A Synthesis Between Mental Attitudes and Social Commitments in Agent Communication Languages

Guido Boella
Università di Torino - Italy
guido@di.unito.it

Joris Hulstijn
VU - Amsterdam
jhulstijn@feweb.vu.nl

Leendert van der Torre
CWI Amsterdam and TU Delft
torre@cwi.nl

Abstract

There are two main traditions in defining a semantics for agent communication languages, based either on mental attitudes or on social commitments. In this paper we show how the role metaphor can be used to bridge the gap between these two approaches. First, we show how dialogues can be modelled as games - a form of normative systems - and how mental attitudes can be attributed not only to agents, but also, in a public manner, to the roles of the game. The dialogue moves allow an agent playing a role to modify the roles' mental states, as specified by the counts-as conditionals (also known as constitutive norms) defining the game. The player of a role is expected to act as if it has the mental attitudes attributed to its role during the dialogue and to prevent its role's mental attitudes from becoming incoherent, as it does for its own private mental attitudes. Secondly, we show how roles as descriptions of expected behavior maintain the normative character of social semantics. Due to the bridge between the two approaches, results and tools from one approach can be used in the other one.

1. Introduction

The modelling of communication in agent theories was originally inspired by Searle [15]'s theory of speech acts. Cohen and Perrault [9] show that their systematic analysis in terms of preconditions and postconditions of speech acts can be modelled straightforwardly in terms of planning operators. The preconditions of these operators refer to the beliefs of the speaker and of the addressee of the speech acts and the postconditions to changes in their beliefs and goals. This view led to the creation of agent communication languages (ACL) of which the semantics are given by preconditions (feasibility conditions) and postconditions (rational effects), formulated in terms of the mental attitudes of the interactants, like FIPA-ACL [11].

In contrast, Singh [17] proposed a social semantics for agent communication languages, based on the notion of *commitment*. A commitment binds a speaker's attitudes to the community and, thus, it has a public character. Commitment is interpreted as a kind of obligation which is undertaken by the speaker through uttering a speech act; for example, Walton and Krabbe [19] argue that: "to assert a proposition may amount to becoming committed to subsequently defending the proposition, if one is challenged to do so by another speaker." Are these two approaches, presented as competing alternatives, really incompatible?

In this paper we ask the following research question: how can we use the role metaphor to bridge the gap between the mental attitudes approach and the social commitments approach to the semantics of agent communication languages? Roles are useful because mental attitudes attributed to roles of a dialogue game instead of attributed to agents capture the public character of meaning which is offered by commitment approaches.

The motivation of maintaining a mentalistic semantics, albeit referred to roles and not to agents, is to reuse the extensive work on the semantics of ACL in terms of mental attitudes [11]. It is sufficient to refocus the model from the agents' beliefs and goals to the roles' beliefs and goals. This shift is also coherent with the separation of concerns requirement put forward by the conceptual modelling community. Separation of concerns means that the behavior of an agent should be specified separately from the interaction capabilities of an agent.

Reconciling the two semantics is not only a theoretical or philosophical problem, since there is still no consensus about which one to choose. Moreover, to combine two legacy systems, one with social semantics and one with mental attitudes semantics, one has a problem. It is also difficult to predict what happens when two agents developed separately interact, one based on mental attitudes semantics and one with social commitments semantics.

The layout of the paper is as follows. In Section 2 we discuss our approach to dialogue by modelling games as normative systems with roles. In Section 3 we introduce a formal model of normative systems for defining dialogue games and an example. Conclusions end the paper.

2. Dialogue as a game and its roles

2.1. Thesis - Antithesis - Synthesis

The most natural interpretation of a speech act is as an action that alters the mental attitudes of the agents involved. By making an inform act, the speaker shows that it intends the addressee to believe some information. This view has become popular by the FIPA semantics of agent communication [11]. However, such mentalistic thesis appears to be flawed, as argued by Singh [17], e.g., because:

1. Communication is intersubjective and public (Singh [17], Walton and Krabbe [19], Brandom [7], Kibble [12], Verdicchio and Colombetti [18]). In contrast, beliefs and goals of the speakers are not accessible by the addressee, since they have a private character.
2. An independent observer cannot verify whether agents conform to the ACL semantics [20].
3. The sincerity assumption that is necessary to reason with mentalistic semantics does not apply to non-fully cooperative dialogue contexts - e.g., negotiation and persuasion - or, more generally, to open systems.

