FIPA Communicative Acts in Defeasible Logic

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

Guido Governatori
University of Queensland
guido@itee.uq.edu.au

Joris Hulstijn
Vrije Universiteit
jhulstijn@feweb.vu.nl

Regis Riveret
Università di Bologna
rriveret@cirsfid.unibo.it

Nino Rotolo
Università di Bologna
rotolo@cirsfid.unibo.it

Leendert van der Torre
University of Luxembourg
leon.vandertorre@uni.lu

Abstract

In agent communication languages, the inferences that can be made on the basis of a communicative action are inherently conditional, and non-monotonic. For example, a proposal only leads to a commitment, on the condition that it is accepted. And in a persuasion dialogue, assertions may later be retracted. In this paper we therefore present a defeasible logic that can be used to express a semantics for agent communication languages, and to efficiently make inferences on the basis of communicative actions. The logic is non-monotonic, allows nested rules and mental attitudes as the content of communicative actions, and has an explicit way of expressing persistence over time. Moreover, it expresses that mental attitudes are publicly attributed to agents playing roles in the dialogue. To illustrate the usefulness of the logic, we reformalize the meta-theory underlying the FIPA semantics for agent communication, focusing on inform and propose. We show how composed speech acts can be formalized, and extend the semantics with an account of persuasion.

1 Introduction

Whereas FIPA speech act syntax (FIPA [2002]) is widely used in agent programming and communication languages, the FIPA semantics based on mental attitudes has not been adopted by the agent community at large. For example, if agent *a* informs agent *b* that there is a price reduction, then the FIPA semantics implies that agent *a* believes that there is such a price reduction, together with various other conditions. It is generally accepted that such assumptions hinder the use of FIPA semantics in various ways, see, e.g., Wooldridge [2000]. First, FIPA semantics can not be verified given common assumptions about open agent systems. Second, the sincerity condition that is assumed in FIPA may be acceptable in cooperative circumstances, but is clearly wrong for persuasion and negotiation dialogues.

During the last years, two ways of dealing with the flaws in the semantics have been discussed. First, it has been suggested that FIPA speech act semantics can be reinterpreted in terms of social attitudes, for example Verdicchio and Colombetti [2006] model FIPA speech acts using the notion of commitment introduced by Singh [2000]. Second, other works reinterpret FIPA semantics in terms of public mental attitudes instead of private ones, trying to preserve its main characteristics. Gaudou *et al.* [2006a,b] use the notion of common ground, Nickles *et al.* [2006] refer to ostensible beliefs and goals and Boella *et al.* [2006a,b] introduce roles. This substitution makes it possible to substantially maintain FIPA semantics without incurring in the verifiability problems.

However, these solutions do not address two other problems of FIPA's semantics:

- FIPA semantics does not specify how an agent adopting it has to deal with the many non-monotonic aspects inherent in communicative actions.
- FIPA semantics does not consider the rule-based character of communication and the conditional nature of mental attitudes of agents using it.

In an agent communication many inferences can be made only in absence of specific information and tentatively, because they can later be retracted or revised when new pieces of information become available. In particular, FIPA does not specify how agents have to deal with the persistence of preconditions and effects of speech acts, and with retraction. Persistence must be intended in two senses. Effects of speech acts persist in the future. However, they persist only as far as the are not cancelled. For example a goal can be cancelled because it has been satisfied, or a belief can be retracted. Inferences about the preconditions persist also towards the past, since they are observations and not caused by the speech act. E.g., it is possible to infer that the speaker believes what he asserted, and, unless other information are available, he believed this also in the past and keeps believing it.

Retraction of speech acts is another non-monotonic mechanism disregarded by FIPA since the standard is focused to cooperative dialogues only, and not towards persuasion or negotiation dialogues.

Last, strict inferences about effects and preconditions of speech acts are only possible in the public beliefs or goals of the participants, while in the real private beliefs or goals of the participants on the dialogue can be made only by default, under the assumption of cooperativity or sincerity.

Concerning the rule-based character of dialogue, consider the semantics of a proposal. When agent *a* proposes a contract α to agent *b*, this means that, if agent *b* also publically adopts the goal α, i.e. accepts the proposal, then agent *a* will adopt this goal too. So when agent *b* accepts the proposal, by informing agent *a* that it has adopted α as a goal, then agent *a* will be publically bound to this goal. We use nested rules to describe how a communicative act implies such a rule.

We therefore introduce in this paper a defeasible logic to reason about the speech act semantics to study the *non-monotonic* and *rule-based* character of role based communicative actions. The defeasible logic combines a variety of recent results in defeasible logic to incorporate multiple agents, nested modalities, time with persistence (Governatori *et al.* [2005]), nested conditionals (Song and Governatori [2005]) and the introduction of roles in a dialogue game. At the same time defeasible logic aims to respect the linear complexity of the basic defeasible logic, so we can consider the resource constraints of agents which use FIPA semantics.

