Integrating Social Commitment-Based Communication in Cognitive Agent Modeling

Philippe Pasquier[1] and Brahim Chaib-draa[2]

[1] AgentLab, University of Melbourne, Department of Information Systems, Australia.
[2] DAMAS lab., Laval University, Computer Science Department, Canada.
{pasquier,chaib}@iad.ift.ulaval.ca

Abstract. In this paper, we extend the classical BDI architecture for the treatment of social commitments based communication by: (1) linking social commitments and individual intentions, (2) providing a model of the cognitive aspect of communication pragmatics in order to automatize social commitment based communication. In particular, we introduce a general decision-making process leading to attitude change in the appropriate cases.

1 Introduction

Cognitive agent modelings rest on the isolation and formalization of private mental states such as beliefs, desires and intentions exemplified by the classic BDI [Beliefs, Desires and Intentions] model. However, social commitments as a way to capture interagent dependencies has founded improved agent communication frameworks.

In this paper, we try to narrow the gap between those two paradigms by proposing an extension of the classic BDI agent model (Section 2), enabling the resulting deliberative-normative agent to communicate using an agent communication language based on the manipulation of social commitments: the DIAlogues Games Agent Language (DIAGAL) (Section 3). This extension involves: refining intention typology, linking individual intention with social commitments (Section 4), and advancing a model of the cognitive aspects of pragmatics (Section 6) that leads to communication moves or attitude change (Section 5).

2 The classic BDI model

Various formulations of the BDI model can be found. The model has been expressed in multimodal logics [18, 22], in first order specification languages [12] or in procedural/algorythmic notation [27, Chapter 4]. In this paper, we will focus on the procedural specifications.

BDI architecture rests on two main processes: deliberation and means-end reasoning. Deliberation is the process by which an agent generates its intentions on the basis of its beliefs and desires, while means-end reasoning consists in planning a sequence of actions to execute as an attempt at satisfying its intentions. The BDI control algorithm (presented in Figure 1) makes a compromise between deliberation (a time consuming cognitive activity), means-end reasoning and acting activities through the Reconsider() function.

Procedure BDICycle(B_0,I_0)

```
 1: Inputs: B_0, set of initial beliefs;
            I_0, set of initially accepted intentions;
 2: Outputs: none, this is not a function!
 3: Local: B := B_0, object that store the agent's beliefs;
           I := I_0, object that stores the agent's intentions;
           D, object that stores the agent's desires;
           List ρ, stores both internal and external percepts;
           List π := null, current plan, sequence of actions;
 4: Body:
 5: while true do
 6:    Get new percepts ρ;
 7:    Update B on the basis of ρ;
 8:    if Reconsider(I,B) then
 9:       D := Options(B,I);
10:       I := Deliberate(B,D,I); // deliberate if necessary
11:    end if
12:    if Empty(π) or Succeeded(I,B) or Impossible(I,B)) then
13:       π := Plan(B,I); // replan if necessary
14:    else
15:       α := Head(π);
16:       Execute(α); // execute an action
17:       π := Tail(π);
18:    end if
19: end while
```

Fig. 1. BDI agent's control loop.

At each cycle of the algorithm, the BDI agent updates his beliefs according to its percepts (lines 6 and 7). If necessary (according to the boolean function Reconsider(), line 8), the agent (re)deliberates in order to update his desires and intentions (line 9 et 10). Then, if the current plan is empty or has become invalid or if the pursued intention has been achieved, has become impossible or has changed (line 11), the agent (re)plans (line 12). Otherwise (if all the preceding conditions are false), the agent executes an action from the current plan (lines 14-16). Notice that this action can be of a complex type.

As stated in [7], intentions are choices to which the agent commits. One of the main characteristics of individual intentions is that they are associated with what has been called an individual action commitment for which resources have been allocated [2, 25]. It means that when an agent has accepted an intention, he is individually committed to achieving particular actions as an attempt to reach the wanted state (described by the intention). This individual commitment should not be confused with social commitments.

These individual commitments are not represented explicitly and it's the intention reconsideration process that ensures intentions' temporal persistence. The mechanism used by an agent in order to decide when and how to reject a formerly accepted intention is called *individual commitment strategies*. One usually distinguishes three main individual commitment strategies [18, 22]:

- *Blind commitment (fanatical)*: agent continues to maintain intention until it has been achieved;

- *Single-minded commitment*: agent will continue to maintain intention until it has been achieved or it is impossible to achieve;
- *Open-minded commitment*: agent will maintain intention as long at it believes it is possible.

Communication in the BDI model In philosophy of language, Grice introduced the fundamental link that lies between intention and communication through the definition of non-natural meaning. This accounts for the fact that literal meaning of a statement does not cover its whole meaning. According to Levinson's formulation [13], non natural meaning can be defined as follows. The locutor A wanted to say z by uttering e, if and only if:

1. A has the intention that e yield to the effect z on B;
2. A has the intention that the previous intention will be achieved by B through its recognizing of it.

