

Logical formalization of social commitments: Application to agent communication languages

(short version)

Benoit Gaudou¹, Andreas Herzig² & Dominique Longin²

¹MSI Lab - IFI
ngo 42, Ta Quang Buu,
Ha Noi, Viet Nam
bgaudou@ifi.edu.vn

²Université de Toulouse, CNRS, IRIT, France
{Andreas.Herzig,Dominique.Longin}@irit.fr

Abstract

Most agent Communication Languages are no longer defined in terms of the agents' mental attitudes, but in terms of social commitments. However, such social approaches have two drawbacks. First, the notion of commitment does not have a clear and unambiguous characterization. Second, commitments are completely unrelated to the agents' reasoning. We remedy this situation by combining a BDI logic with a logic of what is publicly grounded between agents.

Keywords

Agent communication languages, modal logic, grounding, commitments, BDI logic, speech act theory.

1 Introduction

Initially, Agent Communication Languages (ACLs) such as the FIPA-ACL standard [2] had semantics in terms of mental states like belief and intention. However, such a basis lacks verifiability and requires strong hypotheses on agents like sincerity and cooperativity. In order to overcome these problems more recent semantics such as [3] among others are based on commitments. However, some drawbacks remain. First, the notion of commitment does not have a clear

and unambiguous characterization. Second, commitments are completely unrelated to agents' reasoning abilities and in particular to their mental attitudes. We remedy this situation by means of a reductionist logical characterization of social commitments. Contrarily to e.g. Castelfranchi [1] and as explained in section 2, we agree that individual mental attitudes are not enough to characterize social commitments: we have to combine a logic of individual mental attitudes (viz. individual beliefs and intentions) with a logic accounting for the social and public feature of social commitments. To meet this goal, we use the logic of grounding introduced in [4]: it extends a BDI-like logic by a modal operator of what is *publicly established in a group of agents*, as opposed to private mental attitudes (section 3). We then show that it provides a new semantics for commitment-based ACLs by recasting the approach of [3] (section 4).

2 What are commitments?

Commitment has various senses in the AI literature. First we have to distinguish internal commitment and social commitment: the term 'internal commitment' refers to the persistence of an agent's choice to perform an action or to hold some proposition, while 'social commitment' refers to a particular relation between agents. This paper focusses on the latter; however, internal commitment is accounted for *via* the notion of intention (defined in section 4).

Castelfranchi reduces social commitment of the *debtor* i to the *creditor* j w.r.t. the action α using mutual knowledge: " i and j mutually know that i intends to do α and this is j 's goal, and that as for α j has specific rights on i (j is entitled by i to α)." [1, p. 3]. Unlike [3], we agree with Castelfranchi that social commitment is not primitive. In the sequel we discuss his definition in detail.

First, it requires mutual *knowledge* on i 's intention, which logically implies that i 's intention actually holds (under the hypothesis of introspection on intentions). But as Castelfranchi says himself, the actual intention of i to perform the action α is neither necessary nor sufficient for his social commitment to do α . We therefore to weaken his definition by substituting mutual knowledge with the notion of grounding capturing only the public feature of i 's intention. Note that the public ground of an attitude does not imply that this attitude actually holds.

Second, the action α to which i is committed should be a goal of agent j . We argue that this hypothesis is also too strong. Compare this to speech act theory ([6, p. 182–183]) and in particular consider commissive acts. On the one hand, a promise creates a commitment of the speaker toward the hearer and requires what is promised to be good for the hearer, but on the other hand a threat has the same social result (creation of a commitment) but requires in contrary the object of the threat not to be good for the hearer. We therefore have to distinguish what we call desirable social commitment (when the performance of α is a goal of the creditor) from undesirable social commitment (when the performance of α is not a goal of the creditor). Social commitment \dot{a}

la Castelfranchi only covers desirable social commitments. We do not need to define social commitment in such a restrictive way. Hence we do not specify in our definition of social commitment whether the creditor j has the goal that α be performed or not.

Finally, while we agree that the normative aspect of social commitments is important in general, we focus in this paper on the informal commitments involved in ACLs, where the deontic part of commitment can be omitted: after a promise a social commitment is incurred, but rights and duties often stay implicit and undefined.

To conclude, we consider that a social commitment to do α can be characterized as a public ground of both agents on the intention of one agent to perform an action. Castelfranchi argues that commitments are created by communicative acts. This suggests to consider speech act theory and to formalize not only commitment as an effect of speech acts, but also speech acts creating and managing them.

