2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology

Time and Defeasibility in FIPA ACL Semantics

Guido Boella
University of Turin, Italy

Guido Governatori
NICTA, Australia

Joris Hulstijn
Free University Amsterdam, The Netherlands

Régis Riveret
University of Aberdeen, Scotland

Antonino Rotolo
University of Bologna, Italy

Leendert van der Torre
University of Luxembourg

Abstract

Inferences about speech acts are often conditional, non-monotonic, and involve the issue of time. Most agent communication languages, however, ignore these issues, due to the difficulty to combine them in a single formalism. This paper addresses such issues in defeasible logic, and shows how to express a semantics for ACLs in order to make non-monotonic inferences on the basis of speech acts.

1 Introduction

It is widely accepted that the FIPA semantics for Agent Communication Languages (ACL) has some flaws [13]. First, the semantics is expressed in terms of private mental attitudes of agents. Therefore, it cannot be verified given common assumptions about agent systems. Second, the sincerity condition assumed in FIPA may be acceptable in cooperative circumstances, but is clearly wrong for persuasion and negotiation dialogues. Two solutions for these flaws have been discussed. (1) An ACL semantics can be given in terms of social attitudes, like commitments [12, 11]. Social attitudes are maintained in public, as a kind of common data structure. (2) An ACL semantics can be given in terms of public mental attitudes instead of private ones. This preserves the characteristics of the BDI-approach that underlies the FIPA semantics, without incurring the verifiability problems. Thus, [5, 6] use the notion of common ground, [9] refers to ostensible beliefs and goals and [2, 3] introduce public roles. These solutions fail to consider other problems, which emerge when new application domains are considered: in particular, *ACL semantics* (a) do not specify how to deal with *non-monotonicity* in communicative actions; (b) mostly ignore the problem of *time*, i.e. the possibility of timing the different aspects of communication, and the *persistence* of preconditions and effects. Here below are two examples illustrating these issues.

Example 1 (Non-monotonicity) *Consider the dialogue: Seller: "Why don't you buy a Mercedes A class? It is safe." Buyer: "I heard that it may roll over." / Seller: "Yes, maybe it is not safe in that sense, but I know that you are not a driver who takes risks." The seller has to retract its assertion on the safety of the product to make a weaker one. Retraction makes inference about communicative actions non-monotonic, because it is no longer true that every set of speech acts supports the same conclusions as its subsets.*

Example 2 (Time and persistence) *Suppose Ian sells at time 2 a faulty children's chair to John, causing John's child to get hurt. It is legally presumed that when a vendor sells a product, he acts in good faith, i.e., as if he were informing the client that the product is without faults. A witness, Kay, reveals at time 3 to John that the vendor had independently informed her at time 1 that the chair was faulty. Assume that the sincerity precondition of an 'inform' act persists into the future, unless there is evidence to the contrary. Given this assumption, we have at times 2 and 3 two conflicting vendor's beliefs about the quality of the chair and two conflicting client's beliefs about the same.*

This paper shows how a temporal extension of defeasible logic (DL) can express FIPA-ACL semantics. The logic captures essential properties of speech acts. Our requirements for the logic are, besides modeling non-monotonicity and persistence, that it can formalize the key features of ACL semantics using public mental states attributed to the agents via the roles they play. The logic is illustrated by one example, i.e., FIPA's *inform* communicative act, but other FIPA constructs are easily captured.

2 Temporal Defeasible Logic

This paper provides computational-oriented approach to describe the behaviour of FIPA speech acts. We are interested in their non-monotonic character, their temporal aspects and the persistence of their preconditions and effects. [7] proposes an extension of DL [10], an efficient rule-based non-monotonic formalism, able to capture such aspects. In this section we outline how to adjust the logic presented in [7] to model FIPA speech acts.

