whom, how to understand and interpret incoming messages and how to react. These theories articulate around the notion of *cognitive coherence* and address the functional aspect of communication at both the internal and external level. Which are the elements that induce an agent to introduce a certain type of dialogue rather than another? At the internal level, how does an agent cognitively react to a statement in terms of mental states update? At the external, public level (toward the environment), what are the commitments the agent wants to obtain? Why? What is the conversation utility? Is the agent or the group of conversing agents satisfied with the conversation?

In those two approaches, coherence is a central notion. But it is necessary to guard against confusing the structural coherence of the dialogue – is it allowed to pursue the dialogue in this way? – with its cognitive coherence. Is the message content suited to previous messages and agent's mental states? Is the message content coherent with the agent's internal state? Do agents hold relevant dialogues in regard to their goals? Do agents take advantage of the conversation? One should make a difference between the respect of dialogue structural constraints (for example, to satisfy a dialogue game by respecting its rules) and the agents cognitive satisfaction. Even if these two coherence dimensions are different, they are often connected, and working on a cognitive theory does not exclude the need for an interactional theory. Indeed, when one has determined what to say, when and to whom, the question of how conversation can take place remains open. On the other hand, working on cognitive coherence allows us to exceed this level and the ideas advanced in the following sections are valid for any rich enough interactional social commitments based communication framework.

2.2. Interactional frameworks in multi-agent systems

Regarding communication, the multiagent system (MAS) community has been concentrating for some years on building a standard interactional framework (Pasquier & Chaib-draa, 2005). Main current ACLs, KQML (Finin & Fritzson, 1994) and FIPA-ACL (FIPA, 2002), are both based on speech acts theory. Semantics of messages is formulated in terms of mental states, private aspects of agents (Finin & Labrou, 1998). Dialogue is supposed to emerge from the chaining of produced speech acts stemming from agents' intentions by way of recognition and reasoning on others' intentions. This "mentalistic" approach has been criticized (Moulin, 1997; Singh, 1998). It raises the semantic verification problem; agents should be able to verify that the others act according to held dialogues.[1] For messages semantics to be verifiable, it would therefore be necessary to have access to agents' private mental states which is generally not possible. Another problem induced by this formulation is the sincerity assumption. This hypothesis, necessary for the definition of ACL's mentalistic semantics, is considered too restrictive by the MAS community. It forbids certain dialogue types in domains where such hypotheses would not hold, as is the case for negotiation dialogues in electronic business (Dignum & Greaves, 2000).

More recently, some authors have proposed social approaches for agent communication, introducing a public layer expressed explicitly in terms of social commitments (e.g., Colombetti, 2000; Flores & Kremer, 2001; Pasquier, Flores, & Chaib-draa, 2004b; Singh, 2000). These approaches allow (1) resolving the semantic verification problem, (2) getting rid of the sincerity hypothesis and (3) facilitating the treatment of the social aspects of communication.[2] Among these approaches, dialogue games (Dastani,

[1] This semantics verification should not be mistaken with the formal semantics checking: agents are implemented in accordance with the ACL mathematical or logical semantics.

[2] A detailed discussion of those results could be found in previous works (Pasquier & Chaib-draa, 2003a, 2003b).

Hulstijn, & der Torre, 2000; Maudet, 2001; Pasquier, Bergeron, & Chaib-draa, 2004a; Reed, 1998) appear to be a good compromise between strictly speech acts based approaches (with either 'mentalistic' or 'social' semantics) which do not specify anything about dialogue structure (which is supposed to emerge) and protocols which reduce the searching space for possible continuations to its strict minimum, causing the loss of the flexibility and adaptability of conversations. For those reasons as well as others exposed in previous works (Chaib-draa, Maudet, & Labrie, 2003; Maudet & Chaib-draa, 2002), we retain dialogue games as our interactional framework. In the rest of this paper, we have named *conventional* the agent communication tools that are both social commitments based and capture the conventional aspects of dialogue structuration.

2.3. Problem and research objectives

One usually considers three canonical dimensions of language: the syntax, the semantic and the pragmatics. In the MAS field, propositions about the syntactic aspects are quite satisfying: a lot of work has been done and we now have a wide range of ACLs, protocols and dialogue games interactional frameworks. If we consider that social semantics based on social commitments overcomes the major difficulties of mentalistic semantics, the semantic aspects of MAS interactional frameworks are also becoming quite satisfying. In contrast, there has been very little work done concerning the pragmatics aspects of agents communications. If we can consider that the conventional aspects of communication pragmatics are taken into account by dialogue games (due to their normative and predictive structure), *the cognitive aspects of agents communication pragmatics in social commitment based frameworks* have not been investigated at all. If one chooses to use commitment-based communications frameworks, there is a need for a theory of agent communication pragmatics, that is to say a theory for the automation of agents' communicational behaviors.

In other words, this means that the MAS community has investigated interactional theories rather than cognitive theories of social commitments based communication. In previous ACLs, associated with mentalistic semantics, a pragmatics theory based on the work of Grice (1957) in philosophy of language has been proposed. This well-known family of approaches was first formalized by Cohen and Perrault (1979) using planning tools, extended by Grosz and Sidner (1986) toward a better treatment of mental states. Finally, this intentional pragmatics, known as the root of rational interaction theories, has been reformalized within a logical framework by Cohen and Levesque (1990b, 1990a), using intention recognition allowed by (unverifiable) formal mentalistic semantics (under the sincerity assumption). Unfortunately, some computational issues tend to forbid its use in realistic MAS settings (the semantics were specified in terms of multi-modal logics which are as powerful as intractable; Dignum & Greaves, 2000). Despite the criticism and shortcomings of those intentional approaches,[3] they have been successfully applied in the mono-agent case for human computer interaction, that is, in conversational agents (Allen, 1995).

Besides those critics, the mere introduction of the public social commitments layer requires rethinking a pragmatics theory, extended to these new frameworks. Indeed, agents do not directly have to reason solely about others' private mental states (mainly intentions) but also about the social commitments induced by their interactions. These commitments stem from held conversations or from conventions sustaining the interactional framework or the system (Walton & Krabbe, 1995).

It is worth noticing that the interactional tools proposed by conventional approaches such as dialogue games do not include the necessary elements for their automatic and rational use by cognitive agents (they rely on the system designer for that purpose). Besides, conventional interactional frameworks do not supply any guarantee about

[3] Some were mentioned in previous sections, the interested reader could refer to works of Cohen (1996) or Pasquier (2002); Pasquier and Chaib-draa (2005) for others known shortcomings of rational interaction theories and strictly intentional approaches.

the utility of held conversations. Nevertheless, we are more interested in agents' capabilities of having 'useful' conversations in respect to their individual and collective goals rather than in their abilities to structure dialogues.

This is what leads us to re-investigate the cognitive aspects of agent communication pragmatics (Pasquier & Chaib-draa, 2002, 2003a). This includes problematics like: (1) automatizing agents dialogical behavior dynamics, (2) introducing and defining dialogue utility, (3) modelling the social and cognitive consequences of dialogues. Our investigation of those problems have been refined in the following objectives:

(1) to provide a theory of agent communication pragmatics well founded on cognitive sciences that would as much as possible be: computational, formal and minimal. This theory has to introduce a metric to allow the agents to consider the utility of held conversations as well as to guide them in the choice of conversations to hold.
(2) validate the theory in a realistic MAS setting: the theory must be suited to the new social commitment based and conventional interactional frameworks.

In particular, we wanted to validate our theoretical approach using our dialogue game agent language (Section 5 introduces this language named DIAGAL) and the dialogue game simulator (Section 6 presents this tool named DGS) developed in our laboratory. The following sections present our contributions concerning these two objectives beginning with the theoretical one.

3. Cognitive coherence as a motivational model

3.1. The cognitive coherence framework

In cognitive sciences, cognitions include all cognitive elements: propositional attitudes such as beliefs, desires and intentions as well perceptions, feelings, emotional constituents and 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 of maintaining or restoring physiological or psychological constants despite variations in the outside environment. 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 open 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.

The cognitive dissonance theory, initially presented in 1957 by Festinger (1957), is one of the most important theories of social psychology. It has generated hundreds of studies and extrapolations on human attitudes, behaviors, beliefs, values, decision-taking consequences, inter-personal discords and other important psychological phenomena (Harmon-Jones & Mills, 1999, Introduction). This is partially explained by the very general and abstract formulation of this theory which makes it easy to manipulate. In communication theories (Littlejohn, 1992) it appears as one of the main cognitive theories for messages reception and treatment. Numerous formalizations and models of cognitive dissonance have been produced (e.g., Harmon-Jones & Mills, 1999, part 3, Mathematical Models of Dissonance). In this paper, we propose our own, which is explicitly adapted for AI and MAS.

Our formulation is inspired by the coherence theory of the computational philosopher Thagard (2000) which allows us to directly link the cognitive dissonance theory with notions, common in AI and MAS, of elements and constraints. In our formulation, the elements are both the private (beliefs, desires, intentions) and public (social commitments) agent's cognitions. Elements are divided into two sets: set A of accepted elements (which are interpreted as true, activated or valid according to the elements type) and 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, such as: mutual exclusion, incompatibility, inconsistency and all other 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 other hand, 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 an elements partition 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) produces for the agent a tension which incites him to change. The more intense the incoherence, the stronger is 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 the case of incoherence, the agent is not only going to change his cognitions or to try to change those of the others to try to reduce it, but he is also going to avoid all the situations which risk increasing it. Those two hypotheses were verified by a great number of cognitive and social psychology studies and experiments (Wickland & Brehm, 1976). These two assumptions give us a very general *motivational* motor/scheme for our agents.

One of the major advantages of the cognitive dissonance theory captured by our formulation is its ability to supply incoherence (that is dissonance in Festinger's terminology) measures, i.e., a metric for cognitive coherence. This metric is available at every level of the system: for a cognitive element, for a set of elements, for an agent, for a group of agents or even for the whole MAS system. Because a dissonance link in Festinger's model corresponds to a non-satisfied constraint in Thagard's model and a consonance link corresponds to a satisfied constraint, these measures match exactly the dissonance intensity measures first defined by Festinger.