Intention is involved twice in that definition considering the locutor's prior intention as well as his communicative intention, i.e. that the interlocutor recognizes his prior intention and react cooperatively. Consequently, strong *cooperativity* and *sincerity* assumptions are assumed in agent models that use these mentalistic trends. For example, if the agent A wants to know if p holds and believes that B has these pieces of information, he will ask B and hope that B will recognize its intention and answer cooperatively and sincerely according to his own knowledge. Even for assertive speech acts, cooperativeness is present. For example, an assertion involves a belief change as a cooperative answer to it in a context where sincerity is trusted.

Furthermore, computational complexity of the multi-modal logics used for specifying speech-act based ACLs with mentalistic semantics forbids their use by MAS designers (see [8, 16] for discussions on that subject). In practical systems, the communicative behavior of an agent is designed as a simplified reification of the afore-mentioned concepts. For example, in the JACK-BDI agent frameworks [11] (based on dMARS [12]), the agent's communicative behavior is part of its means-end reasoning, which is implanted as follows. Each plan consists of: an invocation condition, which is the event that the plan responds to, a context condition, stating conditions under which to use the plan, and a body that specifies a sequence of actions or subgoals to achieve. Each intention raises a particular internal event type (goal events). Planning consists in selecting one plan with that event as the triggering condition and with a context condition that is believed true. The choice between competitive plans is generally based on meta-plans or hardwired strategies (for efficiency). For example, in the JACK-BDI agent architecture the first eligible plan is chosen by default.

In that setting, dialogical actions are hard-coded in plans as other actions. Dialogue initiative is hard-coded as the primitive action of sending an ACL message that initiates a dialogue. Messages received from other agents are interpreted as external events of a particular type (message events) that are treated in the event queue by updating beliefs and trigger the appropriate plan in order to pursue (or cancel) the conversation.

Social commitment based communication frameworks allow leaving down these cooperativity and sincerity assumptions by providing a treatment of the social aspects of communication that is absent in previously proposed purely mentalistic approaches. The next section will both introduce social commitment based communication and discuss this point.

3 Social commitment based communication

Social commitment has been introduced as a first class concept to represent socially established (and grounded) interagent dependencies. In particular, social commitments can model the semantics of agents' interactions. In that context being able to cancel or modify commitments is a key feature that allows agents to reassess the consequences of past dialogues in the context of dynamic environments. This *semantical flexibility* should not be confused with the commonly considered structural flexibility of dialogues.

Since [17] discusses our modelling of flexible social commitments and their enforcement through sanctions, we simply re-introduce the basic of it here. Conceptually, commitments are oriented responsibilities contracted towards a partner or a group. Following [26], we distinguish *action commitments* from *propositional commitments*. Commitments are expressed as predicates with an arity of 6. Thus, an *accepted* action commitment takes the form:

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

meaning that agent x is committed towards agent y to α since time t, under the sanctions sets s_x and s_y. An accepted propositional commitment would have propositional content p instead α. *Rejected* commitments, meaning that x is not committed toward y to α, takes the form $\neg C(x, y, \alpha, t, s_x, s_y)$. This notation for commitments is inspired from [21], 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, agents keep track of each commitment in which they are debtor or creditor in their *agendas*, which constitutes a kind of distributed "Commitment Store".

In previous work, we proposed a DIAlogue Games Agent Language (DIAGAL) [14] for which our social commitment model offers a complete and valid operational semantics. DIAGAL dialogue games are composed of entry conditions (E), success condition (S), failure conditions (F), all expressed in terms of social commitments and dialogue rules (R) which are expressed in terms of dialogical commitments (C_g) that allow capturing the conventional level of communication. For example, here is DIAGAL's *Request* game (sanctions are avoided for the sake of clarity):

$$\begin{array}{l|l}
E_{rg} & \neg C(y, x, \alpha, t_i) \text{ and } \neg C(y, x, \neg\alpha, t_i) \\
S_{rg} & C(y, x, \alpha, t_f) \\
F_{rg} & \neg C(y, x, \alpha, t_f) \\
R_{rg} & 1)\ C_g(x, y, \mathit{request}(x, y, \alpha), t_j) \\
 & 2)\ C_g(y, x, \mathit{request}(x, y, \alpha) \Rightarrow \\
 & \quad C_g(y, x, \mathit{accept}(y, x, \alpha) | \mathit{refuse}(y, x, \alpha), t_k), t_j) \\
 & 3)\ C_g(y, x, \mathit{accept}(y, x, \alpha) \Rightarrow C(y, x, \alpha, t_f), t_j) \\
 & 4)\ C_g(y, x, \mathit{refuse}(y, x, \alpha) \Rightarrow \neg C(y, x, \alpha, t_f), t_j)
\end{array}$$