Our requirements for a logic of role based agent communication are that it can formalize the essential features of FIPA semantics using public mental states, including composed actions, it can be extended to reason about different assumptions about agents in cooperative dialogues, like information exchange, and non-cooperative dialogues, like persuasion and negotiation.

The layout of this paper is as follows. In Section 2 we introduce the defeasible logic we use. In Section 3 we formalize and extend FIPA speech acts, and we show how to reason about the agent's mental attitudes.

2 Defeasible Logic

This section proposes an extension of Defeasible Logic (DL) (Antoniou *et al.* [2000]) to deal with FIPA dialogues. The next section illustrates the definitions by examples in role based agent communication. We keep the basic distinction among strict and defeasible rules, and defeaters. We have that $\phi \rightarrow \psi$ is a *strict rule* such that whenever the premises ϕ are indisputable so is the conclusion ψ. $\phi \Rightarrow \psi$ is a *defeasible rule* that can be defeated by contrary evidence. $\phi \leadsto \psi$ is a *defeater* that is used to defeat some defeasible rules by producing evidence to the contrary. We extend the language with temporal dimension, mental states (beliefs B and goals G), one level of nested rules, and speech acts.

Definition 1 (Language) *Let* PROP *be a set of propositional atoms,* AA *a set of atomic acts,* Ag *and* Role *the finite sets of agents and roles, and* $\mathscr{T}$ *a discrete totally ordered set of instants of time* $\{t_1, t_2, \dots\}$. MOD, TPROP, *etc., are the smallest sets closed under the following rules:*

modal operators $\mathrm{MOD} = \bigcup_{i \in \mathrm{Ag} \cup \mathrm{Role}} \{\mathrm{B}_i, \mathrm{G}_i\}$;

temporal prop. atoms $\mathrm{TPROP} = \{p : t | p \in \mathrm{PROP}, t \in \mathscr{T}\}$;

temporal literals $\mathrm{Lit} = \mathrm{TPROP} \cup \{\neg p | p \in \mathrm{TPROP}\}$.

modal temporal literals *for any* $X_i \in \mathrm{MOD}$, *(1)* $(X_i l) : t, \neg(X_i l) : t \in \mathrm{ModLit}$ *for* $l \in \mathrm{Lit}$; *(2) if* m *in* ModLit, *then* $X_i(m) : t, \neg X_i(m) : t \in \mathrm{ModLit}$.

If $q : t$ *is a temporal literal,* $\sim q : t$ *denotes the* complementary literal *(if* $q : t$ *is a positive literal* $p : t$ *then* $\sim q : t$ *is* $\neg p : t$*; and if* $q : t$ *is* $\neg p : t$*, then* $\sim q : t$ *is* $p : t$*). If* p *is modal temporal literal* $X_i(q) : t$*, then the set of its complementary literals* $p^{\sim}$ *is the smallest set closed under the following rules: (1)* $\neg X_i(q) : t, X_i(\neg q) : t \in p^{\sim}$ *for* $q \in \mathrm{Lit}$*; (2) if* m *in* ModLit, *then* $\neg X_i(m) : t, X_i(\neg m) : t \in p^{\sim}$*; in this case* $\sim p$ *will denote any element of the set* $p^{\sim}$*. For readability we also write* p^t *for* $p : t$ *and* $X_i^t(m)$ *for* $X_i(m) : t$.

speech act types $ST = \{inform, promise, \dots\}$

temporal speech acts *if* $st \in ST$, $i, j \in \mathrm{Ag} \cup \mathrm{Role}$, $s \in \mathrm{Rul}$, $t \in \mathscr{T}$, *then* $st_{i,j}(s,t) \in \mathrm{SA}$

acts $\mathrm{ACTS} = \mathrm{AA} \cup \mathrm{SA} \cup \{\neg a \mid a \in \mathrm{AA} \cup \mathrm{SA}\}$

atomic persistent rules $\mathrm{Rul}_{atom,\pi}$ *is:*

$$\mathrm{Rul}_{atom,\pi} = \{(\phi \rhd_\pi \psi) : t | \phi \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ACTS}, \psi \in \mathrm{Lit} \cup \mathrm{ModLit}, \rhd \in \{\rightarrow, \Rightarrow, \leadsto\}, t \in \mathscr{T}\}$$

atomic transient rules

$$\mathrm{Rul}_{atom,\tau} = \{(\phi \rhd_\tau \psi) : t | \phi \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ACTS}, \psi \in \mathrm{ACTS} \cup \mathrm{Lit} \cup \mathrm{ModLit}, \rhd \in \{\rightarrow, \Rightarrow, \leadsto\}, t \in \mathscr{T}\}$$

atomic rules Rul_{atom} *is* $\mathrm{Rul}_{atom,\pi} \cup \mathrm{Rul}_{atom,\tau}$.