In commitment-based ACL semantics, commitments are not static but may be in various states. Each speech act is characterized by the commitment it produces, or more precisely by the changes in commitment states it induces. A model in terms of a commitment life cycle is introduced in [3].

3 The grounding logic

The logic of [4] captures what is publicly grounded in a group of agents. It extends the logic of intention developed in [5] with a modal operator Grd , where the formula $Grd_I \varphi$ reads “ φ is publicly grounded for the group I ”. We have shown that the logic provides a new semantics to FIPA ACL [2] and also captures Walton and Krabbe’s notion of propositional social commitment [7] in persuasive dialogue games.

We refer to [4] for the details of the (fairly standard) possible worlds semantics and its axiomatics, and here only give the reading of relevant operators. In the sequel, i, j denote agents and α an action. The expression $(i:\alpha)$ denotes that i is the author of action α .

It is supposed that what is publicly grounded is consistent, public for every subgroup of I , and is formed by a joint agreement of each member. For the sake of simplicity (and also notational economy) we identify $Grd_{\{i\}}$ with the usual belief operator [5]: $Grd_{\{i\}} \varphi$ means that i believes φ . We write $Grd_i \varphi$ for convenience.

$Int_i \varphi$ reads “agent i has the intention that φ holds”. We suppose that Int_i is defined from belief, choice and time operators (cf. [5]). $Done_\alpha \varphi$ expresses that the action α has been performed before which φ held. Therefore $Done_\alpha \top$ means that α has just been performed. Formula $Int_i Done_\alpha \top$ expresses that i has the intention to do action α .

4 Formalization of commitments

In the sequel $C(i, j, \alpha)$ represents a social commitment of i toward j to do action α . Following the discussion of section 2 we define a social commitment to do α as the debtor's publicly grounded intention to perform α :

DEFINITION. $C(i, j, \alpha) \stackrel{def}{=} Grd_{\{i,j\}} Int_i Done_{(i:\alpha)} \top$

Many speech acts can lead to incur such a commitment. Here we only characterize promises, which are a spontaneous commitment of an agent and therefore produce a social commitment whose debtor is the speaker. We impose as precondition of Promise that the speaker stays consistent, *i.e.* that he has not expressed yet that he doesn't want to perform the action. The semantics is provided in terms of their Feasibility Precondition (FP) their and Intentional Effect (IE). FP describes the condition under which an action can be performed, and IE describes the illocutionary intention that, when recognized by the addressee of the act, identifies the performed speech act.

DEFINITION. $\langle i, \text{Promise}, j, Done_{(i:\alpha)} \top \rangle$

FP: $\neg Grd_{\{i,j\}} \neg Int_i Done_{(i:\alpha)} \top$

IE: $C(i, j, \alpha)$

5 Perspectives

The aim of this paper was to use logical framework to describe social approaches of ACLs. In particular we formalized the notion of social commitment. Due to space restrictions only a part of our work has been presented here. In particular, our logical formalism accounts for the whole life cycle of Colombetti *et al.*'s approach: on the one hand we can give a formal characterization of each of its possible states, on the other hand we can provide a semantics for every speech acts leading to the commitments dynamics. Moreover, we concentrate on commitments to do an action, but very interesting links can be done between such commitments and propositional commitment as defined in detail in [7].

References

- [1] C. Castelfranchi. Commitment: from individual intentions to groups and organizations. In V. Lesser, editor, *Proc. of ICMAS*, pages 528–535. MIT press, 1995.
- [2] FIPA. FIPA Communicative Act Library Specification. <http://www.fipa.org/specs/fipa00037/>, Foundation for Intelligent Physical Agents, 2002.

- [3] N. Fornara, F. Viganò, and M. Colombetti. Agent communication and artificial institutions. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 14:121–142, 2007.
- [4] B. Gaudou, A. Herzig, and D. Longin. Grounding and the expression of belief. In P. Doherty, J. Mylopoulos, and C. A. Welty, editors, *Proc. KR 2006*, pages 211–229, Windermere, UK, 2006. AAAI Press.
- [5] A. Herzig and D. Longin. C&L intention revisited . In D. Dubois, C. Welty, and M.-A. Williams, editors, *Proc. KR2004*, pages 527–535, Whistler, Canada, 2004. AAAI Press.
- [6] D. Vanderveken. *Principles of language use*, volume 1 of *Meaning and Speech Acts*. Cambridge University Press, 1990.
- [7] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New-York Press, NY, 1995.