One can wonder in what circumstances incoherence arises. In fact, there are various situations in which incoherence can appear:

- *Initial direct contact with a situation*: a new situation can introduce new elements incoherent with preexisting cognitions.
- *A change in the situation*: a change in the situation can lead coherent cognitions to become incoherent.
- *Communication*: communication with others can introduce cognition elements which are incoherent with those of the agent.
- *Simultaneous existence of various cognitions*: in the general case, a cognition is connected with several others among which some are coherent and others incoherent.

3.2. Incoherence, social influence and attitude change

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 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. But 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 importance of the elements with which it is coherent. This resistance also depends on its type, age, as well as the way in which it was acquired: perception, reasoning, communication. 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 social commitment.

Social commitments are particular cognitions which are not individually modifiable but must be socially established. Dialogue games are formal structures for attempting to establish collectively accepted commitments. That is, in order to modify, reject or accept a social commitment 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 without being further modifiable. They are then social obligations and fix one of the poles of the constraints which are connected to them. To reduce possible incoherence while conforming to taken commitments, agents should then change their private cognitions to restore the coherence. This is the driving principle of the attitude change in our system and it formalizes the vision of the psychologists (Brehm & Cohen, 1962) on this subject, supported by a great number of experiments. A formalization of this attitude change process in the case of the application of the coherence theory in order to automate some BDI agents communicational behavior is proposed in Section 7.2 and exemplified in Section 8.

4. Agent communication as coherence seeking

4.1. Incoherence typology

This section presents a typology of incoherences which aims to introduce a simple but useful vocabulary to handle coherence problems in the explicitly distributed frame of MAS. Incoherence being conceptually close to the notion of conflict, the following typology is borrowed from work on conflicts (Dehais & Pasquier, 2000):

- *Internal and external incoherences*: an incoherence is internal when all the involved cognitions are relative to the same agent, and external when incoherent cognitions involve at least two agents. More concretely, an incoherence is external for an agent if it is an incoherence between his cognitions and that of others or social cognitions. *Shared internal incoherence* is a special case arising when several agents share the knowledge that they experience the same internal incoherence.
- *Explicit and implicit incoherences*: we define explicit as being in "the state of having knowledge of", and implicit as being in "the state of not having knowledge of".[4] An incoherence is explicit for an agent if all the involved cognitions are explicit for that agent. A dissonance is implicit for an agent if at least one of the incoherent cognitions is implicit for him. An implicit incoherence is a potential explicit incoherence. Notice that in MAS internal incoherence will doubtless always be explicit since we do not consider any implicit internal level for software agents.

4.2. Link coherence – initiative, topic and relevance

In AI, dialogue initiative usually raises particularly delicate problems. When should an agent

[4] One can have knowledge of something without being in 'the state of having knowledge of' as is the case with forgetfulness. For example, one can have the knowledge that for driving at night, it is necessary to turn on the lights, but it can happen that one forgets.

initiate a dialogue and why? The answer supplied by our coherence frame is that an agent takes the dialogue initiative if he experiences an incoherence he cannot reduce alone or he failed to reduce alone. He has then to count on the other agents' cooperation, either because he knows that it is an external incoherence which involves other agents, or because it is an internal incoherence he has no capacity to reduce alone. Among the potentially multiple incoherent elements, the agent will choose the most incoherent one as the conversation initial subject (topic).

With the relevance theory, Sperber and Wilson (1986) have advanced the idea that the speaker chooses what he is going to say by dynamically estimating the relevance of his ideas. Every cognition element relevance varies during the conversation. The speaker undertakes a speech act only when its relevance is maximal. Within our approach, according to the coherence principle, an agent who takes initiative is going to attack the incoherence which has the greatest magnitude, because it is the most cognitively relevant choice for him. The following section indicates how the coherence frame allows agents to choose which type of dialogue to engage in.

4.3. Link with dialogue types

In this section, we analyze how dialogue types observed in dialectic can be bound to cognitive coherence. Let us consider the dialogue typology from Walton and Krabbe (1995). These authors distinguish six dialogue types defined by their first purpose (to which interlocutors subscribe) and appropriate private goals of each agent (which can be incompatible, i.e., incoherent):

(1) *Persuasion*: the initial situation of this dialogue type is an external incoherence of point of view and the global purpose is to resolve it. Every participant tries to not change his private cognitions (according to their resistance to change) and to change those of the others. To do this, agents typically resort to argumentation Keefe (1991). Thus, persuasion is an external incoherence reduction technique.

(2) *Negotiation*: starting from a conflict of interest (a type of external incoherence), the global purpose is to conclude a contract, to come to an agreement. Every agent has his own purpose and wants to maximize his profit or interests. The conflict resolution is usually made by an exchange of offers and counter offers. It is frequent that dialogues of persuasion are embedded in a negotiation, offers being thus argued. It is a technique of external incoherence reduction.

(3) *Inquiry/investigation*: participants in this type of dialogue are in an initial situation of shared internal incoherence. They all suffer from the same internal incoherence and they want to inquire together to increase the efficiency of the reduction. Common purpose coincides with individual purposes. Therefore, inquiry is a shared internal incoherence reduction technique.

(4) *Deliberation*: each agent has his own preferences and all agents have to choose together among the potentially mutually incoherent offers of each one. The participants have as a common purpose to make a decision (to choose a plan or an action). Their individual purpose is to influence the decision in their interest (which could match public interest). Deliberation is thus a reduction technique for explicit external incoherence.

(5) *Information seeking*: it is the only dialogue type which is always asymmetrical. An agent tries to obtain information from others. It is a technique of internal incoherence reduction. With this dialogue type only the information applicant agent is in an incoherent state involving the volition to have some piece of information and the knowledge of not having it. The reduction is asymmetrical but to facilitate it, it frequently occurs that an agent tries to clear up his incoherence by indicating to the other agents why he is looking for such information, i.e., by making his incoherence explicit to others. Reduction of this internal incoherence can be made through a dialogue, but it can also take form in other actions (for example: reading a refer-

ence book or searching the Internet) as long as incoherence is reduced, i.e information found.

(6) *Eristic*: eristic dialogue has a highly conflicting and incoherent initial situation. Contrary to the other dialogue types it generally involves feelings and emotions more than reason, rationality and coherence. This is why we will not detail it here, but one can notice that is is indeed an external incoherence reduction attempt.

Fig. 1 summarizes the various incoherence types as well as Walton and Krabbe dialogue types which could be used to reduce them. As one can notice, all dialogue types arise from an incoherent initial situation. Notice that the reverse is false, i.e., all incoherences are not handled by dialogue. Our point is that two agents communicate if an incoherence forces them to do so. From then on, conversation might be seen, in addition to its already known characteristics, as a generic procedure for attempting to reduce incoherence. Due to the conceptual nearness between dissonance, incoherence and conflict notions, this hypothesis is close to the classic position of dialectic: any dialogue arises from a conflict (Hamblin, 1970). For some years, several authors have been insisting again on the role of conflicts in the communicative process:

- For Walton and Krabbe (1995), the question "is there a conflict?" is the base of their dialogues initial situations analysis.
- For Reed and Long (1997), "numbers of dialogues have a conflict for initial situation".
- For Dessalles (1998), "a great number of dialogues find their origins in cognitive conflicts between desires and/or beliefs".
- For Baker (1991), "dialogues result from an opposition between conflicting goals".

Within our approach, the idea is to generalize those intuitions based on results from social psychology and philosophy of mind. Here, we have presented the link between incoherence and dialogue types using Walton and Krabbe's typology but the validation of our theory by automatizing BDI agents communicational behavior using our dialogue games agent language (DIAGAL) presented in Section 7 reconsiders this dialogue type choice in the more specific and formal framework of dialogue games. Furthermore, if we leave aside the internal implicit incoherence which is not considered for artificial agents, the path following implicite external incoherence, labelled explicitation phase of the incoherence remains unexplained by our mapping. The following section is devoted to this particular case.

4.4. Link incoherence – making explicit phase

Prior to any attempt to reduce an incoherence, agents should be aware of that incoherence. We have made the assumption that internal incoherences are always explicit. But we have seen that external incoherence stays implicit for an agent until he is 'in the state of having knowledge of' all the

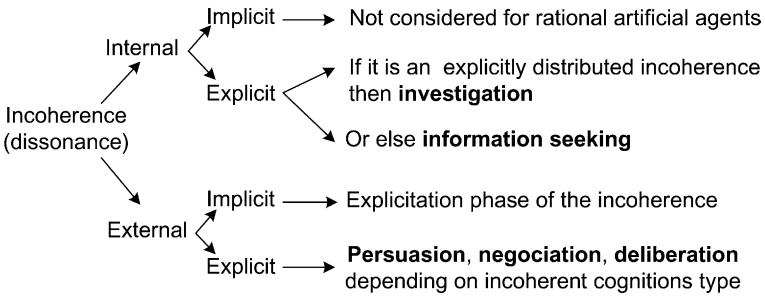

Fig. 1. Typology of cognitive dissonances and link with dialogue types.

involved cognitions. That is to say that in most cases, before a dialogue can occur as an attempt to reduce an external incoherence, the incoherence motivating it should be made explicit.