978-0-7695-3496-1/08 $25.00 © 2008 IEEE
DOI 10.1109/WIIAT.2008.275

A theory in DL comprises a set of *facts* (represented by literals), a set of rules, partitioned in *strict rules*, *defeasible rules* and *defeaters*, and a *superiority relation*, a binary relation over the set of rules. Facts are indisputable statements. Strict rules (a strict rule is noted as $A \rightarrow b$ where A is a possibly empty set of literals) are rules in the classical sense; this means that every time the antecedent holds (i.e., all the elements of A hold) then so does the conclusions b. Defeasible rules (noted as $A \Rightarrow b$) are rules whose conclusion can be defeated by contrary evidence. Defeaters (noted as $A \rightsquigarrow b$) are rules that do not support a conclusion, but which can be used to prevent the derivation of the opposite conclusion. The superiority relation gives the relative strength of rules, and it is used in case of conflicts between rules, e.g., if r and s are rules then $r \succ s$ means that r prevails over s.

To accommodate the temporal aspects and the speech acts we enrich the language of DL with:

- a discrete totally ordered set of instants of time $\mathscr{T} = \{t_0, t_1, t_2, \dots\}$;
- a set of agents $\mathrm{Ag} = \{a, b, \dots\}$;
- a set of agent roles $\mathrm{Role} = \{r_1, r_2, \dots\}$ (given $x \in \mathrm{Ag}$ we use $r(x) \in \mathrm{Role}$ to denote the role played by x).
- the families of modal operators $\{B_x\}_{x \in \mathrm{Ag} \cup \mathrm{Role}}$ (for belief) and $\{G_x\}_{x \in \mathrm{Ag} \cup \mathrm{Role}}$ (for goal);
- a set of speech act types $ST = \{inform, promise, \dots\}$.

Based on the above elements a *modal literal* is an expression $X^t l$ (or its negation, i.e., $\neg X^t l$) where X is a modal operator, $t \in \mathscr{T}$ and l is either a literal or a modal literal. For example the expression $B_a^t G_{r(b)}^{t'} p$ means that agent a believes at t that the goal of the role played by agent b at time t' is (was, will be) p. A *speech act* is an expression $st_{i,j}(s,t)$ or the negation of it (i.e., $\neg st_{i,j}(s,t)$), where $i, j \in \mathrm{Ag} \cup \mathrm{Role}$, s is either a literal or a modal literal, $st \in ST$, and $t \in \mathscr{T}$.

A rule is an expression $lbl : A \hookrightarrow_x m$, where lbl is a unique label of the rule, A is a (possibly empty) set of speech acts and modal literals, $\hookrightarrow \in \{\rightarrow, \Rightarrow, \rightsquigarrow\}$, m is a modal literal and x is either π or τ signaling whether we have a *persistent* or *transient* rule.

A persistent rule is a rule whose conclusion holds at all instants of time after the conclusion has been derived, unless interrupting events occur; transient rules establish the conclusion only for a specific instant of time. Thus $ex_1 : B_a^5 p \Rightarrow_\pi G_{r(a)}^6 q$ means that if agent a at time 5 believes p, then the role played by agent a has the goal q at time 6 and the goal continues to hold after 6 until some event overrides the goal of q. If we change π into τ the resulting rule $ex_2 : B_a^5 p \Rightarrow_\tau G_{r(a)}^6 q$ means that $r(a)$ has the goal q at time 6, but we do not know whether the goal will persist after 6.

DL has a constructive proof theory where conclusions are labelled. The conclusion labels (also called proof tags) indicate the strength as well as the type of the conclusions. Specifically we have the following proof tags:

- $+\Delta^\pi X^t p$ (resp. $+\Delta^\tau X^t p$) we have a definite proof (i.e., a proof using only strict rules and facts) of $X p$ at time t and $X p$ continues to hold after t (resp. we have a definite proof of $X p$ at time t);
- $+\partial^\pi X^t p$ (resp. $+\partial^\tau X^t p$) we have a defeasible proof of $X p$ at time t and $X p$ continues to hold after t until $X p$ is terminated (resp. we have a defeasible proof of $X p$ at time t).