DIAGAL dialogue based communication is grounded and structured through a so-called *contextualisation game* that allows the agents to enter and leave games as well as to structure complex dialogues. All together, our model of flexible social commitment and their enforcement [17] and the DIAGAL [14] language provides a complete agent communication framework that introduces a layered model of agent communication (also formally described in [10]):

1. At the *signal level* (sometimes called attentional level): the contextualisation game allows grounding dialogue games as well as their eventual structuration;
2. At the *message level*: messages (dialogue or speech act) allow fulfilling dialogical commitments and advancing the state of opened dialogue games;
3. At the *dialogic level*: dialogue games allow advancing the state of the social layer of social commitments;
4. At the *social level*: social commitments, if they are respected (which is the case with our model of the enforcement of social commitments) advance the state of activities;
5. At the *activity level*: activities advance the state of the environment in a way that should satisfy the agents or their designers.

According to the principle of *information asymmetry*, what is said does not convey anything about what is actually believed. However, what is said socially commits the locutors toward one another. Social commitments raise action expectations and the enforcement of social commitments through various social control mechanisms take place instead of the sincerity and the cooperativeness assumptions. Social commitments, when modelled with their enforcement mechanism (as in [17]), are not necessarily sincere and don't require the agents to be cooperative. From this perspective, communication serves to coordinate the agents whether or not they are cooperative and whether or not they are sincere.

These social commitment based frameworks, enhancing the social aspects of agents' communications, entail a change of paradigm: agents do not necessarily have to reason on others' intentions anymore but rather they must reason on taken and to be taken social commitments. However, it has not been indicated how agents should dynamically use social commitment based communication and social commitments were not taken into account in previous cognitive agents theory.

In order to fill this gap, we will extend the presented BDI model by: (1) linking private cognitions with social commitments, (2) providing a model of the cognitive aspect of communication pragmatics in order to automatize social commitment based communication. In particular, we will introduce a general decision-making process leading to attitude change in the appropriate cases.

4 Linking Public Cognition and Social Commitments

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 that someone do something) and *intention that* (a proposition holds) [2]. The *intention to* relates to a particular course of action (eventually of a complex and structured type), while *intention that* refers to a propositional statement that the agent wants to became true. Intentions are either accepted ($I_A(p)$) or rejected ($\neg I_A(p)$).

In order to address communication, we will further distinguish between *internal individual intentions* and *social individual intentions*. Internal individual intentions are intentions that the agent can try to achieve alone while social individual intentions are the intentions that relate to other agents' actions. Social individual intentions are intentions concerning goals which require other agents to be worked on. More generally, any intention that is embedded in a somewhat collective activity would be considered as a social individual intention except if

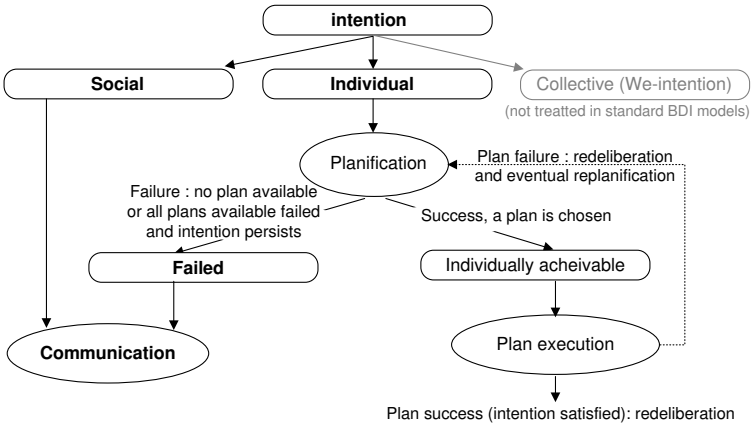

Fig. 2. Operational typology of intentions.

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 a social activity (a problem requiring action, permission or opinion of the others: commerce, exchange, joint action, delegated actions,...). A classic example is delegation where an agent A has the social intention that an agent B achieves a particular action α, $I_A(\alpha_B)$.

Among internal individual intentions, we will also consider *failed individual intentions* which are the intentions that the agent failed to find an individual plan for or for which the available plans failed. This last type matches the case where the agent faces an individual problem he cannot solve alone or he failed to solve alone.

In our framework, failed individual intentions as well as the social individual intentions will be treated through dialogue. The phase of identifying intentions involving a social dimension appears to be crucial for integrating social commitment based approaches with existing cognitive agent architectures. In our approach, all intentions that are not achievable internal intentions will be selected as such. Filtering those failed and social intentions from the other ones is achieved by selecting the intentions for which the mean-end reasoning failed. In particular, in the JACK-BDI framework, intentions that don't match any individual plans or for which all available individual plans have failed fall into those categories. Notice that this implantation implies that trying to achieve individual action (through execution of individual plans) is the prioritized behavior of the agent. Figure 2 sums up this intention typology.