There are numerous ways to make explicit a potential implicit external incoherence. First of all, the external incoherence is generally not implicit for all the involved agents. In that case, the agent "in the state of knowing" all the involved cognitions can inform the others of the incoherence. This is what we call the explicitation phase of dialogue (see Fig. 1). For example, before a buyer and a seller negotiate a price, at least one of them should inform the other about the required price. This is usually achieved through a convention that commits the seller to announcing his starting selling price; and in that case, whether the buyer agrees with the price or not. If the buyer agent agrees with the required price, he will solve his possible internal incoherence between his intention to possess the item and his belief in not having it yet, by making an attempt to buy it. But if he does not agree with the indicated price, he will make explicit his own price, i.e., he will make explicit the external incoherence between the two prices. In that case, several following dialogues can occur: a negotiation dialogue to try to reach a common price; this would perhaps embed a persuasion dialogue where both agents will try to persuade the others that their price is the good one, an investigation in order to determine how the agents will proceed to reach a deal, ... But in all cases, none of such dialogues will occur until the external incoherence has been made explicit. In plenty of similar cases, the explicitation problem will be solved by a social convention of information spreading that will ensure that at least one agent is able to be aware of potential external incoherences. Other cases would need ad-hoc explicitation process.

Even in the case of internal incoherence that an agent i cannot reduce alone (or already failed to reduce alone), the agent i will usually have to make explicit his incoherence in order to allow a cooperative agent j helping him.[5]

[5] The conversing agent j should be cooperative in order to accept to engage resources to help i solve his internal incoherence.

4.5. Link incoherence – common project and grounding

The meaning of a conversation is not reducible to the meaning of the isolated speech acts that compose it. This is why some researchers are trying to analyze dialogue from the conversational level to the speech act level instead of the reverse. Following the ideas of Searle (1990) and Clark (1996), and the concepts of common ground, we-intention and common project, many researchers re-investigate the social aspects of communication. Cohen and Levesque (1991) have proposed persistent goals and mutual belief, whereas Grosz and Kraus (1996) have developed shared plans. Singh (1998) has next introduced social commitments as a central concept in agent communication. We subscribe to this last view by using social commitments for agent conversation semantics. In this section, we will discuss why we think the cognitive coherence approach enforces the idea that dialogues should be seen as grounded common projects (Chaib-draa & Vongkasem, 2000).

From our point of view, the explicitation phase, detailed in the previous section enabled having the explicit incoherence reduction accepted as a common project of the involved agents. In the case of external incoherence, this means that all the involved agents must accept the dialogue initiated in order to attempt reducing it (this could fail) as a common project. In the case of internal incoherence, this means that the dialogue initiated by the agent enduring an internal incoherence in order to reduce it must be adopted by a cooperative agent as their common project.

Furthermore, once the incoherence has been made explicit and a dialogue (seen as a common project) has been initiated toward its reduction, it is important that all dialogical moves are grounded and validated by all participants. This will ensure that each participant follows the same resolution path. Doing so ensures avoiding misunderstandings and mistakes about the way the incoherence is reduced (or not). In that sense, we think that the coherence theory furnishes an original justification for the need of grounding in agent conversation. This need has already been highlighted by other researchers (e.g., Traum, 1994) for differ-

ent reasons which do not preclude the preceding one.

We will not inquire into the need for explicitation and the need for grounding further here. Let us just notice that dialogue games are designed in order to ensure a perfect grounding of all dialogical moves. This is another reason why we choose this particular type of interactional framework in the following sections.

4.6. Link coherence – utility and dialogue dynamics

4.6.1. Dialogue utility

Decision theories as well as micro-economical theories define utility as a property of valuation functions. A function is a utility function if and only if it reflects the agent's preferences. In the cognitive coherence theory, according to the coherence principle (Section 3.1) and to Festinger's theory, coherence is preferred to incoherence, i.e., agent satisfaction is proportional to its coherence degree. It follows that action utility gain can be computed as simple coherence measures differences. More precisely, the expected utility for a conversation is equal to the difference between the intensity degree of the incoherence addressed by dialogue, and the expected incoherence degree after this dialogue – if successful in favor of the agent. Agents can also calculate the utility of a conversation dynamically by working out the incoherence degrees again during the dialogue. When a dialogue unit ends, either incoherence is reduced and the dialogue ends, or the agent can keep on trying to reduce it. This gives us a precious (this is a well known problem in the field) stopping condition. The following subsections explain how those subjective utility measures could be useful in guiding the agent in his communicational behavior for both the intra- and the inter-dialogue dynamic.

4.6.2. Intra-dialogue dynamic

An agent selects a dialogue type (that is a dialogue unit) according to the incoherence type which he wishes to reduce. But during this reduction process, other related incoherences can appear and these sometimes have to be reduced so that the main reduction can continue. This is what makes agents embed sub-dialogues to reduce new incoherences before resuming the main dialogue concerning the initial incoherence. Most dialogue games interactional frameworks provide syntactic facilities for doing so. In other cases, incoherence can move, leading the conversing agents to chain two dialogue units. According to the cognitive coherence theory, the dialogue structuration is determined by the incoherence reduction chaining.

4.6.3. Inter-dialogue dynamic

Dialogues are attempts to reduce incoherence which can fail. More precisely, following a low utility valued dialogue, i.e., incoherence is not reduced, the agent has to decide how to act. The agent will probably persevere in his reduction attempt by taking this failure into account. If it is still possible, he will propose a different dialogue type or a different proposition of the same dialogue type or else he will update his mental states through attitude change as described in Section 3.2. But in all cases he should take note of this failure which can be useful to guide him for the following dialogues. If on the contrary, the dialogue is useful and ends with a success of the incoherence reduction, the agent will have to update his social relationships accordingly.

In particular, in open and heterogenous MAS, an agent is led to communicate with unknown agents, it is then necessary for him to form an idea of dialogues held with them. The agent will be able to take into account the utility of previously held dialogues to select these interlocutors. It might be in the agent's interest to strengthen exchanges with agents with whom dialogues are useful and numerous incoherences (i.e., problems) are resolved. In contrast, he will be able to take into account useless dialogues by weakening his social links with the involved interlocutors. Dialogue utility measures supply information which can be used by a social relations management tool. This aspect has not been developed yet and will be part of our future work.

4.7. Link coherence – mood, intensity

Recently, the need to integrate emotions into artificial agents has appeared (Bates, 1994). Our

cognitive coherence model allows making a direct link between coherence measures and agent's mood. Precisely, our theory supplies a value system in which a coherence state is a comfortable state and coherence gains are satisfactions and reassurance that could lead to happiness, etc. On the contrary, an incoherent state is a discomfort state and an agent can be worried or afraid of potentially future incoherences or disappointed or stressed by a failed attempt at reduction and the persistance of an incoherence, etc. Notice that this view meets the modern characterization of attitudes in social psychology where emotions and cognitions are strongly related (Erwin, 2001).

In addition, following the theoretical approach of Searle and Vanderveken (1985), some MAS interactional frameworks allow using various intensity degrees for speech acts illocutionary forces. But, to our knowledge, no agent theory indicates how this selection should be made. In our case, quantitative measures defined by the cognitive coherence theory supply means to guide the agent in the choice of the suitable intensity degree.

Since a conversation is engaged in as an attempt to reduce an incoherence, its magnitude gives the importance of the resulting conversation. The incoherence intensity influences the choice of intensity degrees of used speech acts in a direct way. For example, an agent who needs information to reduce an internal incoherence is going to enter an information seeking dialogue which includes directive act(s). The intensity degree of the illocutionary force is then going to depend on the intensity of the aforementioned incoherence: (1) an invitation or advice if the incoherence is very light, (2) a recommendation, a demand if it is a little more intense and (3) a plea, an order or an entreaty if the incoherence magnitude is very high and consequently its reduction crucial.

If these parameters of emotions, mood and dialogue acts intensity seem less important for completely artificial MAS, this track is interesting for mixed communities (involving both humans and artificial agents), i.e., human machines interfaces, conversational agents and intelligent tutorial systems, among others. Obviously, this intensity degree selection factor is not unique. Other factors can intervene in this choice. For example, (1) social agreements (it is generally forbidden to give an order to a superior in the hierarchy) or (2) social relations among agents (nearness, confidence, trust, past of the relation) are also important for selecting those intensity degrees.

This concludes our review of the different dimensions that our cognitive coherence approach for agent communication allows being treated. The next sections will be devoted to the presentation of our first application of those ideas to the communicational behavior automation of some belief, desire and intentions (BDI) agents. But prior to that, the next section introduces the agent communication language our agents will use to communicate.

5. 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 communicative acts in terms of public notions, e.g., social commitments (Chaib-draa et al., 2003; Pasquier et al., 2004a). As we saw in Section 2, the use of these public cognitions allows us to overcome classical difficulties of "intentional" agent communication approaches: the sincerity hypothesis does not hold anymore and the semantic verification problem is solved.

5.1. Social commitments

As our approach is based on social commitments, we start with some details about this notion.[6] 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

[6] A more complete account of our model of flexible social commitments and their enforcement can be found in (Pasquier et al., 2004b).

partner or a group. In the line of Walton and Krabbe (1995), we distinguish action commitments from propositional commitments.

Commitments are expressed as predicates with an arity of 6. An accepted action commitment thus takes 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 are usually material sanctions (economical sanctions, repairing actions, etc.). An accepted propositional commitment would 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 Singh (2000), 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 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* (Maudet, 2001; McBurney, Parsons, & Wooldridge, 2002), 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.

5.2. Game structure

We share with others (namely, Dastani et al., 2000; Flores & Kremer, 2002; McBurney et al., 2002) the view of dialogue games as structures regulating the mechanism under which some commitments are discussed through the dialogue. However, unlike Dastani et al. (2000) or McBurney et al. (2002), we adopt a strict commitment-based approach within game structure and express the dialogue rules in terms of dialogical commitments. Unlike Flores and Kremer (2002) 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.

5.3. Grounding and composing the games

The specific question of how games are grounded through the dialogue is certainly one of the most delicate (Maudet, 2003). Following Reed (1998), we assume that agents can use some meta-acts of dialogue to handle the games structuration and thus propose to enter a game, propose to quit a 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. Fig. 2 indicates the current contextualization 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 agent x, agent y is committed to accept (*acc.in*), to refuse (*ref.in*) or to propose entering another game j' (*prop.in*(y,x,j')), which

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

Fig. 2. DIAGAL contextualization game.

would lead to a presequencing type of dialogue games structuration.