For each proof tag we also have the negative version (e.g., $-\Delta^\tau X^t p$) meaning that we can show that it is not possible to build a derivation with the given features.

Definite transient derivations (i.e., conclusions tagged with Δ^τ) are just standard derivations in DL using facts, strict rules and modus ponens. For definite persistent derivations we have an additional clause, that the conclusion was obtained at a previous time and was tagged as a definite persistent. Thus from $+\Delta^\pi B_a^{10} p$ we can conclude $+\Delta^\pi B_a^t p$ for $t > 10$.

A defeasible transient derivation consists of three phases. (1) We have to give an argument for the conclusion we want to prove (and the time we want to prove), thus we need an applicable transient rule whose conclusion has the required features. A rule is applicable if all the elements in the antecedent are provable. (2) We have to consider all possible counterarguments. This means that we have to consider all rules whose conclusion is in conflict with the conclusion we want to prove. (3) We have to rebut the counterarguments. We have two possibilities: (i) we can show that the counterargument does not hold, i.e., some of its premises do not hold; or (ii) we can show that the counterargument is weaker than an applicable argument for what we want to prove.

For defeasible persistent derivations we have an additional case: we can show that the conclusion was proved at a previous instant and that all possible arguments to terminate the validity of the conclusion from the time it was proved to the time of the current derivations either were not valid, or were weaker than the argument supporting the conclusion we want to prove. For the details see [7][1].

Proposition 1 *[8] The extension of a defeasible theory from time t_0 to time t can be computed in time linear to the size of the theory, i.e., $O(|\mathrm{Lit}| \cdot |R| \cdot t)$, where* Lit *is the set of literals in the theory, R is the set of rules in the theory.*

3 Reasoning about FIPA Semantics

We now show how DL formalism can express the semantics for ACLs. For ease of reference, let us focus on FIPA's *inform* and take [4] as a starting point. In FIPA, the meaning of communicative acts is defined in terms of rational effects

[1] In [7] literals are temporalised, i.e., they have the form l^t, while here we have that the modal operators are temporalised but not the literals. However, the temporalisation of the modal operators in this paper can be considered as two disjoint copies of the systems presented in [7].

(REs) and feasibility preconditions (FPs). The REs are the mental states the speaker wants to bring about in the hearer, and the FPs encode the appropriate conditions for issuing a communicative act. For instance, here is the FIPA definition of the *inform* communicative act:

$\langle a, inform(b,p)\rangle$ FP: $B(a,p) \wedge \neg B(a, B(b,p) \vee B(b,\neg p))$
RE: $B(b,p)$

As a feasibility precondition speaker a must believe what he says and he must not believe that hearer b already has an opinion on the conveyed proposition. The rational effect is that hearer b comes to believe p.

Operators like this can be used to generate a dialogue directly, but they can also be used in the interpretation of the utterances of the interlocutor. In FIPA, this methodology relies on axioms ([4], Properties 4 and 5) according to which, when a communicative act is executed, its FPs are assumed to be true, and its RE is wanted by the speaker[2]:

$B(a, done(act) \rightarrow FP(act))$
$B(a, done(act) \wedge agent(b,act)) \rightarrow G(b, RE(act)))$

FIPA, like most other ACL semantics, does not specify some important aspects of communicative acts, such as temporal persistence and non-monotonicity. As regards persistence, in particular, the case of *inform* is instructive: hearer b can infer not only that the precondition that a's belief in the truth of p holds at the moment of execution of the communicative act, but also that this precondition persists into the future or that it held some time before the communicative act, and will hold afterwards.

Note that we adopt [1, 2]'s *role-based* approach to ACL: mental states are publicly attributed to dialogue participants, and can only change according to the rules of the dialogue. In the rules below beliefs and goals are attributed to roles (e.g., $B_{r(a)}$) as well as to individual agents (e.g., B_a).