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. Ideally, an accepted social commitment is the socially accepted counterpart of an accepted intention. Commitments in action are the counterparts of "intentions to" while propositional commitments are the counterparts of "intentions that". In our approach, those links are taken into account by positive and negative binary constraints

that link the agents intentions and social commitments. Positive constraints take into account the correspondence relation introduced above while negative constraints model the incompatibility relations that hold between incompatible intentions or/and social commitments.

Let's take an example to illustrate those relations. If an agent A has the accepted individual social "intention to" that another agent B achieves an action α (noted $I_A(\alpha_B)$), our links mean that the corresponding social commitment from B toward A to achieve α_B (noted $C(B, A, \alpha_B, t, s_B, s_A)$ must be socially accepted as part of this intention satisfaction. This ideal link between those two cognitions is captured with a positive constraint. For this constraint to be satisfied, both elements (the intention and the corresponding commitment) must be accepted or rejected. However, all other possibilities are also important to consider. Furthermore, incompatibility relations are modeled with negative constraints.

Those 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 [2, 25]. 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 lead to the social acceptance of their social commitments counterparts through dialogue. Those links complement Singh's previous work [20], which introduces the idea of linking individual and social commitments. Comparable links have been introduced for so-called normative-deliberative cognitive agent architecture [6, 1, 4]. In particular, following [5], the following axioms have been introduced [19]:

$$S\text{-}COMM(i, j, \tau_i) \rightarrow I_j(\tau_i),^3 \text{ and}$$
$$S\text{-}COMM(i, j, \tau_i) \rightarrow I_i(\tau_i)$$

From which, one can deduce the following theorem:
$\vdash \neg S\text{-}COMM(i, j, \tau) \vee (I_i(\tau) \wedge I_j(\tau))$, which clearly states that either the social commitment is rejected or both i and j have the intention that i achieves the action τ. This formalization is not compatible with the semantic flexibility of social commitments described in Section 3. For example, if i decides to violate or cancel the aforementioned commitment, it is probably because he does not have the corresponding intention accepted. In that case, we have the accepted commitment $S\text{-}COMM(i, j, \tau_i)$ and the rejected intention $\neg I_i(\tau_i)$ that holds which invalidates the second of the above axioms. Symmetrically, if the agent j tries to cancel the accepted commitment $S\text{-}COMM(i, j, \tau_i)$, it can be because he does not have the corresponding intention accepted. In that case, we have $S\text{-}COMM(i, j, \tau_i)$ and $\neg I_i(\tau_i)$ which invalidates the first of the above axioms. In other words, those axioms are not flexible enough to provide a good modelling of the links that lie between intentions and social commitments.

Constraints provide bidirectional and symmetric links that go behind the above mentioned axioms.[4] This is why we used constraints in order to model those links. Examples where a commitment is accepted and the corresponding intention is not or the reverse are very common and just mean that the positive constraint linking those two elements is not satisfied. As a consequence, not only those bidirectional links are more correct than the previously criticized

[3] Sometimes formulated : $S\text{-}COMM(i, j, \tau_i) \rightarrow Goal_j(Does_i(\tau))$.

[4] We refer the interested reader to [23] for a discussion about bidirectionality in cognitive modelling.

axioms but they allow for a new question to be asked. When such a constraint is not satisfied, the agent has to decide which elements' acceptance state he will try to change in order to satisfy this positive constraint: his intention or the corresponding social commitment. This is the basic question of the attitude change process. Since this notion of attitude change has not been yet modelled in the context of AI, we will introduce it here.

5 Attitude Change

In cognitive sciences, cognitions gather together all cognitive elements: perceptions, propositional attitudes such as beliefs, desires and intentions, feelings and emotional constituents as well as social commitments. From the set of all private cognitions result *attitudes* which are positive or negative psychological dispositions towards a concrete or abstract object or behavior.

For contemporary psychologists, attitudes are the main components of cognition. These are the subjective preliminary to rational action [9]. Theoretically, an agent's behavior is determined by his attitudes. The basic scheme highlighted by those researches is that beliefs (cognition) and desires (affect) lead to intentions which could lead to actual behaviors or dialogical attempts to get the corresponding social commitments depending on their nature. From another point of view, it could happen (due to hierarchies, power relations, negotiation, argumentation, persuasion dialogues,...) that an agent becomes socially committed to a counter-attitudinal course of action or proposition. In that case, *attitude change* can occur.

The links between private and public cognitions established in Section 4 allow defining the attitude change process in the way provided by cognitive psychology's classical studies [3]. Ideally, for each accepted or rejected social commitment, the corresponding intention should be accepted or rejected (respectively) in both the creditor and the debtor mental states. For example, we assume that $C(A, B, \alpha_A, t, s_A, s_B)$ holds, indicating that A is committed toward B, since time t, to achieve α_A under the sanction sets s_A and s_B. Then, A and B should ideally have the intention that A achieves α_A, noted $I(\alpha_A)$, accepted in their mental model. If, for example, A doesn't have $I(\alpha_A)$ accepted, he can: (1) revoke or violate the commitment and face the associated sanctions (2) try to modify the commitments through further dialogues or (3) he can begin an *attitude change*, i.e. adopt this intention and possibly reject incompatible ones.