As a consequence, the mentalistic thesis of communication has been countered by an antithesis, the socially oriented perspective: since private mental attitudes cannot be the basis of communication, mental attitudes should be replaced by or embedded into commitments.

The antithesis, however, can be countered in turn. In particular, one could argue that the interpretation of a commitment as an obligation to defend the communicated content is too strong in many circumstances. It makes sense for competitive environments, like argumentation dialogue or negotiation, but it does not in a cooperative one, like information seeking or inquiry dialogues, where a commitment can simply be interpreted as an expectation.

The method we adopt in this paper is to model dialogue as a game in which agents play roles. Speech acts are moves in the game executed by the players of the roles, and their preconditions and effects refer to the mental states attributed to the roles, not to the inaccessible private mental states of the agents themselves. To propose a synthesis of the two approaches based on the notion of a role, we show that roles can support the intuitive notion that speech acts refer to mental attitudes, but also that roles preserve the public character of meaning, and support, when necessary, the commitment of speakers to what they have said. A precondition of this method is that mental attitudes can be attributed to roles as well as to agents. As in [5], we describe roles as agents with mental attitudes, albeit a different kind of agents.

A game is a rule governed social activity. A game is built from constitutive rules, according to Searle [16]'s view of construction of social reality. Rules create activities like playing chess, as well as social institutions like money, property, and normative systems. Institutional facts like these emerge from an independent ontology of "brute" physical facts through constitutive rules of the form "such and such an X "counts as" Y in context C" where X is any object satisfying certain conditions and Y is a label that qualifies X as being something of an entirely new sort. An example of a constitutive rule is "this bit of paper counts as a five Euro bill". By contrast, regulative rules, like obligations and permissions, regulate the newly created games.

Thus, a game is composed of both regulative and constitutive rules: it is a normative system. This view also captures the intuition that in many games illegal moves are associated with sanctions to have obligations. To model games as normative systems, we use and extend our model [5], in which normative systems are considered as agents. The basic methodology of our work is to model complex entities of social reality like groups, normative systems, organizations and roles as a kind of agent [4]. Social entities exist because they are collectively accepted by agents [15]. To define the behavior of social entities, the agents collectively, and thus publicly, attribute mental attitudes to them. The agent metaphor allows us to explain the properties and features of social entities in terms of the properties and features of agents. In [5] a normative system is considered as an agent, where: the regulative norms, like obligations and permissions, are mapped onto the goals of the agent; the constitutive norms are mapped onto the beliefs of the agent. Moreover, a normative system is supposed to behave autonomously to restore the regularities prescribed by norms through a process of monitoring violations and sanctioning them. The metaphor, however, stops here since social entities cannot directly act. Monitoring and sanctioning are carried out by real agents playing roles in the system.

2.2. Constitutive rules and roles

Roles have been introduced in the multiagent systems community as a way to coordinate the behavior of individual agents by means of a normative system or an organization. Roles are associated with expertise, capabilities such as planning rules, and with responsibilities to maintain or achieve some state of affairs. Roles are often also associated with obligations and permissions that restrict the means by which they can fulfill the responsibilities.

In sociology, roles are often defined as descriptions of expected behavior. To describe behavior, agent theory uses beliefs and goals: hence, in [4, 6] we consider roles as descriptions of agents made in terms of beliefs and goals. In the same way as social entities are constructed by the collective attribution of mental entities, roles exist only because they are publicly attributed mental attitudes by the norma-

tive system they belong to. However, roles do not act in the world, and therefore are agents of a special kind. This fact, as first sight, seems in contrast with the possibility to use expressions like "the president has sold the enterprise" where the president is a role in an organization. However, an action like selling is not a physical action. It is the exercise of a power by the agent playing the role of president. Roles are associated with powers which are exercised by the agents that play the roles. Powers allow players of roles to:

- change the normative system which the roles belong to (e.g., by making it commit to a payment),
- change the features of roles themselves (e.g., the president commits itself to new responsibilities), and
- change the features of other roles (e.g., the president creates an obligation for an employee by ordering it).