non-modal rules $\mathrm{NRul} = \mathrm{Rul}_{atom} \cup \{\neg r | r \in \mathrm{Rul}_{atom}\}$

modal rules *for any* $X_i \in \mathrm{MOD}$, *(1)* $X_i r, \neg X_i r \in \mathrm{MRul}$ *for* $r \in \mathrm{NRul}$*; (2) if* r *in* MRul, *then* $X_i r, \neg X_i r \in \mathrm{MRul}$*. The definitions of complementary rules and complementary modal rules is similar to those for complementary literals and complementary modal literals.*

rules $\mathrm{Rul} = \mathrm{NRul} \cup \mathrm{MRul}$.

meta-rules Rul^C*:*

$$\mathrm{Rul}^C = \{\phi \rhd_\pi \psi : t | \phi \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ACTS} \cup \mathrm{Rul}, \psi \in \mathrm{Rul}, \rhd \in \{\rightarrow, \Rightarrow, \leadsto\}, t \in \mathscr{T}\}$$

abbreviations*: if* r *is any rule,* $A(r)$ *denotes the antecedent of* r *while* $C(r)$ *denotes its consequent. Other abbreviations are used, such as subscript for type of rule (persistent vs transient) or for the strength (strict, defeasible, defeater), and* $R[\phi]$ *for rules whose consequent is* ϕ*; for example:*

$\mathrm{Rul}_s = \{\phi \rightarrow \psi \in \mathrm{Rul}\}$ $\quad \mathrm{Rul}_d = \{\phi \Rightarrow \psi \in \mathrm{Rul}\}$
$\mathrm{Rul}_{sd} = \{\phi \rhd \psi \in \mathrm{Rul} | \rhd \in \{\rightarrow, \Rightarrow\}\}$ $\quad \mathrm{Rul}_{dft} = \{\phi \leadsto \psi \in \mathrm{Rul}\}$
$\mathrm{Rul}_\pi[\psi] = \{r \in \mathrm{Rul}_\pi | C(r) = \psi\}$

For readability reasons sometimes we will use $\phi \rhd^t \psi$ *instead of* $\phi \rhd \psi : t$.

Remark 1 The language is built up by introducing different basic components: *temporal literals*, *modal temporal literals*, and *temporal speech acts*. In all cases, the components are labelled by time instants. Facts, such as $It\_Rains : t$ is an example of temporal literal, where t is a time instant. Modal literals have two temporal dimensions, as we want to express the fact that an agent can believe (or have a goal) at a certain time: $\mathrm{B}_i^t\, It\_Rains : t'$, for example, says that agent i believes at t that it rains at t'. Speech acts, too, embed temporal dimensions, the time when the speech act is performed plus the time of the literal in the scope of the speech act operator.

We have different types of rule that we can build using the mentioned basic components. Besides the usual distinction among defeasible, strict rules, and defeaters, we can work with *persistent* and *transient* rules. In addition, rules can be temporalized; the temporal parameter t applied to an entire rule r, such as in the case $r : (a : t' \Rightarrow b : t'') : t$, indicates the time from when such a rule holds. Persistent rules, whose arrow is labelled by π, are such that the conclusions we derive from them persist over time until the opposite conclusion blocks the derivation: $r : (It\_Rains : t' \rightarrow_\pi B_a^t\, It\_Rains : t') : t''$, for example, says that, if it rains at t', then it is a definite conclusion that agent a believes at t that it rains at t' and this belief persists for every time instant next to t; such a rule r holds from t'' onwards. Transient rules, labelled by τ, state conclusions co-occurrent with their premises: $r' : (Church : t \Rightarrow_\tau G_a^t[Pray : t]) : t'$, which says that when in a church at t, agent a, defeasibly, has the goal at t to pray at t, being the rule r' valid from t' onwards. When rules occur within the scope of a modal operator we obtain *modal rules*: $\mathrm{B}_b^{t'}(It\_Rains : t \Rightarrow_\tau G_a^t[Open\_Umbrella : t])$, which states that agent a believes at t' that, if it rains at t, agent b has the goal at t to open at that time the umbrella.