Rules $R_{inf} = \{i_1, i_3, ..., i_{13}\}$ define the meaning of an *inform* communicative act, for a standard type of cooperative dialogue, like information exchange, where one agent is supposed to know more than the other. Adding also rule i_2 allows to capture situations where the knowledge is more symmetric. In what follows, a, b are agents, $r(a)$ and $r(b)$ the role-playing-agents in the dialogue, *inform* is a speech act type, s is either a literal or a modal literal, and $t < t'$ are time points in $\mathscr{T}$. Rules are prioritized as follows: $\prec_{inf} = \{i_2 \succ i_1, i_9 \succ i_8, i_9 \succ i_{10}, i_{11} \succ i_{10}, i_{13} \succ i_{12}\}$.

Rule i_1 describes how an *inform* act is performed by an agent a through an utterance event. The rule is defeasible since the communicative act can be retracted later, as indicated in rule i_2. There are different ways of handling retraction. The solution is to withdraw the original communicative act by means of a defeater. This means that all the consequences that can be inferred from the act, expressed in rules $i_3 - i_6$, are also withdrawn. If we would take the alternative solution of only retracting the content of the *inform*, we would need additional explicit rules to withdraw those consequences too.

i_1 $utter_{a,b}(inform_{r(a),r(b)}(s,t),t) \Rightarrow_\tau inform_{r(a),r(b)}(s,t)$
i_2 $retract_{a,b}(inform_{r(a),r(b)}(s,t),\, t') \rightsquigarrow_\tau \neg inform_{r(a),r(b)}(s,t)$

The agent of actions *utter* and *retract* is the individual agent a and not its role $r(a)$. The rules i_1 and i_2 are used to connect individual agents to their roles. These rules are non-persistent, since the action of uttering only temporally coincides with execution of an *inform* communicative act. If the agent makes an inform at a given time, it is not possible to infer that it is making the inform again at the next time instant.

The defeasible character of i_1 is necessary in applications where an agent is trying to persuade the other one and the other one can reply (see Example 1), or where two agents are trying to cooperatively find a solution and to advance arguments which are probed by the other one. In these types of dialogues the speaker can retract his assertions to avoid a contradiction. In information gathering scenarios like the ones considered by FIPA this is not necessary, since an agent is supposed to know more than the other.

Rules $i_3 - i_5$ represent the FPs of *inform*. Following Properties 4 and 5 of [4], they are interpreted as strict rules. Only rule i_3 is persistent towards the future, since its effect is not affected by the consequent of other rules. Instead, since the *inform* possibly changes the beliefs of the hearer b, $i_4 - i_5$ are not persistent and they refer to the situation before the execution of the speech acts[3]:

i_3 $inform_{r(a),r(b)}(s,t) \rightarrow_\pi B^t_{r(a)}\, s$
i_4 $inform_{r(a),r(b)}(s,t) \rightarrow_\tau \neg B^{t-1}_{r(b)}\, s$
i_5 $inform_{r(a),r(b)}(s,t) \rightarrow_\tau \neg B^{t-1}_{r(b)} \sim s$

Note that the beliefs in the consequent of the rules are not attributed to the agents as private beliefs, but to the roles they play: thus they have a public character[4].

Rule i_6 represents the RE of *inform*: its propositional content is embedded in a goal of the speaker that the hearer believes it:

i_6 $inform_{r(a),r(b)}(s,t) \rightarrow_\pi G^t_{r(a)} B^t_{r(b)}\, s$

FIPA does not allow explicit inferences about the success of the RE, but in our model of cooperative information exchange R_{inf}, rule i_7 can represent that the hearer publicly adopts the information conveyed, if he believes that the speaker is reliable:

i_7 $G^t_{r(a)} B^t_{r(b)} s, B^t_{r(b)} reliable(a) \rightarrow_\pi B^t_{r(b)}\, s$

[2] In FIPA notation, *act* stands for any action, $done(act)$ is the proposition that expresses completion of *act*, and $agent(b,act)$ represents that b is the agent who executes action *act*.