A player of a role can exercise its powers only because its physical actions "count as" institutional acts, for example, as the creation of an obligation for another role. Constitutive rules, thus, explain how an utterance by a player "counts as" a move in the game and how the move impacts on the beliefs and goals of the roles: the moves performed by a role in a game are the powers of the player of the role.

To play a role an agent is expected to adopt the goals that correspond to its responsibilities and to carry out them according to the beliefs that correspond to its expertise. In other words: the agent must act *as if* the beliefs and goals of the role were its own beliefs and goals, and maintaining them coherent, as it does for its own mental attitudes. The pressure of coherence is basic principle of cognition, and, as shown in next section, according to Pasquier and Chaib-draa [14], a basic point in the explanation of dialogue.

Hence, roles possess both the features that we need to model dialogue: they have mental attitudes and they represent expectations to which a player is bound. Finally, a role does not force a player to actually have the mental attitudes of its role: roles are flexible enough to represent dialogues where sincerity is not assumed or necessary.

2.3. The rules of the dialogue

The FIPA-ACL semantics [11] of speech acts is composed of feasibility conditions and rational effects. For the *inform* speech act the feasibility conditions are that, first, the speaker believes that the propositional content is true, and, second, it does not already believe that the addressee has any knowledge about the truth of the propositional content. The rational effect is that the addressee also comes to believe the propositional content. Interpreting the first feasibility condition as a precondition means that the move cannot be played if the proposition is not believed. In our model we interpret this condition not as a precondition but as a presupposition: if the speaker plays an *inform* move, this means that it must believe the proposition. This presupposition is modelled by a constitutive rule and must be accommodated, if possible, otherwise a contradiction is created, e.g., when the speaker asserts first p and then $\neg p$ without retracting its first move. The speaker makes its role's beliefs incoherent.

The belief we are talking about is the belief publicly ascribed to the role played by the speaker, not a private belief of the player. Thus, it is possible that the proposition is accommodated in the beliefs of the role, while it cannot be accommodated in the player's beliefs: the agent is lying. This explains how it is possible that role's beliefs can become contradictory, while the agent is not in an inconsistent state.

In contrast, we do not model the second precondition, since, as noticed by Amgoud *et al.* [1], it belongs to the reasoning level of the agents and, thus, it is not part of the game: it concerns the strategies used by an agent in playing a role. In the game an agent is free to inform the addressee about propositions it believes the addressee already believes: this is a legitimate, even if dumb, move.

Concerning the rational effect that the addressee comes to believe what communicated, we do not model it in terms of the beliefs of the receiver agent. Rather, we model only the public effect of an *inform* via a constitutive rule expressing a power of the addressee role: *inform* affects the beliefs publicly attributed to the role played by the receiver. Moreover, the addressee role is held to believe the proposition only if the proposition is not challenged by the addressee. This idea relies on Brandom's view of dialogue: "when a commitment is attributed to an interlocutor, entitlement to it is attributed as well, by default" [7, p.177].

We adopt Pasquier and Chaib-draa [14]'s view that dialogue arises from the need of maintaining coherence of mental states: "two agents communicate if an incoherence forces them to do so. [...] Conversation might be seen [...] as a generic procedure for attempting to reduce incoherence". An agent engaged in the dialogue goes on to avoid entering in contradiction. Thus, if a new belief may be in conflict with other ones, the agent is compelled to challenge the *inform* to avoid a contradiction in its role. In contrast with [14], we do not consider here the coherence of the private mental states of the agents, but the coherence of the mental attitudes of the roles. Referring only to the agents' beliefs would not be realistic, since there is no way to ensure that the addressee, as an autonomous agent, accepts to believe a proposition. This is one of the limitations of the traditional mentalistic approach. We overcome it using rules referring to mental attitudes, but those of the roles composing the game, rather than to the agents' private ones. As long as an agent plays a game, it cannot refuse that what it has said will be considered as a public display of its position in the game, according to its role.

