

Modelling the Links Between Social Commitments and Individual Intentions

Philippe Pasquier
DAMAS Laboratory
Laval University
G1K 7P4, Québec, Canada
pasquier@iad.ift.ulaval.ca

Brahim Chaib-draa
DAMAS Laboratory
Laval University
G1K 7P4, Québec, Canada
chaib@iad.ift.ulaval.ca

ABSTRACT

Social commitments have been increasingly used to model inter-agent dependencies and normative aspects of multi-agent systems such as the semantics of agent communication. However, current cognitive agent architecture rest on a formalization of private mental states. In this paper, we propose a modelling of the links between private mental states resulting in individual intentions and social commitments.

1. INTRODUCTION AND MOTIVATIONS

On one hand, social commitment has been introduced as a first class concept to represent socially established (and grounded) interagent dependencies and normative aspects such as roles, authority relations, social norms and the semantics of agent communication [13]. On the other hand, main current cognitive agent architectures rest on a formalization of private mental states [18]. In this paper, we propose a modelling of the links between private mental states (resulting in individual intentions) and social commitments when used in the context of agent communication.

2. SOCIAL COMMITMENTS

Social commitments are useful to 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 [9] discusses our modelling of flexible social commitments and their enforcement through sanctions, we simply re-introduce the basic of it here.

The notion of commitment is a social one, and should not be confused with the notion of individual commitment used to emphasize individual intention persistence. Conceptually, commitments are oriented responsibilities contracted towards a partner or a group. Following [17], 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 allows 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 have proposed a dialogue games based agent communication language (DIAGAL [7]) which is sound and complete according to the creation, cancellation, modification and discharge of social commitments. In that context, social commitments allow to get ride of the main problems associated with the mentalistic semantics previously proposed. Indeed, 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 [9]), 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, social commitments were not taken into account in previous cognitive agents theory. As a first step to fill this gap, we propose to model the links between private cognitions and social commitments.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS'06 May 8–12 2006, Hakodate, Hokkaido, Japan.
Copyright 2006 ACM 1-59593-303-4/06/0005 ...\$5.00.

3. INDIVIDUAL INTENTIONS

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 become true. Intentions are either accepted ($I_A(p)$) or rejected ($\neg I_A(p)$).

In order to address the social dimension of communication, we will further distinguish between *internal individual intentions* and *social individual intentions*¹. 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 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 approach, 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. For example, in the JACK-BDI framework [6], intentions that don't match any individual plans or for which all available individual plans have failed fall into those categories. Notice that this selection process implies that trying to achieve individual action (through execution of individual plans) is the prioritized behavior of the agent. Figure 1 sums up this intention typology.

4. LINKING INTENTIONS AND SOCIAL COMMITMENTS

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

¹With 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 [11] or [15]. Here, intentions are classical private intentions.

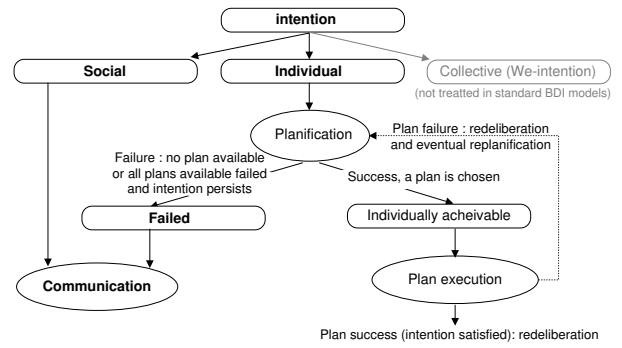


Figure 1: Operational typology of intentions.

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 modelled with negative constraints.

5. RELATED WORK

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, 16]. 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 [12], which introduces the idea of linking individual and social commitments. Comparable links have been introduced for so-called normative-deliberative cognitive agent architecture [5, 1, 3]. In particular, following [4], the following axioms have been introduced [10]:

$$\begin{aligned} S-COMM(i, j, \tau_i) &\rightarrow I_j(\tau_i),^2 \text{ and} \\ S-COMM(i, j, \tau_i) &\rightarrow I_i(\tau_i) \end{aligned}$$

From which, one can deduce the following theorem:
 $\vdash \neg S-COMM(i, j, \tau) \vee (I_i(\tau) \wedge I_j(\tau))$, which clearly states

² Sometimes formulated : $S-COMM(i, j, \tau_i) \rightarrow Goal_j(Does_i(\tau))$

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 2. 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 flexible social commitments.

Constraints provide bidirectional and symmetric links that go behind the above mentioned axioms.³ 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 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 that we modelled in previous works [8].

6. CONCLUSION

While social commitments capture the social interdependencies of the agent, individual intentions make the links between the agent private cognition and its individual or social behavior. This short paper advances the state of the art by motivating and proposing a modelling of the links between individual intentions and social commitments. While this idea is not new, it is shown how the proposed links improve previous proposals.

7. 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: Cambridge, MA, USA, 1990.
- [3] 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.
- [4] 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.
- [5] C. Castelfranchi, F. Dignum, C. Jonker, and J. Treur. Deliberative normative agents: Principles and architecture. In *Intelligent Agents V, Proceedings of the International Workshop on Agent Theories, Architectures, and Languages (ATAL)*, volume 1757 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 364–378. Springer-Verlag, 1999.
- [6] N. Howden, R. Rönquist, 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.
- [7] 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.
- [8] P. Pasquier and B. Chaib-draa. Agent communication pragmatics: The cognitive coherence approach. *Cognitive Systems*, 6(4):364–395, December 2005.
- [9] 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.
- [10] 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 11th International Conference on Databases and Expert Systems Applications*, volume 1873 of *Lecture Notes in Computer Science (LNCS)*, pages 302–311. Springer-Verlag, 2000.
- [11] J. R. Searle. Collective intentions and actions. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*, pages 401–416. The MIT Press: Cambridge, MA, USA, 1990.
- [12] 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.
- [13] M. P. Singh. An ontology for commitments in multiagent systems: Toward a unification of normative concepts. *Artificial Intelligence and Law*, 7:97–113, 1999.
- [14] P. Thagard. *Coherence in Thought and Action*. The MIT Press: Cambridge, MA, USA, 2000.
- [15] R. Tuomela and K. Miller. We-intentions. *Philosophical Studies*, 53:367–389, 1988.
- [16] G.H. von Wright. *Freedom and determination*. North Holland Publishing Co., 1980.
- [17] D. N. Walton and E. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Suny Press, 1995.
- [18] M. Wooldridge. *An Introduction to MultiAgent Systems*. Wiley, 2001.

³We refer the interested reader to [14] for a discussion about bidirectionality in cognitive modelling.