Concerning the possibility of combining the games, the seminal work of Walton and Krabbe (1995) and the follow-up formalization of Reed (1998) have focused on the classical notions of *embedding* and *sequencing*. Even if recent works, including ours, extend this to other combinations (Chaib-draa et al., 2003; McBurney et al., 2002), in our present simulation framework, we only consider the three games' compositions allowed by the previous contextualization 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 explicit 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.

5.4. Basic games

5.4.1. Completeness

While in (Pasquier et al., 2004a) we have introduced a set of 12 games that gives a complete and sound operational semantics of the social commitments model described in Pasquier et al. (2004b). Here, we will only introduce four basic building dialogue games, which are exactly those which lead (in the 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*).

This means that under the assumption that social commitments will be created sequentially, *this set of games is complete for commitments creation*. In other words, using our social commitments typology, a state of the public layer that is not reachable through the use of those four dialogue games does not exist.

The next subsections detail those four games. Sanctions were omitted in our games specifications for better readability. Within commitments, time is expressed using a simple instant theory with $<$ as the precedence relation. Notice that the game rules structure provides an elegant turn-taking mechan-

ism by entailing that $t_j < t_k < t_f$, where t_j is the time when the game is opened and t_f the instant where the last turn of the game has been played.

5.4.2. Request game (rg)

This game captures the idea that the initiator (x) "requests" the partner (y) an action α and the latter can "accept" or "reject". The conditions and rules of the request game are indicated by Fig. 3.

5.4.3. Offer game (og)

An offer is a promise that is conditional on 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 accepts α. Conditions and rules of the DIAGAL offer game are presented in Fig. 4.

E_{rg}	$\neg C(y,x,\alpha,t_i)$ and $\neg C(y,x,\neg\alpha,t_i)\ \ \forall\ t_i,\ t_i < t_j$
S_{rg}	$C(y,x,\alpha,t_f)$
F_{rg}	$\neg C(y,x,\alpha,t_f)$
R_{rg}	1) $C_g(x,y,request(x,y,\alpha),t_j)$
	2) $C_g(y,x,request(x,y,\alpha) \Rightarrow$
	$\quad C_g(y,x,accept(y,x,\alpha)\|refuse(y,x,\alpha),t_k),t_j)$
	3) $C_g(y,x,accept(y,x,\alpha) \Rightarrow C(y,x,\alpha,t_f),t_j)$
	4) $C_g(y,x,refuse(y,x,\alpha) \Rightarrow \neg C(y,x,\alpha,t_f),t_j)$

Fig. 3. Conditions and rules for the *request game*.

E_{og}	$\neg C(x,y,\alpha,t_i)$ and $\neg C(x,y,\neg\alpha,t_i)\ \forall\ t_i,\ t_i < t_j$
S_{og}	$C(x,y,\alpha,t_f)$
F_{og}	$\neg C(x,y,\alpha,t_f)$
R_{og}	1) $C_g(x,y,offer(x,y,\alpha),t_j)$
	2) $C_g(y,x,offer(x,y,\alpha) \Rightarrow$
	$\quad C_g(y,x,accept(y,x,\alpha)\|refuse(y,x,\alpha),t_k),t_j)$
	3) $C_g(x,y,accept(y,x,\alpha) \Rightarrow C(x,y,\alpha,t_f),t_j)$
	4) $C_g(x,y,refuse(y,x,\alpha) \Rightarrow \neg C(x,y,\alpha,t_f),t_j)$

Fig. 4. Conditions and rules for the *offer game*.

E_{ig}	$\neg C(x,y,p,t_i)$ and $\neg C(x,y,\neg p,t_i)\ \forall\ t_i,\ t_i < t_j$
S_{ig}	$C(x,y,p,t_f)$
F_{ig}	$\neg C(x,y,p,t_f)$
R_{ig}	1) $C_g(x,y,inform(x,y,p),t_j)$
	2) $C_g(y,x,inform(x,y,p) \Rightarrow$
	$\quad C_g(y,x,accept(y,x,p)\|refuse(y,x,p),t_k),t_j)$
	3) $C_g(x,y,accept(y,x,p) \Rightarrow C(x,y,p,t_f),t_j)$
	4) $C_g(x,y,refuse(y,x,p) \Rightarrow \neg C(x,y,p,t_f),t_j)$

Fig. 5. Conditions and rules for the *inform game*.

5.4.4. Inform game (ig)

Notice that a human partner can be disposed to be in accord or agreement with someone without uttering a word. He can also agree by performing 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 the DIAGAL inform game are those of Fig. 5.

5.4.5. 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 structure indicated by Fig. 6 for the ask game.

5.4.6. Commitments reneging and non-monotonicity

A special reneging action is available at any time for both the debtor or the creditor of each established social commitment. This dialogical

E_{ag}	$\neg C(y,x,p,t_i)$ and $\neg C(y,x,\neg p,t_i)\ \forall\ t_i,\ t_i < t_j$
S_{ag}	$C(y,x,p,t_f)$
F_{ag}	$\neg C(y,x,p,t_f)$
R_{ag}	1) $C_g(x,y,ask(x,y,p),t_j)$
	2) $C_g(y,x,ask(x,y,p) \Rightarrow$
	$\quad C_g(y,x,agree(y,x,p)\|disagree(y,x,p),t_k),t_j)$
	3) $C_g(y,x,agree(y,x,p) \Rightarrow C(y,x,p,t_f),t_j)$
	4) $C_g(y,x,disagree(y,x,p) \Rightarrow \neg C(y,x,p,t_f),t_j)$

Fig. 6. Conditions and rules for the *ask game*.

action allows suppressing a commitment ensuring that the system is non-monotonic. Notice that the reneger will have to cope with the associated sanctions (which are different for the debtor and for the creditor). Their role is precisely to be an incitement toward the respect of social commitments, which are not necessarily strict obligations in the general case. If the sanction explicitly attached with negotiated commitments are usually material ones: economical sanctions, repairing actions, etc. There could also be implicit social sanctions acting in time associated with a renege (trust, reputation, etc.). Since we did not investigate a whole agent architecture in this article, we leave sanctions as a realistic conceptual abstraction along with the social commitment specification.

While this section has given sufficient information for the purpose of this paper, we refer the interested reader to Pasquier et al. (2004b) for a more complete specification of our social commitment model and to Pasquier et al. (2004a) for a more complete specification of DIAGAL and its various extra-features and properties. The next section describes the dialogue game simulator in which our implementation and validation work took place.

6. The dialogue game simulator

We have developed a toolbox, the dialogue game simulator (DGS), 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 aims to be an effective tool for games testing and validation as well as a means 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™ agent technology (Howden, Rönnquist, Hodgson, & Lucas, 2001). In this section, we briefly present the various components of DGS.

6.1. 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 these files. That gives designers and users a means 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 load them when connecting to the DGS.

6.2. 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 the commitments of other 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 each agent's operation conforms to the current contextualization and opened dialogue games.

6.3. 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 is closed and withdrawn from the stack. This stack is also used to manage priority between the games: the top element having more priority over the bottom element.

6.4. 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. There could be several dialogue workspaces open in parallel, allowing several pairs of agents to hold dialogues at the same time.

In Fig. 7, we present a simplified overview of the DGS framework including two agents interacting through 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 place them in their agendas.

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

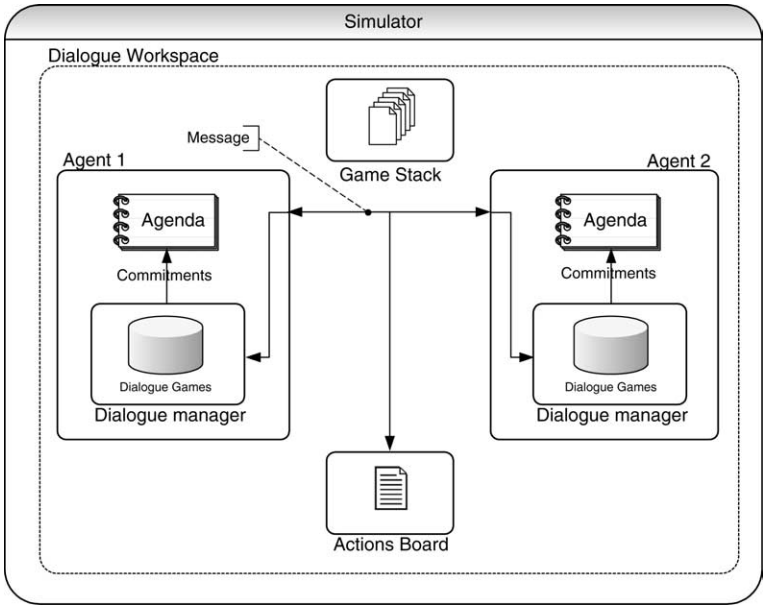

Fig. 7. Simulator overview.

coherence theory to automate dialogues between BDI agents. Those dialogues would take place in the DGS framework using precisely DIAGAL dialogue games presented in the previous sections.

7. Integrating coherence theory to BDI agents

7.1. Linking private and social cognitions

Since we do not propose a whole coherentist approach for agent modelling, we validate our approach by extending the classical belief, desire and intention (BDI) framework[7] so that it can fit with our approach. In particular, traditional BDI frameworks do not involve social commitments treatment.

Choosing a social commitments based approach for BDI agent communication leads us to extend the intentional paradigm for agent practical reasoning issued from rational interaction theories: *a 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 pragmatics 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 intentions we receive are either *social individual intentions* or *failed individual intentions*.[8] Social individual intentions are intentions concerning goals which require social aspects to be worked on. For example, an employee who 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 be considered as 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. On the other hand, we call failed individual intention, individual intentions which do not match any individual plan or whose associated plan has already failed. This last type 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 to integrate 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:[9]

- According to the classic practical reasoning scheme, private cognitions end up in intentions through deliberation and we make the usual distinction between *intention to* (do something or make someone do something) and *intention that* (a proposition holds) as introduced in (Bratman, 1990).
- Regarding public cognitions, we distinguish *commitments in action* from *propositional commitments* (Walton & Krabbe, 1995).
- An accepted social commitment is the socially accepted counterpart of an intention. Commitments in action are the counterparts of "intentions to" while propositional commitments are the counterparts of "intentions that".