3. The formal model

For a formal definition of the agents in a multiagent system we are inspired by the rule-based BOID architecture [8]. Beliefs and goals are represented by conditional rules gathered in different sets representing the mental and motivational states of an agent.

We assume that the base language contains boolean variables and logical connectives. The variables are either *decision variables* of an agent, which represent the agent's actions and whose truth value is directly determined by the agent, or else *parameters* - among which *institutional facts*, which describe the state of the world, whose truth value can only be determined indirectly via belief rules. The agent's motivational state contains two sets of rules. *Goal rules* G_a express the attitudes of agent a towards a given state, depending on the context. Beliefs B_a are expressed as rules, too. Consequences of beliefs are produced via a closure operator out.

Definition 1 (Multiagent system)
A multiagent system MAS *is a tuple* $\langle RA, NS, RO, X, B, G, AD, MD, \geq, I, PL\rangle$ *such that:*

- *Real agents* RA*, games as normative systems* NS*, roles* RO*, variables* X*, beliefs* B*, goals* G*, and institutional facts* $I \subset X$ *are finite disjoint sets. We write* $A = RA \cup NS \cup RO$ *for the set of all agents.*
- $Lit(X)$ *is the set of literals from* $X \cup \{\neg x \mid x \in X\}$*.* $Lan(X)$ *is the set of logical formulas built out of* X *with the usual connectives. The set of rules built from* X*, written as* $R(X) = Lan(X) \times Lan(X)$*, is the set of pairs* $\langle \chi, \phi\rangle$*, written as* $\chi \rightarrow \phi$ *where* $\chi, \phi \in Lan(X)$*.*
- *An agent description* $AD : A \rightarrow 2^{X \cup B \cup G \cup A}$ *is a complete function that maps each agent to a set of decision variables, beliefs and goals, and agents. For each agent* $a \in A$*, we write* X_a *for* $X \cap AD(a)$*,* B_a *for* $B \cap AD(a)$*,* G_a *for* $G \cap AD(a)$*, and* A_a *for* $A \cap AD(a)$*.*
- *The function* AD *assigns decision variables only to real agents* RA*: normative systems and roles are not autonomous and can act only through the intermediation of role players.* AD *does not necessarily assign each variable in* X *to at least one agent in* RA*. We write* $P = X \setminus \cup_{a \in RA} X_a$ *for the set of parameters.*
- *The function* AD *associates also agents with agents, because normative systems and roles exist only as they are described as agents by, respectively, real agents and normative systems. Formally, a socially constructed agent* $b \in NS \cup RO$ *exists only as some other agents attribute mental attitudes to it:* $\forall b \in NS \cup RO\ \exists a \in A : b \in AD(a)$*.*
- *The mental description* $MD : (B \cup G) \rightarrow R(X)$ *is a complete function from the sets of beliefs and goals to the set of rules built from* X*.*
- *A priority relation* $\geq$*:* $A \rightarrow 2^G \times 2^G$ *is a function from agents to a transitive and reflexive relation on the powerset of goals containing at least the subset relation.*
- *The role playing function* $PL : RO \rightarrow RA$ *associates a role to its player.*

A game modelled as a normative system includes both regulative and constitutive rules. Regulative norms are based on the notion of a conditional obligation with an associated sanction as we do in [5]. Obligations are defined in terms of goals of both the bearer of the norm and the normative system. We do not discuss regulative norms here in detail, because for the purpose of this paper we only need constitutive norms. We formalize the "counts-as" conditional of constitutive rules as a belief rule of the normative agents NS or of roles. For the logic of rules we use an input/output logic, called out_3 [13], which provides us with the consequences of a set of belief rules.