Finally, the language permits to introduce meta-rules, namely, rules whose consequents are other rules among those we mentioned above. For example: $r : (inform_{a,b}(It\_Rains : t' \Rightarrow G_a Open\_Umbrella, t) \rightarrow_\pi (B_a^t[It\_Rains : t'] \Rightarrow G_a Open\_Umbrella) : t) : t''$, which states the from the point of view t'' if agent a informs b about rule, then from time t, the rule $B_a^t[It\_Rains : t'] \Rightarrow G_a Open\_Umbrella)$ will hold from time t. For readability, when rules have no external temporal parameter this means that they hold from t_0 onwards.

Definition 2 (Defeasible ACL theory) *A defeasible ACL theory for some dialogue type* d*, such as information seeking, is a tuple* $(\mathscr{T}, F, \mathrm{Ag}, \mathrm{Role}_d, r, R_d, R_d^C, \succ_d)$*, where*

- $\mathscr{T}$ *is a discrete totally ordered set of times* $\{t_1, t_2, \dots\}$;
- $F \subseteq \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{ACTS}$ *is a finite set of facts,*
- $\mathrm{Ag} = \{a, b, ..\}$ *is a set of agents,* $\mathrm{Role}_d = \{r_1, r_2, ..\}$ *a set of roles, and r is a dialogue instance which determines a function* $r : Ag \mapsto Role_d$,
- $R_d \subseteq \mathrm{Rul}$ *and* $R^C_d \subseteq \mathrm{Rul}^C$ *are the rules for d, and*
- $\succ_d \subseteq (\mathrm{Rul} \times \mathrm{Rul}) \cup (R^C \times R^C)$ *the priority relation for d, which is an acyclic binary relation over the set of rules.*

2.1 Proof theory

DL is based on a constructive inference mechanism based on tagged conclusions. A tagged conclusion has one of the following (basic) forms

- $+\Delta p$ ($-\Delta p$), meaning that we have (we can show that we do not have) a definite derivation for p; a definite derivation is derivation obtained using only facts and strict rules.
- $+\partial p$ ($-\partial p$), meaning that we have (we can show that we do not have) a defeasible derivation of p.

Rules are partitioned into persistent and transient rules according to whether, respectively, the consequent persists until an interrupting event occurs, or it is temporally co-occurrent with the premises. Thus, proof tags ∂ and Δ can be labelled with π or τ. Second, since rules can be temporalized (indicating the time of their validity, their "viewpoint"), the conclusions we infer from them can be temporalized as well in this sense: using an applicable rule holding at t, we write, e.g., $+\partial@t\ q$ meaning that q is defeasibly provable with viewpoint t, i.e., using the rules that are derivable (in force or effective) at t. When proof tags are not labelled, we generically refer to any type of proof tag. Analogously, when no viewpoint is specified in proof tags, the conclusion has been obtained with an "untimed" rule, which implicitly holds at any times.

Definition 3 (Derivation) *Given a defeasible ACL theory D, a derivation P from D is a finite sequence* $P(1), \dots, P(n)$ *of tagged conclusions* $\pm\#p$, *where* $\# \in \{\Delta, \partial\}$, $p \in \mathrm{Lit} \cup \mathrm{ModLit} \cup \mathrm{Rul} \cup \mathrm{ACTS}$, *and where each* $P(i) = \pm\#p$, $1 \le i \le n$ *satisfies the proof conditions (based on the elements of D) given in the rest of this section.*

With $P(1..n)$ we denote the initial subsequence P of length n. As usual we will use $D \vdash \pm\#p$ to denote that there is a derivation P from D of $\pm\#p$.

Definition 4 *If* $D = (\mathscr{T}, F, \mathrm{Ag}, \mathrm{Role}, R, R^C, \succ)$ *is an ACL theory, the rule-range is* $\mathscr{R} = R \cup R'$ *of D such that* $R' = \{r | r \in C(r'), r' \in R^C\}$.

Definition 4 provides the rule-range which is considered when we prove a literal or a rule. For example, when we defeasibly prove a literal, we have to find an applicable rule r whose consequent is such a literal, and check that all rules that provide the opposite conclusion are defeated or cannot be applied. These rules, r and the "attacking rules", can be in theory from the beginning (in R), or can be derived using appropriate meta-rules. $\mathscr{R}$ is the domain that include them.

Definition 5 *Let* $i, j \in \{a, b, r(a), r(b)\}$, $\# \in \{\Delta, \partial\}$, *and* $P = (P(1), \dots, P(n))$ *be a proof in D.*

- *A rule* $r : t \in \mathscr{R}$ *is* #-applicable-at-t' *in P iff*
  - $r : t \in \mathscr{R}_{atom}$, *and*
  - $+\#^C_\pi@t'r : t \in P(1..n)$, *and*
  - $\forall p \in A(r), +\#@t'p \in P(1..n)$;
- *A rule* $r : t \in \mathscr{R}$ *is* #-discarded-at-t' *in P iff either*
  - $r : t = \sim r' : t$ *for some* $r' : t \in R$, *or*
  - $r : t \in \mathrm{MRul}$, *or*
  - $-\#^C_\pi@t'r : t \in P(1..n)$, *or*
  - $\exists p \in A(r)$ *such that* $-\#@t'p \in P(1..n)$.