Let us take an example to illustrate these relations. If an agent X has the accepted individual social intention that another agent Y achieve an

[7] We refer the reader to work of Rao and Georgeff (1995) or Wooldridge (2001) for a complete introduction of the BDI agents model.

[8] 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 (1990) or Tuomela and Miller (1988). Here, intentions are classical private intentions.

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

action α, this intention is linked by a positive constraint to the potential social commitment from Y toward X to achieve α. Notice that this particular social commitment is only a potential one, used by X to reason on the public social layer. The link we are establishing between an intention and the corresponding social commitment is reifying in a positive constraint. Other 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. It is worth noticing that those links do not say anything about the acceptance or rejection of those potential commitments that would follow from a dialogue. According to our closed world hypothesis for the acceptance of social commitments (all non-explicitly accepted commitments are rejected), social commitments are rejected by default.

These relations between the private and public cognitions are not completely new since many authors have already considered individual intentions as a special kind of individual commitment (von Wright, 1980; Bratman, 1990). Our links extend this to reach the social level in the appropriate cases by saying that social individual intentions or failed individual intentions should ideally be associated with the corresponding social commitments. These links complement previous work from Singh (1991), which introduces the idea of linking individual and social commitments.

7.2. 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. According to both their nature and their resistance to change, all the cognitions are not equally modifiable. 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 in 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 sanctions associated with withdrawing or reneging an accepted commitment. Furthermore, we assume that a discussed commitment which is still rejected at the end of the dialogue (the failure conditions have been reached) will gain in resistance to change. The point here is that an agent could not make attempts to have the desired commitment accepted indefinitely. Consequently, 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 tend to 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. As we have seen in Section 3.2, this is the driving principle of the *attitude change* in our system and it formalizes the vision of social psychologists on this issue.

In the present BDI oriented framework, the only private cognitions we consider within our coherence system are the intentions, but we assume that in case of attitude change (the acceptance state of an intention is to be modified) 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 8.

In MAS, knowing when an agent should try to modify the environment (the public social commitments layer, among others) to satisfy his intentions,

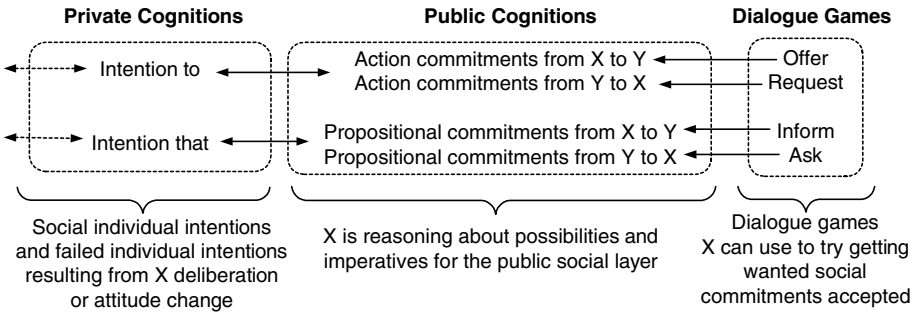

Fig. 8. Links between private cognitions, public cognitions and DIAGAL dialogue games.

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 takes the form: when should an agent reconsider his intentions 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 persistence types as introduced by Rao and Georgeff (1992). The main novelty is that this choice, like others, would be dynamically based on cognitive coherence based expected utility.

In Fig. 8, we sum up (hiding the quantitative level of calculus) the means by which we link intentions, social commitments and DIAGAL dialogue games. From the left to right we have the failed individual intentions and the social individual intentions (which are refined in the two types introduced in Section 7.1) linked with the four possible corresponding commitments types (the four ones seen in Section 5.4.1). 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 his 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 unifies with the wanted commitment. Depending on the issue of the dialogue the second way, from right to left could be used and an attitude change process could be initiated.

7.3. The expected utility function

As we have seen in Section 3.1, the whole agent cognitive coherence is expressed as the sum of weights of satisfied constraints divided by the sum of weights of all constraints.[10] 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 cognitive element is a social commitment, the agent will attempt to change it through dialogue and if it is an intention, it will be changed through attitude change. In this 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 selects the cognition which can *in the case of a successful change* be the most useful to change. A state is said to be reachable if it can

[10] Notice that the general coherence problem: to give the elements partition between A and R that maximize coherence, is NP-complete as formally demonstrated by Thagard and Verbeurgt (1998).

be obtained from the current state by modifying only one cognition acceptance state. Since all cognitions 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*, g, expressed as below:

$$g(exploredState) = coherence(exploredState) - coherence(currentState) - cost(cognitionChanged),$$

where *coherence* computes the cognitive coherence of a state, *exploredState* is the evaluated state, *currentState* is the current state, *cognitionChanged* is the cognition whose change we are examining, 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 a rejected commitment, its cost of change equals its resistance to change (which is initially low but which is possibly 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.

7.4. 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:

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

As we have seen in Section 5.1, we distinguish between two types of commitments: the dialogical ones and the extra-dialogical ones. The procedure for treating the extra-dialogical commitments (line 3) consists in updating the cognitive model of the agent by browsing extra-dialogical commitments in the agenda and operates 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 algorithm calls the underlying BDI architecture for an 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 eventually after several unsuccessful attempts, this commitment will be so expensive to establish that it will not constitute a useful change of cognition.[11] This last case could also 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 5) consists in searching for the most useful cognition to change.[12] 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.[13]

[11] This point is discussed in Section 8.2.
[12] There could be none, for example if the coherence is already maximal.
[13] Due to their completeness propriety detailed in Section 5.4.1, the four DIAGAL dialogue games ensure that there will be one.

Treating dialogical commitments (line 7) consists in exploring all the possible actions that are determined by dialogue games and selecting the one which has the best consequences for coherence. If the extra-dialogical commitment which is concerned with 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 dialogical commitments order in the agents' agendas. Finally, this algorithm is called each time:

- The underlying BDI architecture finishes a deliberation process (or a re-deliberation process after a call initiated by our algorithm for an 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 has something in his agenda. This ensures that the agent re-executes this algorithm until all dialogues are closed and that the agent will treat dialogue initiated by others. For example, when the agent receives a *prop.in* message to enter a particular dialogue game, the corresponding dialogical commitment given by the contextualization game is 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 have implemented JACK™ BDI[14] agents using this pragmatics framework to manipulate DIAGAL dialogue games within the DGS. The next section provides an example of the resulting system execution.

8. Example

Let us assume that we have two agents, Paul and Peter, who have agreed on a common plan to go to a concert of their favorite band and split the bill. A subtask of this plan is to go 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 take a cab and the accepted one that Paul go by foot.

Those social individual intentions may be associated with two corresponding potential commitments (according to links established in Section 7.1): 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 walking are incompatible, as well as the commitment to 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. Fig. 9 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.[15]

[14] JACK is a commercial JAVA agent framework from Agent Oriented Systems (AOS) which implements BDI concepts (Howden et al., 2001), in particular those of PRS (procedural reasoning system) and dMars (distributed multi agent reasoning system).

[15] Considerations about the hybrid symbolic connextionist knowledge representation techniques would take us beyond the scope of this article. Sun (1997) provides a good introduction for the interested reader.

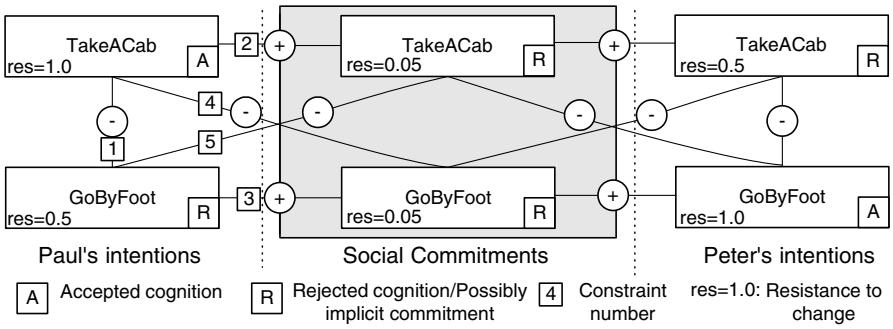

Fig. 9. Cognitive models of Paul and Peter.

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 Fig. 9, Paul has three satisfied constraints (number 1, 3 and 4) out of 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. Fig. 10 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. Peter will be the dialogue partner, since the wanted commitment is a commitment toward him. Paul will then choose between the available dialogue games whose success condition unifies with the desired commitment. The only DIAGAL dialogue game which has a success condition of the form $C(initiator, partner, action)$ is the *offer* game.

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

Before replying, Peter will check if he does not have a higher incoherence to reduce by searching his own most useful change of cognition and locate

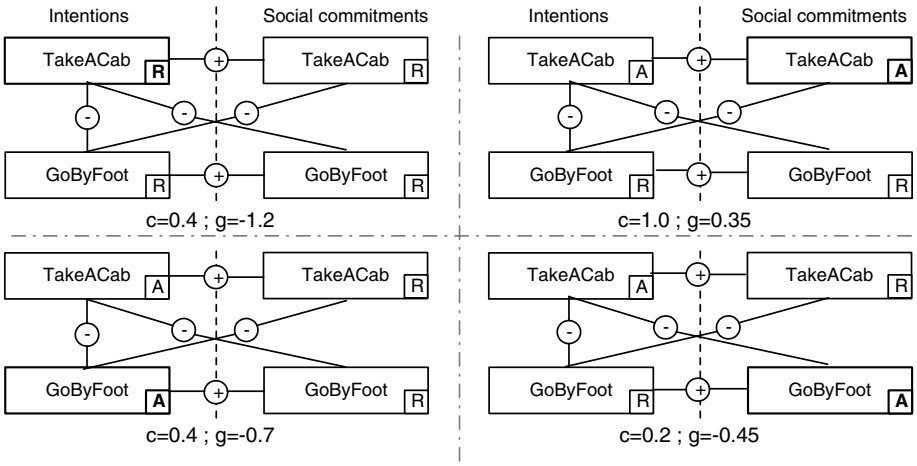

Fig. 10. States explored by Paul at the initial stage.