Definition 2 (Counts-as relation)
Let $MAS = \langle RA, NS, RO, X, B, G, AD, MD, \geq, I, PL\rangle$ *be a normative multiagent system, and let* $out(S)$ *be the smallest set of conditionals that contains the set of rules* S *and is closed under left and right replacement by logical equivalents, and the following inference rules:*

$$\frac{\chi \rightarrow \phi}{\chi \wedge \psi \rightarrow \phi} SI \qquad \frac{\chi \rightarrow \phi, \chi \wedge \phi \rightarrow \psi}{\chi \rightarrow \psi} CT \qquad \frac{}{\phi \rightarrow \phi} Id$$

where $\chi, \phi, \psi \in Lan(X)$*. We write* $y \in out(S, x)$ *iff* $x \rightarrow y \in out(S)$*. A literal* $x \in Lit(X)$ *counts as* $y \in Lit(I)$ *in context* $C \in Lan(X)$ *for agent* c*, written as* $MAS \models counts\text{-}as_c(x, y|C)$*, iff* $C \wedge x \rightarrow y \in out(B_c)$*.*

3.1. The rules of dialogue

In this section, we focus on persuasion dialogues. We consider how an *inform* move can be played, and how such a move can be challenged or accepted. We introduce the constitutive rules which the semantics of *inform* in terms of mental attitudes.

Although technically the multiagent formalism is based on atomic propositions, we use a notation convention for speech acts that makes the speaker, addressee, content and type of speech act explicit. Since at no place we use quantifiers and variables that range over these placeholders, they can all be regarded as constants.

We distinguish between the locutionary level of the utterances, represented by $utter$, and the illocutionary acts like *inform*, $accept$, $retract$ and *why-question*. We distinguish two kinds of moves in the dialogue game: those di-

Turn	Moves of role a	Consequences of B for role a	Moves of role b	Consequences of B for role b
1	*inform*(a,b,p) (1)			
		p (5)		p if $\neg challenge(b,a,p)$ (6)
2			*inform*$(b,a,q \supset \neg p)$ (1)	
			inform(b,a,q) (1)	
			$challenge(b,a,p)$ (3)	
		$q, q \supset \neg p, \neg p, p$ if $\neg challenge(a,b,q)$ or if $\neg challenge(a,b,p \supset q)$ (6)		$q, q \supset \neg p, \neg p$ (5)
3	*why-question*(a,b,q) (2)			
	$challenge(a,b,q)$ (4)	$p, q \supset \neg p$		
				$\neg q, q \supset \neg p$ (8) if $\neg challenge(b,a,\neg q)$
4			$retract(b,a,$*inform*$(b,a,q))$ (2) hence $\neg$*inform*(b,a,q) hence $\neg challenge(b,a,p)$	$q \supset \neg p$
			$accept(b,a,\neg q \wedge p)$ (2)	$\neg q, p$ (7)

Figure 1. The consequences of applying rules in the example of Section 3.1.

rectly played by making an utterance and those which are played by means of other moves.

Definition 3 (Speech acts) In the propositional variables $X \setminus P$ we distinguish two subsets of speech acts $PSA \cup CSA = SA$, where $a, b \in RO$, $\phi \in Lan(X \setminus SA)$, $\alpha \in SA$, and the integer t is the time the move is uttered:

- $PSA \subseteq I$ are the primitive speech acts: *inform*(a,b,ϕ,t), $accept(a,b,\phi,t)$, $retract(a,b,\alpha,t)$, *why-question*(a,b,ϕ,t) are directly played by means of making an utterance (Rules 1 and 2 below).
- $CSA \subseteq I$ are the complex speech acts generated by means of other speech acts; e.g., $challenge(a,b,\phi,t)$ can be performed by means of sets of *inform* moves or by a *why-question* (Rules 3 and 4 below).

For each $\Psi(a,b,\phi,t) \in PSA$ there exists a corresponding decision variable of a real agent a in RA which is the act of making an utterance e.g., $utter(x,y,\Psi(a,b,\phi,t),t) \in X_x$, where $x, y \in RA$, $x = PL(a)$ and $y = PL(b)$.

Since speech acts are institutional facts in I, and, hence, parameters in P, they cannot be made true directly by agents. They are made true, and, hence, moves are played, indirectly, as specified by the constitutive rules of the game. The constitutive rules, attributed to the roles' beliefs, are public, thus the consequences of those rules according to the out operation are public too.