6 Dialogue pragmatics

6.1 The cognitive coherence framework

All attitude theories, also called cognitive coherence theories appeal to the concept of homeostasis, i.e. the human faculty to maintain or restore some physiological or psychological constants despite the outside environment variations. All these theories share as a premise the *coherence principle* which puts coherence as the main organizing mechanism: *the individual is more satisfied with coherence than with incoherence*. The individual forms an opened system whose purpose is to maintain coherence as much as possible.

Our pragmatics theory (presented in [15]) follows from those principles by defining a formal theory of cognitive coherence. Here, elements are both failed

or social intentions and social commitments. Elements are divided in two sets: the set $\mathcal{A}$ of *accepted elements* (accepted, failed or social, intentions and socially accepted social commitments) and the set $\mathcal{R}$ of *rejected elements* (rejected social intentions and socially rejected social commitments). 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 like the correspondence relation described in Section 4.
- *Negative constraints*: negative constraints are inferred from negative relations like mutual exclusion and incompatibility relations considered in Section 4.

These constraints can be satisfied or not: a positive constraint is satisfied if and only if the two elements that it binds are both accepted or both rejected. On the contrary, a negative constraint is satisfied if and only if one of the two elements that it binds is accepted and the other one rejected. For each of these constraints a weight reflecting the importance of the underlying relation can be attributed.[5]

Given a partition of elements among $\mathcal{A}$ and $\mathcal{R}$, one can measure the *coherence degree*, $\mathcal{C}(\mathcal{E})$, of a non-empty set of elements, $\mathcal{E}$, 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 weight of concerned constraints. The general coherence problem is then to find a partition between the set of accepted elements $\mathcal{A}$ and the set of rejected elements $\mathcal{R}$ that maximize cognitive coherence. It is a constraint optimization problem shown to be NP-complete by [24]. In our case the coherence problem is solved in an iterative manner by the local search algorithm.

6.2 Local search algorithm

Decision theories as well as micro-economical theories define utility as a property of some valuation functions. A function is a utility function if and only if it reflects the agent preferences. In the cognitive coherence theory, according to the afore-mentioned coherence principle, coherence is preferred to incoherence.

In order to try to maximize its coherence, at each step of his pragmatics' reasoning, an agent will search for a cognition's acceptance state change which maximizes the coherence increase, taking into account the resistance to change of that cognition (technically a 1-optimal move). If this attitude is a commitment, the agent will attempt to change it through dialogue and if it is an intention, it will be changed through attitude change. In that last case, we call the underlying architecture of the agent to spread the attitude change and re-deliberate.

In our implementation, an agent determines which is the most useful cognition's acceptance state change by exploring all states reachable from its current state and selects the cognition which can *in case of a successful change* be the most useful to change. A state is said to be reachable if it can be obtained from the current state by modifying only one cognition. A notion of cost has been

[5] This is a way of prioritizing some cognitive constraints as is done in the BOID architecture [4].

introduced to advocate for the fact that all cognitions cannot be equally modified. All explored states are so evaluated through an *expected utility function*, g, expressed as below:

$$g(ExploredState) = \mathcal{C}(exploredState) - \mathcal{C}(currentState) - r(cognitionChanged)$$

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

1. if *cognitionChanged* is an intention, its cost of change equals its resistance to change that reflects the underlying individual commitment strength;
2. if *cognitionChanged* is a rejected commitment, its cost of change equals its resistance to change, which is initially low but which could be increased at each unfruitful attempt to establish it;
3. if *cognitionChanged* is an accepted commitment, its cost of change is increased by its associated sanctions (which could be null, positive or negative).

The local search algorithm is an informed breath first search algorithm with the afore-mentioned expected utility measure as its greedy heuristics. We don't have a proof of correctness of this algorithm in regards to the general coherence problem but, as [24] (who used it in another context), it was shown to be optimal on tested examples.

6.3 Pragmatic Treatment Algorithm

The dialogic behavior of the agent is based on his cognitive coherence calculus involving failed and social intentions as well as social commitments. Social commitments and their state are memorized in the agent agenda which is maintained by the DIAGAL dialogue manager. Figure 3 presents the agent pragmatic treatment algorithm that integrates pragmatics reasoning and social commitments' treatment.