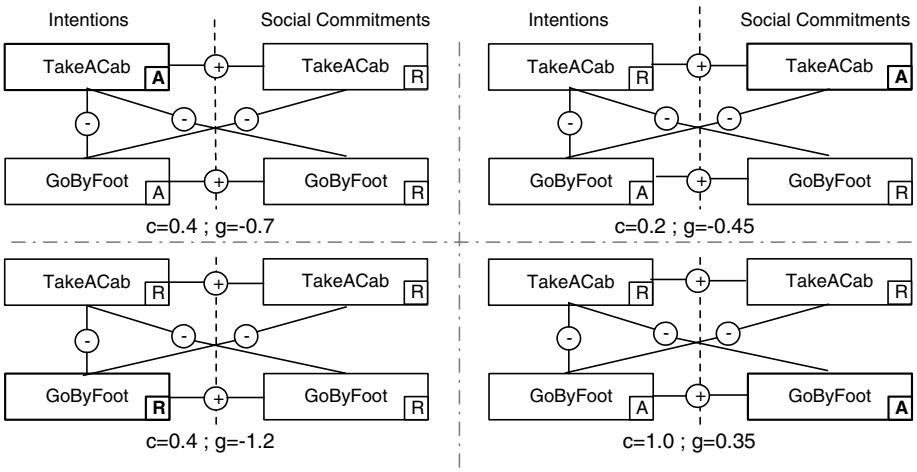

Fig. 11. States explored by Peter before replying.

the commitment from Paul toward him to go by foot, as shown in Fig. 11. He will thus 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 will reject Peter's directive act. The resistance to change of the still rejected commitment from Paul toward Peter to go by foot will increase. The embedded request game is then closed since the failure conditions have been reached. To illustrate the attitude change, we have drastically increased the resistance of change of the explicitly rejected commitment to go by foot so that Peter will not persist in trying to make it accepted. At the end of this embedded dialogue game, Peter's treatCommitments() procedure will then recall the underlying BDI architecture for an attitude change (a re-deliberation which would include the rejection of Peter's "intention to" that Paul went by foot and the acceptance of the "intention to" that Paul went by car) as explained in Section 7.4.

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 Fig. 12. Paul's intentions remain unchanged since no established social commitment conflicts with its intentions while Peter's ones have been reevaluated.

Peter, according to his expected utility calculus over his new set of cognitions will then accept Paul's offer to take a cab and they will finally quit the embedding dialogue *offer* game. After this dia-

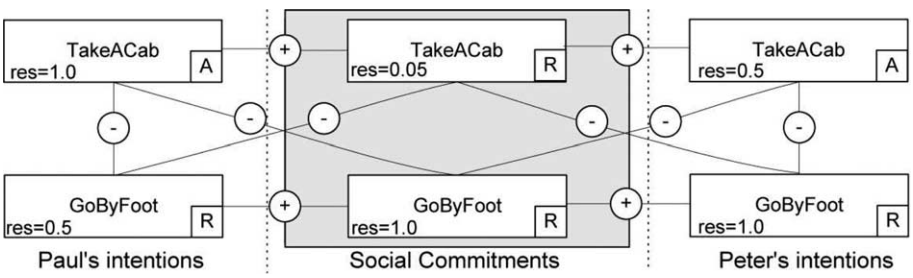

Fig. 12. Cognitive models of Paul and Peter after Peter's attitude change.

logue, both agents will have all their constraints satisfied (i.e., a cognitive coherence of 1).

8.1. Resulting dialogue

The diagram of sequence shown in Fig. 13 illustrates the messages exchanged between Paul and Peter as detailed above. This diagram is actually part of the action board that DGS fills during the execution so that the user can see what the agents are doing.

Contextualization acts (issued from DIAGAL contextualization game presented in Section 5.3) for the two dialogue games initiated by Paul and Peter are presented in Fig. 13 as well as resulting speech-acts used by both agents. Notice that all those interactions 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 dialogue results in the opposite situation. This is normal since, for this example, we consider that the commitments socially rejected by dialogue gain a very high resistance to change as previously stated. It results in a non-persistance of intentions in case of refusal (i.e., a highly influenceable open-minded commitment strategy in classic BDI vocabulary). In that particular case (chosen in order to simplify the example), dialogue initiative plays a crucial role.

8.2. Summing up the resulting behavior

Since our algorithms as well as the dialogue manager are based on the treatment of the agent's

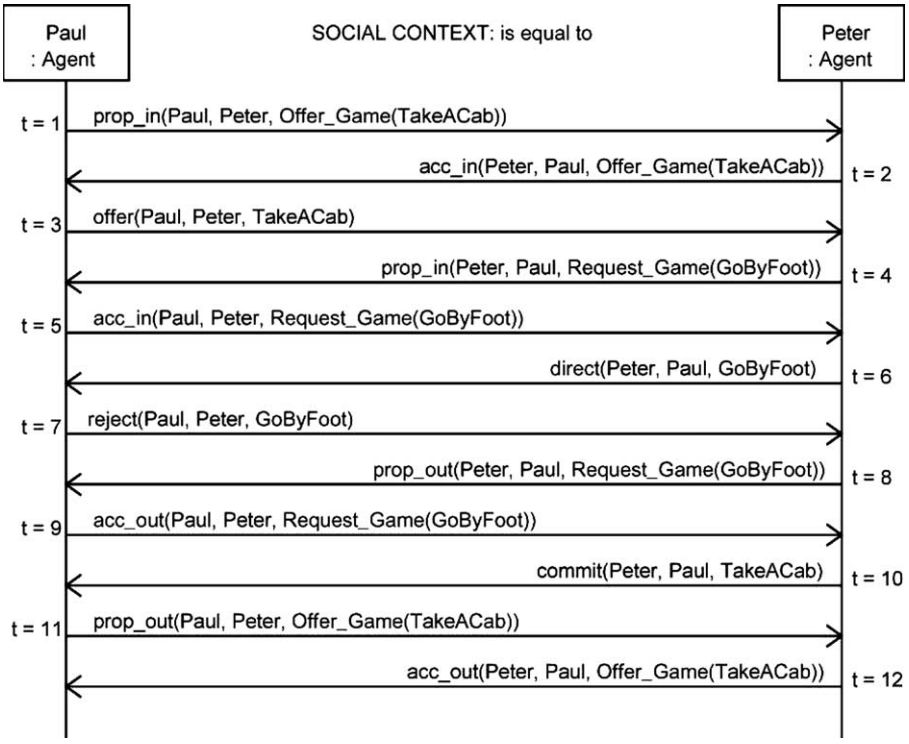

Fig. 13. Dialogue between Paul and Peter.

agenda content, it is somehow difficult to have a clear idea of how our coherence driven BDI agent behaves in a general way. Fig. 14 sums up the emerging behavior of our agents based on their expected utility calculus. At each step of his reasoning, the agent will try to find the element change that would maximise his coherence based expected utility function (involving coherence gain as well as some possible costs).

If this element is an intention, then we have a case of the attitude change. If the intention is an accepted one, he will refuse it and call underlying BDI architecture for redeliberation. If the intention is a rejected one, he will accept it (some talk about intention adoption) and call the underlying agent architecture in order to propagate the attitude change. This is the case where changing the outside world (the social commitment layer in our case) is too costful for the agent, in this case, he will try to change his private cognitions in order to restore coherence with the environment.

If this element is an already accepted commitment (issued from previous dialogue or from system's conventions) the agent will have to renege or withdraw (depending if he is the debtor or the creditor of this particular commitment). In that case the agent has to cope with the associated sanctions but this is not a surprise since those were taken into account in the expected utility calculus as having some costs.