Definition 4 (Constitutive rules of the game) The following rule schemata are general beliefs of both the roles a and b (where $\sigma \in \{a,b\}$), involved in a game, i.e., the rule instances are included in $AD \cap B_\sigma$. In the following, x is an agent participating in the dialogue with role a, and y is an agent with role b. $utter(x,y,\Psi(a,b,\phi,t),t)$ is a decision variable referring to an action of agent x, *inform*(a,b,ϕ,t) and $\Psi(a,b,\phi,t)$ (where $\Psi \in \{retract,$ *why-question*$, accept\}$) are illocutionary acts performed by role a played by agent x, directed to a role b played by agent y, ϕ is a formula of $Lan(X \setminus SA)$, and t, t' are time instants, where $t < t'$.

1. $(\; utter(x,y,$*inform*$(a,b,\phi,t),t) \wedge \neg retract(a,b,$*inform*$(a,b,\phi,t),t') \rightarrow$ *inform*$(a,b,\phi,t) \;) \in B_\sigma$
2. $(\; utter(x,y,\Psi(a,b,\phi,t),t) \rightarrow \Psi(a,b,\phi,t) \;) \in B_\sigma$
3. $(\;$ *inform*$(a,b,\chi,t) \wedge$ *inform*$(a,b,\chi \supset \neg\phi,t) \rightarrow challenge(a,b,\phi,t) \;) \in B_\sigma$
4. (*why-question*$(a,b,\phi,t) \rightarrow challenge(a,b,\phi,t)) \in B_\sigma$

Rule 1 explains how an agent x can perform an $inform$ illocutionary act as role a by uttering a sentence. Note that this rule does not have any effect if the agent retracts subsequently the move it played.

Rule 2 explains how an agent x can perform the remaining illocutionary acts in PSA by means of its utterances. This game does not allow the retraction of these moves.

Rules 3 and 4 model two ways of playing the complex speech act $challenge$: either by introducing a counterargument by means of two *inform*s or by asking a justification of the speaker. The meaning of $challenge$ is below.

Given a move played by role a, the following rules describe the specific effects on role a and role b. Symmetrical rules exist when the two roles are inverted:

5. $(\;$ *inform*$(a,b,\phi,t) \rightarrow \phi \;) \in B_a$
6. $(\;$ *inform*$(a,b,\phi,t) \wedge \neg challenge(b,a,\phi,t') \rightarrow \phi \;) \in B_b$
7. $(\;$ *inform*$(a,b,\phi,t) \wedge accept(b,a,\phi,t') \rightarrow \phi \;) \in B_b$
8. $(\;$ *why-question*$(a,b,\phi,t) \wedge \neg challenge(b,a,\neg\phi,t') \rightarrow \neg\phi \;) \in B_b$

Rule 5 expresses a precondition of *inform*. Informing that ϕ implies that the speaker a believes the presupposition that ϕ. Note that the belief that ϕ holds only as long as the *inform* is not retracted (Rule 1).

Rule 6 models Brandom's intuition that an unchallenged *inform* by default entitles the speaker role a to believe that the addressee b believes (in its role in the game) the conveyed proposition, unless b does not explicitly challenge the information conveyed. Rule 7, instead, models explicit acceptance by b of the proposition ϕ which a informed about.

Rule 8 models the effect on b of being challenged by a *why-question* move of a. Such an inference will be made unless b challenges the challenge. We do not claim that this is the general case, since all these rules work for a specific kind of dialogue game. A more general rule would be that if the speaker is unable or unwilling to provide justifications after a *why-question*, then it is committed not to use the challenged proposition to support its arguments.

Consider the following example:

1. a:	The president will win the election.	*inform*$(a, b, p, 1)$
2. b:	But there is fraud,	*inform*$(b, a, q, 2)$
	so the president will not win.	*inform*$(b, a, q \supset \neg p, 2)$
3. a:	Fraud? But why!	*why-question*$(a, b, q, 3)$
4. b:	Fair enough, no fraud.	*retract*$(b, a,$*inform*$(b, a, q, 2), 4)$
	So you're right.	*accept*$(b, a, \neg q \wedge p, 4)$

The consequences of the constitutive rules driving the game, due to the out operation, are shown in Figure 1, which illustrates the example above. Notice that the number of the applied rule is shown on the right while the time instants are omitted. For clarity, in the second and fourth columns the moves are shown separately, even if they are consequences of beliefs of both roles; in the third and fifth column the consequences on the beliefs of role a and role b are shown, respectively. For simplicity we do not report the utterances of the agents. Representing the agent's own mental states is not important, since the dialogue is determined by roles' mental states only.