As seen in Section 3, we distinguish extra-dialogical commitments (assigned to a List on line 5) from dialogical commitments (assigned line 6). Dialogical commitments result from dialogue games' rules as well as from the contextualisation game. Extra-dialogical commitments are processed by TreatCommitments() (line 8) which consists in updating the agent representations of commitments by taking into account dialogical as well as extra-dialogical action of the agents that has been reported by the agent's dialogue manager in the agenda. Three cases are then distinguished :

1. *dialogue initiative* : there is no active dialogic commitment in the agenda and the *initiate* boolean is true (test, line 9), which means that the underlying BDI control loop just called the Pragmatic treatment algorithm. The InitiateDialogue() procedure is called (line 11);
2. *ending of a dialogue* : there is no more active dialogic commitment in the agenda and the *initiate* boolean is false (test, line 13), which means that the dialog segment is finished. The underlying BDI control loop is called again (ModifiedBDICycle(), line 14);
3. *pursuing a dialogue* : there are some dialogical commitments to process, the TreatDialogCommitment() (line 16) procedure is called.

Procedure CommunicationPragmatics(*initiate*)

```
 1: Inputs: initiate, boolean variable (true when called by the
            underlying BDI architecture, false otherwise)
 2: Outputs: none, this is not a function!
 3: Global: agenda, object that stores the agent's agenda
 4: Local:
 5: List commitments:=agenda.GetCommitments();
 6: List dialogCommitments:=agenda.GetDialogCommitments();
 7: Body:
 8: TreatCommitments(commitments);
 9: if dialogCommitments.IsEmpty() and initiate=true then
10:    initiate:=false;
11:    InitiateDialogue(); // initiate a dialogue
12: else
13:    if dialogCommitments.IsEmpty() and initiate=false then
14:       ModifiedBDICycle(); // dialog finished
15:    else
16:       // pursue a dialogue
17:       TreatDialogCommitments(dialogCommitments);
18:    end if
19: end if
```

Fig. 3. Pragmatic treatment algorithm.

In order to *initiate a dialogue*, InitiateDialogue(), generates the intentions and commitments network according to the principles of representation enunciated and argued in Section 4. Then, the local search algorithm is called and elements' acceptance states are changed until a social commitment is encountered and a dialogue is initiated as an attempt to realize the desired change.[6] The appropriate DIAGAL game is chosen by unifying *currentState* and the games entry conditions and *exploredState* with the success conditions of the game (see [14] for details). The different fields of the commitment indicate the partner and the subject of the dialogue.

In order to *pursue a dialogue*, TreatDialogCommitments(), consists in treating the remaining dialogical commitments. This is done by evaluating the consequences of all the outcomes allowed by the current dialogue games rules on the cognitive coherence. The resulting choice utility is compared to the local search choice utility. If the modification allowed by the current dialogue game is less usefull than the one proposed by local search, then the agent will imbricate a subjectively more appropriate sub-dialogue game.

In case a *dialogue ended*, control is given back to the underlying BDI control loop through the ModifiedBDICycle() (line 14) procedure call. The modified BDI control loop will take into account the eventual partial or complete attitude change and will deliberate again eventually generating new intentions that will be treated according to their nature as indicated by the algorithm of Figure 4.

Finally, the CommunicationPragmatics() procedure is called each time:

[6] Notice that the local search can return nothing (e.g., if coherence is already maximal).

Procedure ModifiedBDICycle(B_0, I_0)

```
1: Inputs: B_0, set of initial beliefs;
           I_0, set of initial intentions;
           Those inputs are optional (used for the first call)
2: Outputs: none, this is not a function!
3: Global: B := B_0, object that stores the agent's beliefs;
           I := I_0, stores the agent's accepted intentions;
           I_s, stores the agent's social or failed intentions;
           D, object that stores the agent's desires;
           List ρ, stores both internal and external percepts;
           List π := null, current plan, sequence of actions;
4: Body:
5: while true do
6:    ρ.GetNewPercepts(); // get new percepts ρ
7:    B.Update(ρ); // update B on the basis of ρ
8:    if Reconsider(I,B) then
9:       D := Options(B,I);
10:      I := Deliberate(B,D,I); // deliberate if necessary
11:   end if
12:   if Empty(π) or Succeeded(I,B) or Impossible(I,B)) then
13:      π := Plan(B,I); // replan if necessary
14:      I_s := Filter(B,I); // assign failed or social intentions
15:   else
16:      α := Head(π);
17:      Execute(α); // execute an action
18:      π := Tail(π);
19:   end if
20:   if agenda.Modified()=true then
21:      CommunicationPragmatics(false); // pursue a dialogue or answer a new dia-
         logue offer
22:   end if
23:   if not Empty(I_s) then
24:      CommunicationPragmatics(true); // initiate a dialogue
25:   end if
26: end while
```

Fig. 4. Modified BDI control loop

- the underlying BDI control loop deliberation produces either social or failed intentions that the agent cannot fulfill by itself (and thus need to communicate).
- the DIAGAL dialogue manager modifies the agent agenda and this modification is not the fulfillment or violation of an extra-dialogical commitment (which are taken into account as specified in [17]). This ensures that: (1) the agent executes the CommunicationPragmatics() algorithm until all ongoing dialogue segments are closed and (2) the agent treats dialogues initiated by other agents.