[3] In i_5, $\sim s$ stands for the complement of s. If s is a positive literal p then $\sim s$ is $\neg p$; if s is $\neg p$, then $\sim s$ is p [7].

[4] The conclusions of FP rules could also persist from the past. We do not have the space to discuss here this issue.

This does not necessarily mean that the hearer privately believes what was said. Only if there is no evidence to the contrary, we assume that individual agents believe what their roles believe ($i_8 - i_{13}$). Rule i_8 assumes that the speaker individually believes what he says, unless he is believed to be insincere (i_9). Rule i_{10} assumes that a hearer believes what has been asserted, unless he is believed not to be a trusting character (i_{11}). Rule i_{12} assumes sincerity for goals in a similar way. In all these cases, the cooperative behavior is the default, but it can be overruled by evidence to the contrary. Hence we have $i_9 \succ i_8$, $i_{11} \succ i_{10}$ and $i_{13} \succ i_{12}$.

i_8 $\quad inform_{r(a),r(b)}(s,t) \Rightarrow_\pi B^t_a s$

i_9 $\quad inform_{r(a),r(b)}(s,t), B^t_{r(b)} \neg sincere(a) \rightsquigarrow_\pi \neg B^t_a s$

i_{10} $\quad inform_{r(a),r(b)}(s,t), B^t_{r(b)} s \Rightarrow_\pi B^t_b s$

i_{11} $\quad inform_{r(a),r(b)}(s,t), B^t_{r(b)} s, B^t_{r(a)} \neg trusting(b) \rightsquigarrow_\pi \neg B^t_b s$

i_{12} $\quad inform_{r(a),r(b)}(s,t), G^t_{r(a)} B^t_{r(b)} s \Rightarrow_\pi G^t_a B^t_b s$

i_{13} $\quad inform_{r(a),r(b)}(s,t), B^t_{r(b)} \neg sincere(a) \rightsquigarrow_\pi \neg G^t_a B^t_a s$

Strict inferences about REs and FPs of communicative acts are only possible regarding the public beliefs or goals of the participants, while inferences about the private mental states of the participants can be made only by default. Sincerity, trust and cooperativity are the assumptions to pass information from the public level to the private one.

Example 3 (Time and persistence; cont'd) *Consider the agents i (Ian), j (John) and k (Kay) and their roles $v(i)$ (vendor), $c(j)$ (client) and $w(k)$ (witness). Literal s means "the chair is without faults". The theory contains R_{inf}, $\succ_{inf}$, plus the following facts and additional rules:*

Facts: $utter_{i,k}(inform_{v(i),c(j)}(\neg s,1),1)$, $sell(i,j,2)$, $B^3_{c(j)}(reliable(k))\}$, $utter_{k,j}(inform_{w(k),c(j)}(\neg s,3),3)$

Rules: $r_1 : sell(i,j,2) \Rightarrow^\tau inform_{v(i),c(j)}(s,2)$

$r_2 : sell(i,j,2) \Rightarrow^\pi B^2_{c(j)}(reliable(i))$.

r_1 states that when $v(j)$ sells the chair at 2, he acts as if he were informing $c(j)$ that the chair is without faults; r_2 assumes that $c(j)$ defeasibly takes $v(i)$ as reliable. Consider the following relevant derivations