Definition 5 states the conditions of applicability for basic rules (i.e., rules which are not meta-rules). A rule r is applicable at a time t' if: (1) r does not have the form $\neg(a \rhd b)$; the negation of a rule cannot be used to derive any literal; (2) r is derivable at t, namely it not defeated, otherwise it cannot be used; (3) all antecedents of r must be provable. If these conditions are not met, r is not applicable (it is discarded) at t'. An additional condition for discarding r is that it has the form $X(a \rhd b)$ ($X \in \{\mathrm{B}, \mathrm{G}\}$), namely, it is modalized: only when we admit –as we do in the next sections– that rules such as $B_{r(a)}[p \to G_{r(a)}q^t]$ can be transformed into $B_{r(a)}p \to G_{r(a)}q^t$ we can apply them (see Definition 8).

Let us focus on proof conditions. For space reasons, we provide only proof conditions for positive persistent conclusions. Conditions for transient conclusions roughly follow the inference patterns of standard DL. The negative proof condition can be obtained from the positive one by applying the principle of strong negation (Antoniou *et al.* [2000]), i.e., the negative proof conditions are constructive conditions showing that it is not possible to satisfy the corresponding positive proof condition.

The inference conditions for positive persistent definite proofs of literals and modal literals are as follows.

$+\Delta_\pi$: If $P(n+1) = +\Delta_\pi@t'q : t$, then
(1) $q : t \in F$; or
(2) $\exists r \in \mathscr{R}_{\pi,s}[q : t]$: r is Δ-applicable-at-t'; or
(3) $\exists t'' \in \mathscr{T}$:$t' < t$, $+\Delta_\pi@t'q : t'' \in P(1..n)$; or
(4) $\exists t''' \in \mathscr{T}$: $t''' < t$, $+\Delta_{pi}@t'''q : t \in P(1..n)$.

Remark 2 A temporal literal or a temporal modal literal $q : t$ is strictly provable with the viewpoint t' in a theory when (1) $q : t$ is a fact (and, as such, it is assumed to hold for every viewpoint); (2) strict rules in D can be applied one after the other and the chain of reasoning arrives at a persistent rule which is applicable at t' and that has as a consequent $q : t$; (3) $q : t''$ is strictly and persistently provable with the viewpoint t'; since q is persistent, it holds for any instants subsequent to t'', and so at t; (4) $q : t$ has been proved for a viewpoint t'' preceding the current viewpoint t'. Here is an example of the last two cases:

$$
\begin{aligned}
F &= \{a : t_2\} \\
R &= \{r_1 : (a : t_2 \to_\tau b : t_1) : t_1 \\
&\quad\ r_2 : (b : t_1 \to_\pi q : t_3) : t_1\} \\
&\quad\ t_3 < t,\ t_1 < t
\end{aligned}
$$

We want to know whether $+\Delta@tq : t$, i.e., whether q holds at t, when we consider the evidence and rules that hold at t (the time at which we consider the derivation, $@t$). The fact makes r_1 applicable and so we obtain $+\Delta_\tau@t_1\ b : t_1$. This makes r_2 applicable and, since r_2 is persistent and $t_3 \le t$, $+\Delta_\pi@t_1\ q : t$, and then $+\Delta_\pi@t\ q : t$. Hence, $+\Delta@t\ q : t$.

The inference conditions for positive persistent defeasible proofs are as follows.

$+\partial_\pi$: If $P(n+1) = +\partial_\pi@t_xq : t$ then
(1) $+\Delta_\pi@t_xq : t \in P(1..n)$, or
(2) $-\Delta@t_x \sim q : t \in P(1..n)$, and
  (2.1) $\exists r : t_r \in \mathscr{R}_{\pi,sd}[q : t]$: r is ∂-applicable-at-t_x, and
  (2.2) $\forall s : t_s \in \mathscr{R}[\sim q : t]$: if $+\partial^C_\pi@t_x\ s : t_s \in P(1..n)$, then either
    (2.2.1) $s : t_s$ is ∂-discarded-at-t_x; or
    (2.2.2) $\exists w : t_w \in \mathscr{R}_\pi[q : t]$: w is ∂-applicable-at-t_x and $w : t_w \succ s : t_s$; or