7 Examples

Returning to the example of delegation, suppose the modified BDI control loop of an agent A just generated the intention that B achieves an action α ($I_A(\alpha_B)$). This intention is filtered (line 14) as a social intention and CommunicationPragmatics() (line 24) is called which in turn calls the InitiateDialogue() procedure which produces the coherence framework of Figure 5,a. In these examples, we assume that constraints' weights are unitary and that elements rejected by default have a lower default cost of change (0.05) than accepted elements' default resistance to change (which is 0.2, plus the eventual associated sanctions strength). Update, reified by an increased by 0.2 of this resistance to change, occurs at each attempt of change (according to Section 5).

The local search algorithm returns that the best change would be to have a social commitment from B to A to $achieve(\alpha)$ accepted (as indicated by the decision tree of Figure 5,b). The appropriate DIAGAL game is the *Request* game, that is proposed by A through the contextualisation game. Suppose B refuses A's request, the resistance to change of the still rejected commitment would be updated and redeliberation will occur. If B accepts, the social commitment would be marked as socially accepted and the enforcement mechanism would be activated in order to foster its satisfaction.

In a more rich setting involving three agents, suppose that A is already committed toward a third agent F not to achieve β and has the corresponding intention accepted when an agent B orders him to achieve β (one can suppose that there is a permanent commitment to accept B's request because of his authority position, ...). Despite the fact that A has the intention to achieve β rejected, the counter attitudinal commitment toward B to do so is accepted. This situation is presented by the coherence framework of Figure 5,c. The decision tree of Figure 5,d indicates that an attitude change occurred. Following our

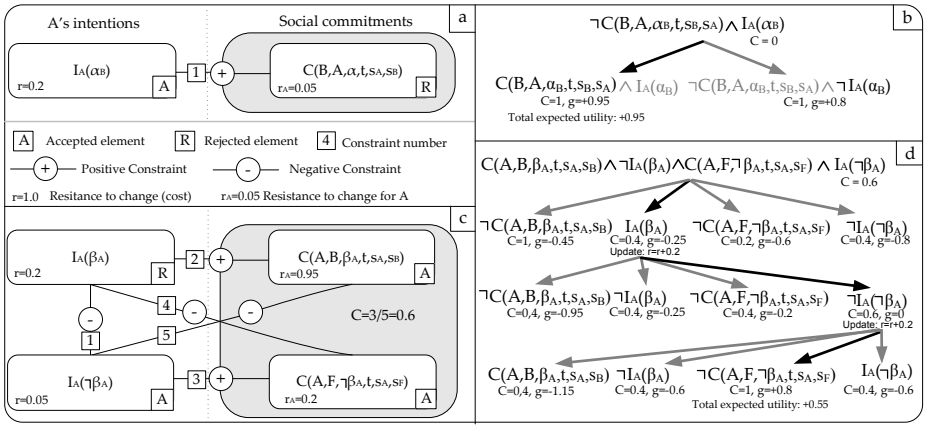

Fig. 5. Parts b and d indicate A's reasoning as computed by the local search algorithm from the states described by in parts a and c respectively. For each reachable state, the cognitive coherence and expected utility measures are indicated. The black path indicates the change(s) returned by the local search algorithm (presented section 6.2).

algorithms, agent A has rejected the intention not to achieve β and accepted the intention to achieve β and is now about to initiate a dialogue with F in order to cancel the previously accepted social commitment toward him.

Notice that the choice of the default resistances to change and update rules extends individual commitment strategies (presented in Section 1). In the proposed approach, intention persistence not only depends on the chosen default resistance to change (the higher it is, the more fanatic the agent is) but also on accepted commitments' resistance to change (reflecting sanctions and rewards). This models social pressure and allows to introduce the concept of attitude change that is central for the study and modelling of agent behavior changes.

8 Conclusion

In this paper, we unify – both at the theoretical (Section 4) and practical level (Section 5) – two important trends in MAS modelling: cognitive agents based on BDI models and social commitment based communication (a model of flexible social commitment and their enforcement [17] and the DIAGAL agent language [14]). Note that the resulting framework automatizes the agent pragmatic reasoning and communication behavior by giving him tools to measure himself the expected utility of possible communicative behaviors. The proposed model rests on solid cognitive sciences' results that allow to take into account the motivational aspects of agent communication. This approach models the persuasive dimension eventually present in all communications by reifying attitude change when necessary.

Notice that the proposed approach cumulates the advantages of past contributions. Resulting agents can be used in an open system (as long as the other agents use DIAGAL), no sincerity (of the others) is assumed, no hard-coded cooperation is needed. Notice that our pragmatic coherence approach includes the reasoning on sanctions (taken into account in the expected utility function) so that the chosen punishment strategy influences agent behavior as discussed in [14]. This is thus a major improvement over the hand written communication behavior of classical agent implantation (as those described in Section 2).