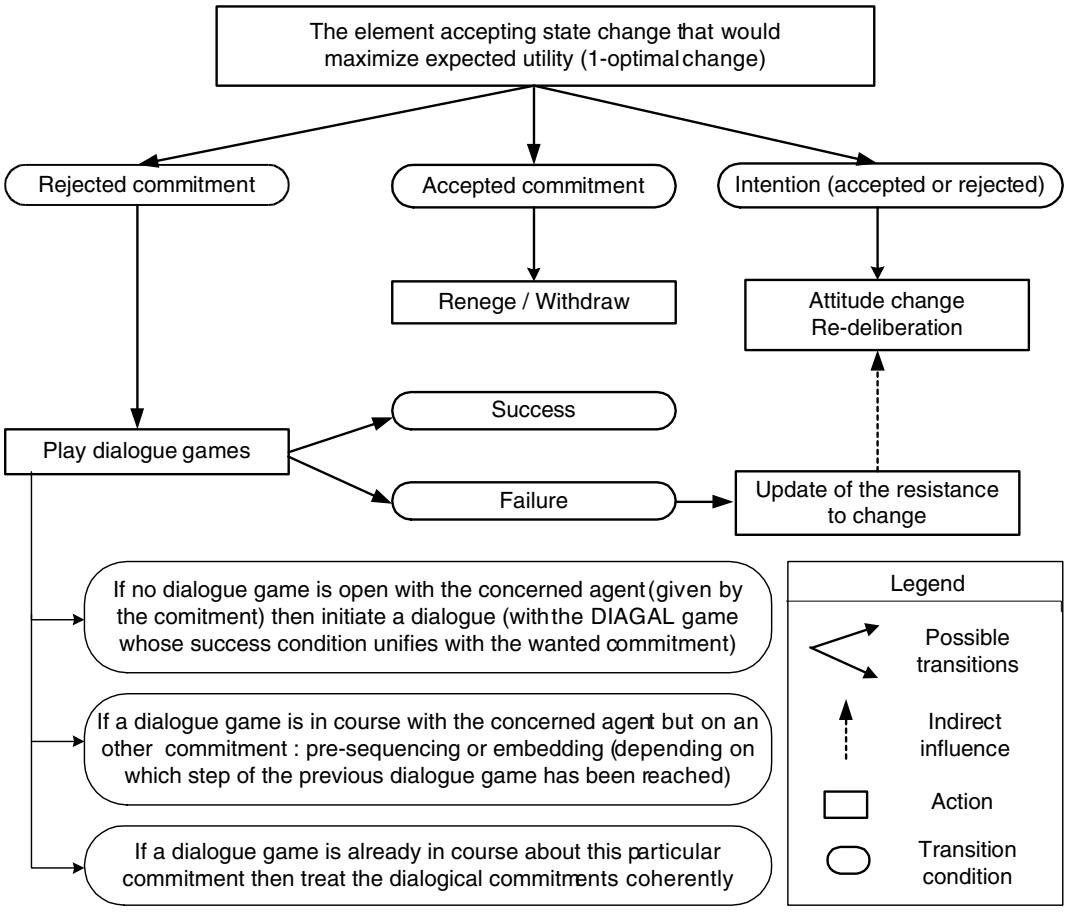

Fig. 14. Resulting decision tree of a BDI agent using cognitive coherence for his communicational behavior.

If this element is a rejected commitment then we should consider three cases:

- If no dialogue game is open with the concerned agent (given by the commitment), then the agent will initiate a dialogue (with the DIAGAL game whose success condition unifies with the wanted commitment).
- If a dialogue game is in course with the concerned agent but on another commitment: the agent will initiate a pre-sequencing or an embedding (depending on which step of the previous dialogue games has been reached).
- If a dialogue game is already open about this particular commitment then the agent will treat his dialogical commitments (coherently).

Finally, each dialogue game will end with a success or a failure for the initiator agent depending if they reach the success or the failure conditions. In the case of success, the commitment is accepted with its attached sanctions. In the case of failure, the initiator agent should update the resistance to change of this explicitly rejected commitment in order to take into account that failure. This update results in increasing the resistance to change so that he will tend not to repeat the same moves in the next reasoning steps: he will try to get a weaker commitment accepted or will try to get the same one but with another agent (e.g., if your father does not want, try with your mother) or else will select an attitude change (e.g., if your mother does not either, change your mind). Notice that following Rao and Georgeff (1992) vocabulary, the amount of the increase in resistance to change will lead to the different individual commitment strategies: if this increase in the resistance to change is null the agent will be blindly committed to trying to get this social commitment accepted, if the increase is drastically important this individual commitment will be an influenceable open-minded one and in between, we would get a wild range of single minded commitment strategies toward the wanted social commitment. Notice that those commitment strategies could dynamically depend on: the incoherence magnitude, the dialogue topic, the partner, etc.

9. Related work and prospects

As previously stated, little work has been done on the cognitive aspects of agent communication pragmatics. We have already mentioned the "intentional pragmatics" approaches which are of historical importance, but we have seen that they are not suited to social commitment based interactions. For the sake of completeness, we could otherwise mention some more technical work using decision theory to reason about commitments and sanctions in MAS due to Excelente-Toledo, Bourne, and Jennings (2001). We should also notice that the cognitive dissonance theory from Festinger was first introduced as an AI tool for truth maintenance systems (TMS) by Schwartz (2001).

Due to both its wide coverage and genericness, our approach seems to be a good candidate to provide foundations for other systems. This is probably why our theory has been used by Sansonnet and Valencia (2003a) from the LIMSI laboratory (in Paris) to automate non-task-oriented agents communicational behaviors. The authors have extended it for social simulation (Sansonnet & Valencia, 2003b). Another use of our pragmatics framework is due to the IRIT laboratory in Toulouse where the Graal team, working on natural language processing, has used our approach in order to develop their own dialogue games simulator (Adam, 2003). We refer the interested reader to the publications devoted to those particular works and we now turn to a brief description of some of our prospects.

Although the architecture presented in this paper is quite satisfying, 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, social control and social relations dynamic management,[16] (3) extend the current framework with argumentation seen as constraints propagation allowing agents to reason

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

about others' cognitive constraints and thus taking them into account, introducing cooperation.[17]

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 agents architecture. This would permit taking more advantage of the power of coherentist approaches (Thagard, 2000), using the powerful hybrid symbolic-connexionist formalisms attached with them. A preparatory step would be to extend the actual formalism by allowing using more complex domains than the mere accepted or rejected state for cognitions and using more complex constraints than the simple positive and negative ones. This could be done by using some classical constraint satisfaction algorithms and would increase the expressive power of our formalism.

10. Conclusion

In this article, we presented an agent communication pragmatics theory: the cognitive coherence theory. Proposed as a new layer above classical cognitive agent architecture, it supplies theoretical and practical elements for automating agent communication. The incoherence and utility measures defined within the cognitive coherence framework provide the necessary mechanisms to answer (even partially) the following questions which are usually poorly treated in the AI and MAS literature:

(1) *Why should agents 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.[18]

(2) *When should an agent take a dialogue initiative, on which subject and with whom?* An agent engages in a dialogue when an incoherence appears or when an incoherence magnitude exceeds a fixed level[19] 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 (Section 4.3), the theory is generic enough to be applied to any conventional communicational framework. In Section 7, 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 Section 4.6, 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 minus the cost of the dialogue moves. Furthermore, we define the expected utility of a dialogue as the incoherence reduction in the 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, how to pursue it?* The dialogue stops when the incoherence is reduced or, either it continues with a structuration according to the incoherence reductions chain or it stops because things cannot be re-

[17] It is worth noticing that the present framework does not allow cooperation per se but that the BDI application does, because the BDI agents holds a model of others.

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

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

discussed anymore (this case where incoherence persists can lead to attitude change depending on the way resistance to change is updated as discussed in Section 7.2).

(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 3.2).

(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[20] 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 deduce emotional attitudes 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 other 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 – have been implemented and exemplified in Section 8 as part of our exploration in applying the cognitive coherence pragmatics 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 in a general scheme for automatizing agents communicational behavior.

In addition, we stressed the importance of the explicitation phase of the dialogue which is usually forgotten and we have shown how cognitive coherence approaches imply the need to consider conversation as a common project with grounded dialogical moves.

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 through the expected utility function extends the deliberation process to take into account the social level, whereas selecting a dialogue game by unifying its success conditions with the wanted social result is part of the means-end reasoning. We also insisted on the dialogue's effect on agent's private mental states through the attitude change process. This process is activated by a kind of *reconsider*() function (see Rao & Georgeff, 1992) which has been modelled and integrated into our expected utility function and whose results depend on the chosen individual commitment strategy (which is taken into account when the resistance to change of explicitly rejected commitments are updated).

We hope that we have shown how a motivational reasoning and calculation on elements and their associated constraints allows automating agent communicational behavior. In MAS, agent autonomy is a critical point, we think that providing agents with tools to compute their dialogues utility and the dynamic of their communications themselves is a new step in that direction. This is the technical stake of our work. In the future, this type of model will perhaps enable agents designers to take agents relevant communicational behavior for granted when designing their agents, this is the

[20] Actually, this is not the only factor, as we exemplify in Section 4.7, other factors could also matter: social role, hierarchical positions, etc.

technological stake of such work in the long term range.

Finally, the cognitive coherence approach is a generic (wide spectrum) agent communication pragmatics theory well-founded over cognitive sciences results (namely, social psychology and computational philosophy of mind) and there is a scientific stake in referring back to cognitive sciences to see if our extensions and specializations could be meaningful.

Obviously, all the dimensions of our work could not be fully discussed here, but our purpose was rather to give an overview of our approach for agent communication pragmatics emphasizing its wide coverage. Because there is much more to say about it, future publications will be dedicated to more specific aspects of the cognitive coherence theory sketched here.

References

Adam, C. (2003). Structure et simulation du dialogue. Master's thesis, IRIT, Institut de Recherche en Informatique de Toulouse.

Allen, J. (1995). *Natural language understanding: Chapter 17: Defining a conversational agent* (pp. 540–576). Menlo Park, CA: The Benjamin/Cummings Publishing Compagny, Inc..

Baker, M. (1991). *Knowledge acquisition in physics and learning environments. Chapter: An analysis of cooperation and conflict in student's collabortive explanations for phenomena in mechanics*. Heidelberg, Germany: Springer.

Bates, J. (1994). The role of emotion in believable agents. *Communications of the ACM, 37*(7), 122–125.

Bratman, M. E. (1990). What is intention? In P. R. Cohen, J. L. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 15–32). Cambridge, MA: MIT Press.

Brehm, J., & Cohen, A. (1962). *Explorations in cognitive dissonance*. New York: Wiley.

Chaib-draa, B., Maudet, N., & Labrie, M. A. (2003). Diagal, a tool for analysing and modelling commitment-based dialogues between agents. In *Proceedings of Canadian AI 2003. Lecture Notes in Artificial Intelligence* (Vol. 2671, pp. 353–369).

Chaib-draa, B., & Vongkasem, L. (2000). ACL as a joint project between participants: A preliminary report. In F. Dignum & M. Greaves (Eds.), *Issues in agent communication. Lecture Notes in Artificial Intelligence* (Vol. 1916, pp. 235–248). Heidelberg, Germany: Springer.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cohen, P. R. (1996). *Survey of the state of art in natural language technology. Chapter: Discourse and dialogue* (pp. 234–241). Universitat des Saarlandes.

Cohen, P. R., & Levesque, H. J. (1990a). *Intentions in communication. Chapter: Persistence, intention and commitment*. Los Altos, CA: Morgan Kaufmann.