(3) $\exists t' \in \mathscr{T}$: $t' < t$ and $+\partial_\pi @t_x\ q : t' \in P(1..n)$, and
(3.1) $\forall t'', t' < t'' \leq t, \forall s : t_s \in \mathscr{R}[\sim q : t'']$:
if $+\partial^C_\pi @t_x\ s : t_s \in P(1..n)$, then
(3.1.1) $s : t_s$ is ∂-discarded-at-t_x or
(3.1.2) $\exists v : t_v \in R_\pi[q : t']$, v is ∂-applicable-at-t_x and
$v : t_v \succ s : t_s$; or
(4) $\exists t''' \in \mathscr{T}$: $t''' < t_x$ and $+\partial_\pi @t'''q : t \in P(1..n)$, and
(4.1) $\forall t''''\ t''' < t'''' \leq t_x\ \forall s : t_s \in \mathscr{R}[\sim q : t]$:
if $+\partial^C_\pi @t_x\ s : t_s \in P(1..n)$, then
(4.1.1) s is ∂-discarded-at-t_x or
(4.1.2) $\exists v : t_v \in R^X[q : t''']$: v is ∂-applicable-at-t_x and
$v : t_v \succ s : t_s$.

Remark 3 Proof conditions for defeasible persistent literals runs as follows. In general, notice that each time a rule is used, this requires that it is applicable, which in turn requires, among other conditions, that it is derivable. Clause 1 allows us to infer a defeasible persistent conclusion from a strict persistent conclusion with the same mode. Clause 2 requires that the complement of the literal we want to prove is not definitely provable (or definitely provable for $-\partial$), but it does not specify whether it is persistent or transient: remember that what we want to achieve is to see whether the literal or its complement are provable at t but not both; in the same way, and for the same reason, q can be attacked by any rule for the complement of q (clause 2.2). An important issue in all clauses of this proof condition is that each time we have to use a rule (either to support the conclusion (2.1), to attack it (2.2.1) or to rebut the attack (2.2.2)) we must have that the rule is provable at time t of the derivation ($@t$). Clauses 3 and 4 are the clauses implementing persistence (i.e., the conclusion has been derived at a previous time and carries over to the current time). Essentially clause 3 ensures that the conclusion has been derived at a previous time t'' and no interrupting event occurred between t'' and t; while clause 4 takes care of the case where q is derived persistently for a time before t', and that no interrupting event will occur between the effectiveness of q and the time q is expected to hold according to the current derivation. Let us see a small example:

$$
\begin{aligned}
F &= \{a : t'', p : t''\} \\
R &= \{r_1 : (a : t'' \Rightarrow_\pi q : t) : t' \\
&\quad r_2 : (b : t'' \Rightarrow_\tau \neg q : t''') : t' \\
&\quad r_3 : (p : t'' \Rightarrow_\pi s : t'''') : t'''' \\
&\quad r_4 : (p : t'' \Rightarrow_\pi b : t'') : t' \\
&\quad r_5 : (s : t' \Rightarrow_\pi \neg b : t'') : t'\} \\
\succ &= \{r_5 \succ r_4\} \\
&\quad t'''' < t'
\end{aligned}
$$

The facts make applicable r_1, r_3, r_4, and r_5. r_4, in particular, would permit to make r_2 applicable, which would attack r_1, as we do not know whether it is stronger than r_2. However, r_5 is stronger than r_4, which permit to discard r_2. In fact, r_5 is applicable thanks to r_3, from which we obtain $+\partial_\pi @t''''s : t''''$. By persistency (with regard to the viewpoint and the time of the consequent s; clauses 3 and 4) we get $+\partial_\pi @t's : t'$, which, as we said, makes r_5 applicable.

The proof conditions for transient rules have the same structure as the corresponding proof conditions for permanent rules but without the persistence conditions. Namely the condition for $+\Delta_\tau$ is the same as the first two conditions of $+\Delta_\pi$, while the condition for $+\partial_\tau$ corresponds to clauses (1) to (2) of $+\partial_\pi$.

Definition 6 *Two rules $r : t, s : t' \in$ Rul are incompatible iff*

1. *$t = t'$ and*
2. *$r = \sim s$, or*
3. *$A(r) = A(s)$ and $C(r) = \sim C(s)$.*

Two modal rules $X(r) : t, Y(s) : t' \in$ MRul, such that $r, s \in$ Rul, are incompatible iff $t = t'$ and either

1. *$X = Y$, and r and s are incompatible; or*
2. *$Y = \neg Y$, and $r = r'$.*

Given two rules r and s, we write $r \preceq s$ to denote that s and r are incompatible and s is at least as strong as r, where the strength of rules is such that strict rules are stronger than defeasible rules and defeaters, and defeasible rules and defeaters have the same strength.