References

1. G. Boella and L. Lesmo. *Social Order in Multi-Agent Systems*, chapter Deliberative normative agents, pages 85–110. Kluwer Academic, 2001.
2. M. E. Bratman. What is intention? In P. R. Cohen, J. L. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 15–32. The MIT Press, 1990.
3. J. Brehm and A. Cohen. *Explorations in Cognitive Dissonance*. John Wiley and Sons, inc, 1962.
4. J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. Van der Torre. The BOID architecture: Conflicts between beliefs, obligations, intention and desires. In *Proceedings of the Fifth International Conference on Autonomous Agent*, pages 9–16. ACM Press, 2001.
5. C. Castelfranchi. Commitments: from individual intentions to groups and organizations. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 41–48, San Francisco, CA, USA, 1995.
6. C. Castelfranchi, F. Dignum, C. Jonker, and J. Treur. Deliberative normative agents: Principles and architecture. In *Intelligent Agents V, Proceedings of the Internationnal Workshop on Agent Theories, Architectures, and Languages (ATAL)*,

volume 1757 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 364–378. Springer-Verlag, 1999.
7. P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42:213–261, 1990.
8. F. Dignum and M. Greaves. Issues in agent communication: An introduction. In F Dignum and M. Greaves, editors, *Issues in Agent Communication*, volume 1916 of *LNAI*, pages 1–16. Springer-Verlag, 2000.
9. P. Erwin. *Attitudes and Persuasion*. Psychology Press, 2001.
10. R. A. Flores, P. Pasquier, and B. Chaib-draa. Conversational semantics with social commitments. *Journal of Autonomous Agent and Multiagent Systems, Special Issue on Agent Communication*, 2005. To appear.
11. N. Howden, R. Ronnquist, A. Hodgson, and A. Lucas. Jack intelligent agents: summary of an agent infrastructure. In *Proceedings of the Fifth International Conference on Autonomous Agents*. ACM Press, 2001.
12. M. Inverno, M. Luck, M. Georgeff, D. Kinny, and M. Wooldridge. The dmars architechure: A specification of the distributed multi-agent reasoning system. *Journal of Autonomous Agents and Multi-Agent Systems*, 1-2(9):5–53, 2004.
13. S. C. Levinson. Activity type and language. *Linguistics*, 17:365–399, 1979.
14. P. Pasquier, M. Bergeron, and B. Chaib-draa. DIAGAL: a Generic ACL for Open Systems. In *Proceedings of The Fifth International Workshop Engineering Societies in the Agents World (ESAW'04)*, volume 3451 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 139–152. Springer-Verlag, 2004.
15. P. Pasquier and B. Chaib-draa. The cognitive coherence approach for agent communication pragmatics. In *Proceedings of The Second International Joint Conference on Autonomous Agent and Multi-Agents Sytems (AAMAS'03)*, pages 544–552. ACM Press, 2003.
16. P. Pasquier and B. Chaib-draa. Modèles de dialogue entre agents cognitifs : un état de l'art. *In Cognito : Cahiers de Sciences Cognitive*, 1(4):77–135, 2004.
17. P. Pasquier, R. A. Flores, and B Chaib-draa. Modelling flexible social commitments and their enforcement. In *Proceedings of the Fifth International Workshop Engineering Societies in the Agents World (ESAW'04)*, volume 3451 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 153–165. Springer-Verlag, 2004.
18. A. S. Rao and M. P. Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of Knowledge Representation and Reasoning (KR&R-91)*, pages 473–484. Morgan Kaufmann Publishers: San Mateo, CA, April 1991.
19. L. Royakkers and F. Dignum. No organisation without obligation: How to formalise collective obligation? In M. Ibrahim, J. Kung, and N. Revell, editors, *Proceedings of the* 11^{th} *Internationnal Conference on Databases and Expert Systems Applications*, volume 1873 of *Lecture Notes in Computer Science (LNCS)*, pages 302–311. Springer-Verlag, 2000.
20. M. P. Singh. Social and psychological commitments in multiagent systems. In *AAAI Fall Symposium on Knowledge and Action at Social and Organizational Levels*, pages 104–106, 1991.
21. M. P. Singh. A social semantics for agent communication languages. In *Issues in Agent Communication*, Lecture Notes in Artificial Intelligence (LNAI), pages 31–45. Springer-Verlag, 2000.
22. M. P. Singh, S. Rao, and M. P. Georgeff. *Multiagent Systems : A Modern Approach to Distributed Artificial Intelligence*, chapter Formal Methods in DAI: Logic-Based Representation and Reasoning, pages 331–376. The MIT Press, 1999.
23. P. Thagard. *Coherence in Thought and Action*. The MIT Press, 2000.
24. P. Thagard and K. Verbeurgt. Coherence as constraint satisfaction. *Cognitive Science*, 22:1–24, 1998.
25. G. H. von Wright. *Freedom and determination*. North Holland Publishing, 1980.
26. D. N. Walton and E. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Suny Press, 1995.
27. M. Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, 2001.