In Turn 1, role a informs b that p. The rules of role a now state that p is a presupposed belief of a. The rules of role b state that the belief p can be attributed to b unless b challenges a's *inform*.

In Turn 2, the agent playing role b challenges a's information by means of an argument of the form $q \wedge (q \supset \neg p)$.

In Turn 3, a is facing the risk of entering in an incoherent state: if it does not do anything, it will get into a contradiction, since b's argument supports $\neg p$, which is in contrast with its current beliefs (p). It has some alternatives: retracting its first *inform* or challenging b's argument. So a decides to challenge b's challenge by asking for justification (*why-question*).

Finally, in Turn 4 role b retracts the *inform* about q, thus giving up its challenge to p, and, subsequently, it accepts p.

Note that while *inform* introduces beliefs, similar definitions can be given for other speech acts, like *propose*, *promise* or *request*, which introduce or remove goals and obligations via suitable constitutive rules. Like for *inform*, these speech acts can create entitlement by default or by explicit acceptance, depending on the rules of the game. In case of goals and obligations, the role can become incoherent because the agent commits to conflicting goals or it behaves in a way that leaves unsatisfied the goals of the role it is playing, which represent its expected behavior.

3.2. Commitments and their public character

We introduce the rules of a dialogue game to show that they can safely refer to beliefs since the beliefs they refer to are not private and inaccessible. Rather they are the beliefs attributed to the role the agents are playing in the dialogue game. These beliefs are publicly attributed following the constitutive rules which define the role in the game.

There is still an open question: how do we capture the appealing idea of the commitment of a speaker to defend what it said, an idea which has been put forward by [17, 19]?

We do not introduce an explicit commitment referring to actions like defending, but we use only the beliefs and goals of the role. The commitment emerges from the necessity of maintaining the coherence of the role's beliefs.

After an *inform* move the speaker should be committed to defend its assertion from challenges. In our model this commitment does not derive as an effect of the *inform*. Assume that role a does not reply to a challenge by role b to its *inform*. This means that, by default, it accepts the content of the challenge. But since a challenge negates the content of the previous *inform* of a, accepting the content of the challenge amount for a to entering in an inconsistent belief state. Thus a is compelled to defend its position. The same mechanism is used to explain why an agent is compelled to retract some previous information it is committed to if it is not able to challenge it. Retracting a preceding *inform* is just another way of preventing the contradiction.

Since commitment is not directly part of the model, we can defend the model against the objection that social commitment semantics are too strong in cooperative situations. In non-cooperative contexts we can strengthen the expectations concerning the roles by adding sanctions to define obligations which apply to non-legal moves.

In this way we get a parametric framework: it does not only allows us to model different types of dialogue by using specific constitutive rules, and different initial goals and beliefs, but also to model different types of commitments ranging from mere expectations in cooperative dialogue, to formal obligations enforced by a kind of normative system i.e., the community of players of the game.

4. Conclusion and future research

In this paper we propose a synthesis between the approaches to the semantics of agent communication based on mental attitudes and those based on social commitments. As a bridge between the mentalistic approach and the social commitments approach, we use the role metaphor. We show how mental attitudes can be attributed to roles – which are public – instead of to agents. First, the behavior of an agent in the role it is playing is formalized by "counts-as" conditionals, also known as constitutive norms. The player of a role is expected to act as if it has the mental attitudes that are attributed to it during the dialogue. Secondly, we show how roles as descriptions of expected behavior maintain the normative characterization given by social semantics.

The motivation of maintaining a mentalistic semantics, albeit referred to roles, is to be able to reuse the extensive work on the semantics of ACL in terms of mental attitudes, e.g., the FIPA standards [11]. It is sufficient to refocus the model from the agents' beliefs and goals to the roles' beliefs and goals. By using roles we solve the three issues of criticism that were raised against the mentalistic approach.

While Bentahar et al. [3]'s model integrate mental attitudes and social commitments, we claim that social commitments, even if necessary, can be directly explained by mental attitudes, if attributed to roles.