Definition 6 states the incompatibility conditions between rules, namely, the criteria to see when the derivations of two rules are in conflict. For non-modal rules, they are incompatible when (1) they have the same external time of validity, and, either (2) one is the negation of the other, or (3) they have the same antecedents, but one has, as its consequent, the complement of the consequent of the other. Two modal rules are incompatible when, having the same external time of validity, they have the same modality and the rules in the scope of the modal operator are incompatible, or one is the negation of the other.

Definition 7 *Let $\# \in \{\Delta, \partial\}$, and $P = (P(1), \dots, P(n))$ be a proof in D.*

- *A rule $r \in R^C$ is #-applicable-at-t' in P iff $\forall p \in A(r)$*
  - *if $p \in$ Lit $\cup$ ModLit $\cup$ ACTS, $+\#@t'p \in P(1..n)$;*
  - *if $p \in \mathscr{R}$, $+\#^C@t'r \in P(1..n)$.*
- *A rule $r :\in R^C$ is #-discarded-at-t' in P iff $\exists p \in A(r)$ such that*
  - *if $p \in$ Lit $\cup$ ModLit $\cup$ ACTS, $-\#@t'p \in P(1..n)$;*
  - *if $p \in \mathscr{R}$, $-\#^C@t'r \in P(1..n)$.*

Applicability conditions for meta-rules (Definition 7) are intuitive. A meta-rule is applicable when its antecedent is provable and the rule is derivable.

Let us now state the positive definite proof procedures for deriving rules using meta-rules. Remember that we assumed to work only with persistent meta-rules.

$+\Delta^C_\pi$: If $P(n+1) = +\Delta^C_\pi @t'r : t_r$, then
(1) $r : t_r \in R$, or
(2) $\exists r' \in R^C_{\pi,s}[r : t_r]$: r' is Δ-applicable-at-t'; or
(3) $\exists t'' \in \mathscr{T}$: $t'' < t$, $+\Delta^C_\pi @t'r : t'' \in P(1..n)$; or
(4) $\exists t'' \in \mathscr{T}$: $t''' < t'$, $+\Delta^C_\pi @t'''r : t \in P(1..n)$.

Remark 4 Strict persistent derivations of rules are based on the use of meta-rules, namely, rules in R^C. These proof conditions obey the same criteria of strict persistent derivations of literals (see Remark 2 for a detailed comment). The main difference is that we do not have that $r \in R^X$ is a condition for definitely deriving r. This happens because the mere inclusion of a rule r in the set of rules of a theory does not exclude that r be attacked by a meta-rule permitting to derive another rule which is incompatible with r.

Defeasible derivations of persistent rules using persistent meta-rules are as follows.

$+\partial^C_\pi$: If $P(n+1) = +\partial^C_\pi @t\ \rho : t_r$, then
(1) $+\Delta^C_\pi @t\ r : t_r \in P(1..n)$, or
(2) $\forall z \in \mathscr{R}$ such that $z : t_z \succeq r : t_r$, $-\Delta @t\ z : t_z \in P(1..n)$, and
(2.1) $\exists s \in R^C[r : t_r]$, and s is ∂-applicable at t and
(2.2) $\forall v \in R^C[x : t_x]$, $x : t_x \succeq r : t_r$ either
v is ∂-discarded at t or $s \succ v$; or
(3) $\exists t' \in \mathscr{T}$, $t' < t_r$, $+\partial^C_\pi @t\ r : t'$ and $\forall t'' \in \mathscr{T}, t' < t'' < t_r$,
$\forall s \in R^C[x : t_x]$, $x : t_x \succeq r : t_r$ either s is ∂-discarded at t or $t_s < t_r$; or
(4) $\exists t'' \in \mathscr{T}$, $t'' < t$, $+\partial^C_\pi @t''\ r : t_r$ and $\forall t''' \in \mathscr{T}, t'' < t''' < t$,
$\forall s \in R^C[x : t_x]$, $x : t_x \succeq r : t_r$ either s is ∂-discarded at t'' or $t_s < t_r$.

Remark 5 Defeasible persistent derivations obey the same criteria of defeasible persistent derivations of literals (see Remark 3 for a detailed comment). The main difference is that conflicts here are based on the notion of incompatibility (see Definition 6).