Cohen, P. R., & Levesque, H. J. (1990b). Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 221–256). Cambridge, MA: MIT Press.

Cohen, P. R., & Levesque, H. J. (1991). *Teamwork. Technote 504*. Menlo Park, CA: SRI International.

Cohen, P. R., & Perrault, C. R. (1979). Elements of a plan based theory of speech acts. *Cognitive Science, 3*, 177–212.

Colombetti, M. (2000). Commitment-based semantic for agent communication languages. In *1st Workshop on the history and philosophy of logic, mathematics and computation*.

Craig, R. (1983). *Conversational coherence: Form, structure and strategy*. Beverly Hills, CA: Sage.

Dastani, M., Hulstijn, J., & der Torre, L. V. (2000). Negotiation protocols and dialogue games. In *Proceedings of the Belgium/Dutch Artificial Intelligence Conference*, BNAIC 2000.

Dehais, F., & Pasquier, P. (2000). Approche Générique du Conflit. In D. L. Scapin & E. Vergisson (Eds.), *ErgoIHM 2000*. France, ESTIA: Biarritz.

Dessalles, J.-L. (1998). *Formal semantics and pragmatics of dialogue. Chapter: The interplay of desire and necessity in dialogue*. (Vol. TWLT-13, pp. 89–97). Enschede: University of Twente.

Dignum, F., & Greaves, M. (2000). Issues in agent communication: An introduction. In F. Dignum & M. Greaves (Eds.), *Issues in agent communication. Lecture Notes in Artificial Intelligence* (Vol. 1916, pp. 1–16). Heidelberg, Germany: Springer.

Erwin, P. (2001). *Attitudes and persuasion*. Psychology Press.

Excelente-Toledo, C. B., Bourne, R. A., & Jennings, N. R. (2001). Reasoning about commitments and penalties for coordination between autonomous agents. In J. P. Müller, E. Andre, S. Sen, & C. Frasson (Eds.), *Proceedings of the fifth international conference on autonomous agents* (pp. 131–138). Montreal, Canada: ACM Press.

Festinger, L. (1957). *A theory of cognitive dissonance*. CA: Stanford University Press.

Finin, T., & Fritzson, R. (1994). Specification of the KQML agent communication language plus example agent policies and architectures. Technical Report, DARPA Knowledge Sharing Initiative External Interface Working Group.

Finin, T., & Labrou, Y. (1998). Semantics for an ACL. *Agents, architecture and languages*. Heidelberg, Germany, Berlin: Springer.

Flores, R., & Kremer, R. (2001). Bringing coherence to agent conversation. In M. Wooldridge, P. Ciancarini, & G. Weiss (Eds.), *Agent-oriented software engineering II. Lecture Notes in Computer Science* (Vol. 2222, pp. 50–67). Berlin: Springer.

Flores, R., & Kremer, R. C. (2002). A formal theory for agent conversations for actions. *Computational Intelligence, 18*(2).

FIPA [Fondation for Intelligent Physical Agents] (2002). FIPA ACL message structure specification. Available from http://www.FIPA.org.

Grice, H. P. (1957). Meaning. *Philosophical Review, 66*, 377–388.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics, 12*, 175–204.

Grosz, B. J., & Kraus, S. (1996). Collaborative plans for complex group action. *Artificial Intelligence, 86*, 269–357.

Hamblin, C. (1970). *Fallacies*. London, UK: Methuen.

Harmon-Jones, E., & Mills, J. (Eds.). (1999). *Cognitive dissonance: Progress on a pivotal theory in social psychology*. American Psychological Association.

Howden, N., Rönnquist, R., Hodgson, A., & Lucas, A. (2001). Jack intelligent agents: Summary of an agent infrastructure. In *5th International conference on autonomous agents*.

Keefe, J. (1991). *Cognitive dissonance. Chapter: Persuasion: Theory and research* (pp. 61–78). Newbury Park,CA: Sage.

Littlejohn, S. (1992). *Theories of human communication*. Waldsworth Publishing Company.

Maudet, N. (2001). Modéliser les conventions des interactions langagières: la contribution des jeux de dialogue. PhD thesis, Université Paul Sabatier, Toulouse, France.

Maudet, N. (2003). Negociating dialogue games. *Journal of Autonoumous Agents and Multi-agent Systems, 7*(2), 229–233.

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

McBurney, P., Parsons, S., & Wooldridge, M. (2002). Desiderata for agent argumentation protocols. In *Proceedings of the first international conference on autonomous agents and multi-agents, AAMAS'01* (pp. 402–409). Bologna, Italy: ACM Press.

Moulin, B. (1997). *Agent and multi-agent systems. Chapter: The social dimension of interactions in multi-agent systems. Lecture notes in artificial intelligence (LNAI)* (Vol. 1441, pp. 169–222). Berlin: Springer.

Pasquier, P. (2002). La cohérence cognitive comme fondement pour la pragmatique des communications agents. PhD proposal, Computer Science and Software Engineering Department, Laval University. Available from http://www.damas.ift.ulaval.ca/~pasquier/publications.html.

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

Pasquier, P., & Chaib-draa, B. (2003a). The cognitive coherence approach for agent communication pragmatics. In *Proceedings of the second international joint conference on autonomous agents and multiagents systems, AAMAS'03* (pp. 544–552). ACM Press.

Pasquier, P., & Chaib-draa, B. (2003b). Engagements, intentions et jeux de dialogue. In *Actes des Secondes Journées Francophones des Modèles Formels de l'Interaction MFI'03*, Cépaduès (pp. 289–294).

Pasquier, P., & Chaib-draa, B. (2005). Modèles de dialogue entre agent: un état de l'art. In *Cognito: Cahiers Romans de Sciences Cognitive* (to appear).

Pasquier, P., Bergeron, M., & Chaib-draa, B. (2004a). DIAGAL: A Generic ACL for Open Systems. In M.-P. Gleizes, A. Omicini, & F. Zambonelli (Eds.), *Proceedings of the fifth international workshop engineering societies in the agents world (ESAW). Lecture Notes in Artificial Intelligence (LNAI)* (Vol. 3451, pp. 139–152). Berlin: Springer.

Pasquier, P., Flores, R., & Chaib-draa, B. (2004b). Modelling flexible social commitments and their enforcement. In M.-P. Gleizes, A. Omicini, & F. Zambonelli (Eds.), *Proceedings of the fifth international workshop engineering societies in the agents world (ESAW). Lecture Notes in Artificial Intelligence (LNAI)* (Vol. 3451, pp. 153–165). Berlin: Springer.

Rao, A. S., & Georgeff, M. P. (1992). An abstract architecture for rational agents. In C. Rich, W. Swartout, & B. Nebel (Eds.), *Proceedings of knowledge representation and reasoning (KR&R-92)* (pp. 439–449).

Rao, A. S., & Georgeff, M. (1995). BDI agents: from theory to practice. In *Proceedings of the first international conference on multi-agent systems (ICMAS-95)*, San Francisco, CA (pp. 312–319).

Reed, C. (1998). Dialogue frames in agent communication. In *Proceedings of the third international conference on multi-agent systems (ICMAS)*.

Reed, C., & Long, D. (1997). Collaboration, cooperation and dialogue classification. In *Proceedings of the International Joint Conference in Artificial Intelligence, IJCAI'97*, Nagoya, Japan. Morgan Kaufmann.

Sansonnet, J., & Valencia, E. (2003a). Dialogue between non-task oriented agents. In *The 4th workshop on agent based simulation, ABS 04*, Montpelier, France. Available from http://www.limsi.fr/Individu/jps/research/buzz/buzz.htm.

Sansonnet, J., & Valencia, E. (2003b). Agents informationnels pour l'étude expérimentale de concepts de socio-cognition: vers une approche agent de la socio informatique. In *Journées francophones des systèmes multiagents, JFSMA'03*, Hamamet, Tunisia.

Schwartz, P. (2001). Truth maintenance with cognitive dissonance. Technical Report, University of Maryland at College Park.

Searle, J. R., & Vanderveken, D. (1985). *Foundations of illocutionary logic*. New York: Cambridge University Press.

Searle, J. R. (1990). Collective intentions and actions. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 401–416). Cambridge, MA: MIT Press.

Singh, M. P. (1991). Social and psychological commitments in multiagent systems. In *AAAI fall symposium on knowledge and action at social and organizational levels*, *Lecture Notes in Artificial Intelligence (LNAI)* (Vol. 1916, pp. 104–106).

Singh, M. P. (1998). Agent communication languages: rethinking the principles. *IEEE Computer, 12*(31), 40–47.

Singh, M. (2000). A social semantics for agent communication languages. In F. Dignum & M. Greaves (Eds.), *Issues in agent communication* (pp. 31–45). Heidelberg, Germany: Springer.

Sperber, D., & Wilson, D. (1986). *Relevance*. Cambridge, MA: Harvard University Press.

Sun, R. (1997). *Connectionist–Symbolic integration. Chapter: An introduction to hybrid connectionist–symbolic models*. Lawrence Erlbaum Associates.

Thagard, P., & Verbeurgt, K. (1998). Coherence as constraint satisfaction. *Cognitive Science, 22*, 1–24.

Thagard, P. (2000). *Coherence in thought and action*. Cambridge, MA: MIT Press.

Traum, D. R. (1994). A computational theory of grounding in natural language conversation. PhD thesis, Department of Computer Sciences, University of Rochester.

Tuomela, R., & Miller, K. (1988). We-intentions. *Philosophical Studies, 53*, 367–389.

von Wright, G. (1980). *Freedom and determination*. Amsterdam: North-Holland.

Walton, D., & Krabbe, E. (1995). *Commitment in dialogue*. State University of New York Press.

Wickland, R., & Brehm, J. (1976). *Perspectives on cognitive dissonance*. New York: Halsted Press.

Wooldridge, M. (2001). *An introduction to multiagent systems*. New York: Wiley.