A different tradition in agent communication is dialogue games [10], which describe patterns and the order in which moves must be made. However, dialogue rules modelled as hard constraints can not deal with conflicts between applicable rules, and prevents contrary-to-duty reasoning. We need a way to represent expectations towards moves, that can distinguish between violations and inconsistencies, as we can do modelling games as normative multiagent systems.

Here, we present a simple formal framework to illustrate our point. We focus on the *inform* move in persuasion dialogues since they are more troublesome for mentalistic approaches. Extensions to other speech acts, like requests or proposals, and other types of dialogue can be dealt with, e.g., by using the more complex model of self-modifying normative systems proposed in [5]. Other issues which are not addressed here are the problem of logical omniscience, mutual beliefs, sequences of speech acts, or undercutters. Moreover, issues of like non-monotonicity for the retraction of speech acts are dealt very simply in this paper.

Finally, future work concerns studying how our approach can be used as the basis for modelling dialog among agents independently developed with different semantics. In particular, it is interesting to understand if by using a framework like Reo [2] it is possible to model the dialog protocol as an exogenous coordination mechanism, so to be able to enforce the rules of the game regardless of the behavior of the coordinated agents.

References

[1] L. Amgoud, N. Maudet, and S. Parsons. An argumentation-based semantics for agent communication languages. In *Procs. of ECAI'02*, pages 38–42. IOS Press, 2002.

[2] F. Arbab. Abstract behavior types: A foundation model for components and their composition. In *Formal Methods for Components and Objects*, LNCS 2852, pages 33–70. Springer Verlag, Berlin, 2003.

[3] J. Bentahar, B. Moulin, J. Meyer, and B. Chaib-draa. A logical model for commitment and argument network for agent communication. In *Procs. of AAMAS'04*, pages 792–799. ACM Press, 2004.

[4] G. Boella and L. van der Torre. Organizations as socially constructed agents in the agent oriented paradigm. In *Procs. of ESAW'04*, LNAI n. 3451, pages 1–13, Berlin, 2004. Springer Verlag.

[5] G. Boella and L. van der Torre. Regulative and constitutive norms in normative multiagent systems. In *Procs. of KR'04*, pages 255–265. AAAI Press, 2004.

[6] G. Boella and L. van der Torre. The ontological properties of social roles. In *Procs. of LOAIT at ICAIL'05*, 2005.

[7] R. B. Brandom. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard Univ. Press, 1994.

[8] J. Broersen, M. Dastani, J. Hulstijn, and L. van der Torre. Goal generation in the BOID architecture. *Cognitive Science Quarterly*, 2(3-4):428–447, 2002.

[9] P. R. Cohen and C. R. Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3:177–212, 1979.

[10] U. Endriss, N. Maudet, F. Sadri, and F. Toni. Protocol conformance for logic-based agents. In *Procs. of IJCAI'03*, pages 679–684. Morgan Kaufmann, 2003.

[11] FIPA. FIPA ACL communicative act library specification. Technical Report SC00037J, FIPA, 2002.

[12] R. Kibble. Beyond BDI? brandomian commitments for multi-agent communication. In *Procs. of Normative Multiagent Systems Workshop at AISB'05 (NorMas'05)*, 2005.

[13] D. Makinson and L. van der Torre. Input/output logics. *Journal of Philosophical Logic*, 29:383–408, 2000.

[14] P. Pasquier and B. Chaib-draa. The cognitive coherence approach for agent communication pragmatics. In *Procs. of AAMAS '03*, pages 544–551. ACM Press, 2003.

[15] J. Searle. *Speech Acts: an Essay in the Philosophy of Language*. Cambridge University Press, Cambridge (UK), 1969.

[16] J. Searle. *The Construction of Social Reality*. The Free Press, New York, 1995.

[17] M. P. Singh. A social semantics for agent communication languages. In *Issues in Agent Communications*, LNCS 1916, pages 31 – 45. Springer-Verlag, 2000.

[18] M. Verdicchio and M. Colombetti. A logical model of social commitment for agent communication. In *Procs. of AAMAS'03*, pages 528–535. ACM Press, 2003.

[19] D. N. Walton and E. C. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, 1995.

[20] M. J. Wooldridge. Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1):9–31, 2000.