Time	Conclusion & Argument
1	1. $+\partial^\tau inform_{v(i),w(k)}(\neg s,1)$ (a1): Facts, i_1
1	2. $+\partial^\pi B^1_{v(i)} \neg s$ (a2): 1, i_3
2	3. $+\partial^\tau inform_{v(i),c(j)}(s,2)$ (a3): Facts, r_1
2	4. $+\partial^\pi B^2_{c(j)}(reliable(i))$ (a4): Facts, r_2
2	5. $+\partial^\pi G^2_{v(i)} B^2_{c(j)} s$ (a5): 3, i_6
2	6. $+\partial^\pi B^2_{c(j)} s$ (a6): 4, 5, i_7
2	7. $-\partial^\pi B^2_{v(i)} \neg s$ (a7): conflict between (a2) and 3, i_3
2	8. $-\partial^\pi B^2_{v(i)} s$ (a8): conflict between (a2) and 3, i_3
3	9. $+\partial^\tau inform_{w(k),c(j)}(\neg s,3)$ (a9): Facts, i_1
3	10. $+\partial^\pi G^3_{w(k)} B^3_{c(j)} s$ (a10): 9, i_6
3	11. $-\partial^\pi B^3_{c(j)} \neg s$ (a11): conflict between (a6) and Facts, 10, i_7
3	12. $-\partial^\pi B^3_{c(j)} s$ (a12): conflict between (a6) and Facts, 10, i_7

For conclusions 7-8 and 11-12, due to the persistence of $B^1_{v(i)} \neg s$ and $B^2_{c(j)} s$, these beliefs should also hold, respectively, at 2 and 3. But they are defeated by opposite arguments, which are in turn defeated by the former ones. (DL is a sceptical non-monotonic formalism: with two conflicting defeasible conclusions DL refrains to take a decision.) 7-8 show that FPs of some inform acts are violated, 11-12 that some REs are not successful.

4 Summary

We used DL to study non-monotonicity and time in role-based agent communication. Non-monotonicity occurs in reasoning about the persistence of FPs and REs of speech acts. Whereas FIPA makes strong assumptions about the private states, the alternative of using public mental attitudes does not make any assumptions about them; using non-monotonic reasoning we can make inferences about the private mental attitudes of the agents which hold only by default and can always be revised. Finally, non-monotonicity can be used for challenges, concessions, and retractions. E.g., an inform is accepted –its content becomes part of the public beliefs of the addressee– unless it is challenged.

References

[1] G. Boella, R. Damiano, J. Hulstijn, and L. van der Torre. Public mental attitudes for ACL semantics. In *Procs. AC'06*. 2006.

[2] G. Boella, R. Damiano, J. Hulstijn, and L. van der Torre. Role-based semantics for agent communication: Embedding of the mental attitudes and social commitments semantics. In *Procs. AAMAS'06*. 2006.

[3] G. Boella, R. Damiano, J. Hulstijn, and L. van der Torre. A common ontology of agent communication languages: Modelling mental attitudes and social commitments using roles. *Applied Ontology*, 2007.

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

[5] B. Gaudou, A. Herzig, and D. Longin. A logical framework for grounding-based dialogue analysis. In *Procs. of LCMAS'05*. 2006.

[6] B. Gaudou, A. Herzig, D. Longin, and M. Nickles. A new semantics for the FIPA agent communication language based on social attitudes. In *Procs. ECAI'06*. 2006.

[7] G. Governatori, A. Rotolo, and G. Sartor. Temporalised normative positions in defeasible logic. In *Procs. ICAIL05*, 2005.

[8] G. Governatori and P. Terenziani. Temporal extensions to defeasible logic. In *Procs. Aust. AI 2007*, 2007.

[9] M. Nickles, F. Fischer, and G. Weiss. Communication attitudes: A formal approach to ostensible intentions, and individual and group opinions. In *Procs. LCMAS'05*. 2006.

[10] D. Nute. Defeasible logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming*. 1994.

[11] M. P. Singh. A social semantics for agent communication languages. In *Procs. IJCAI Workshop on ACLs*. 2000.

[12] M. Verdicchio and M. Colombetti. From message exchanges to communicative acts to commitments. In *Procs. LCMAS'05*. 2006.

[13] M. J. Wooldridge. Semantic issues in the verification of agent communication languages. *JAAMAS*, 3(1):9–31, 2000.