Proposition 1 *Let D be an acyclic ACL theory (i.e., a theory where the transitive closure of the superiority relation is acyclic); then for every* $\# \in \{\Delta, \partial\}$*:*

- *for not p it is the case that both* $D \vdash +\#p$ *and* $D \vdash -\#p$*;*
- *if* $D \vdash +\partial p$ *and* $D \vdash +\partial{\sim}p$*, then* $D \vdash +\Delta p$ *and* $D \vdash +\Delta{\sim}p$*.*

The above proposition shows the soundness of the logic in the sense that it is both consistent and coherent. i.e., it is not possible to derive a tagged conclusion and its opposite, and that we cannot defeasibly prove both p and its complementary unless the strict part of the theory prove them; this means that inconsistency can be derived only if the theory we started with is inconsistent, and even in this case the logic does not collapse to the trivial extensions (i.e., everything is provable).

While it is possible to show "soundness" of the logic as we have explained above, completeness needs a semantics. It is possible to give an argumentation semantics for basic DL (Governatori *et al.* [2004]) and extensions of the semantics have been proposed for temporalised defeasible logic. In addition Antoniou *et al.* [2006] proposed a semantic characterisation of defeasible logic in terms of Kunen semantics. We believe it is possible to extend this characterisation to cover the logic presented here, but we believe that this kind of work, while valuable to understand the computational properties of the logic and to establish relationships with other formalism, does not contribute to the intuitive understanding of the logic. This would provide a different type of computation to the logic. Instead of combinatorial means one could use set-theoretic notions.

3 Reasoning about FIPA Semantics

In this section we show how the defeasible logic of Section 2, can be applied as a kind of meta-language, to express the semantics of an agent communication language, and to make inferences about communicative acts. For ease of reference, we take FIPA [2002] semantics as a starting point.

In FIPA, communicative acts are defined in terms of rational effects (RE) and feasibility preconditions (FP). The rational effect is the mental state the speaker wants to bring about in the hearer, and the feasibility preconditions encode the appropriate conditions for issuing a communicative act. For instance, here is the FIPA definition of the *inform* communicative act:

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

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 agent a wants to achieve is that hearer b comes to believe p.

Planning operators like this can be used L for enabling agents to generate a dialogue, but they can also be used in the interpretation of the utterances of the interlocutor. In the FIPA framework, this methodology relies on axioms (Property 4 and 5) according to which, when a communicative act is executed, its feasibility preconditions are assumed to be true, and its rational effect is wanted by the speaker[1].

[1] 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*.

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

However, FIPA does not specify some important aspects of communicative acts: persistence, roles, and the rule-based character of dialogue.

Persistence. FIPA does not specify how to deal with the persistence of preconditions and effects of communicative acts. In the case of the above *inform*, hearer b can infer not only that the precondition that a believes p is true at the moment of execution of the communicative act, b can also infer that this precondition held before the communicative act, and will hold afterwards. However, these inferences have a defeasible character, since a can later retract the *inform*.

Concerning rational effects, not only can effects persist towards the future, but they can also be successful or not, which does not only depend on the speaker. FIPA does not allow explicitly to make inferences about the success of a rational effect, even though informally the FIPA specification (p.10) says: "Whether or not the receiver does, indeed, adopt belief in the proposition will be a function of the receiver's trust in the sincerity and reliability of the sender." Nevertheless, in many circumstances the speaker can in fact attribute to the hearer the proposition he has asserted. For example, in collaborative situations when he knows that the hearer considers him reliable, or in persuasion dialogues when the hearer does not challenge the *inform*.

Roles. Strict inferences about effects and preconditions of communicative acts are only possible regarding the public beliefs or goals of the participants (if at all), while inferences about the private beliefs of the participants can be made only by default, under the assumption of cooperativity, trust or sincerity. It is not possible to model a dialogue on the basis of private beliefs and goals only, since in many situations dialogue goes on correctly, even if the participants are not cooperative, sincere or do not trust each other.

We adopt Boella *et al.* [2006a,b]'s *role-based* approach to agent communication, where mental attitudes are publicly attributed to dialogue participants, and can only change according to the rules of the dialogue. Public beliefs and goals represent the expected behavior of an agent, hence they are called a role. Each dialogue is associated with a set of roles to index the public mental attitudes of the participants. Besides the fact that the verifiability problem is solved, in this way, each participant can engage in different dialogues at the same time by playing different roles, and dialogues can obey different kinds of rules associated with each role. For example, in a persuasion dialogue, the proponent of a proposition can have a different 'burden of proof' than the opponent.

In the rules below, the beliefs and goals are therefore attributed to the roles (e.g., $B_{r(a)}$), and not only to the individual agents (e.g., B_a). Attitudes of individual agents can be unknown, or can be different from the attitudes publicly attributed to their roles. Using DL we can connect public and private attitudes in a defeasible manner.

Rule-based character. Defeasible logic is based on rules. Thus, it allows us to model the conditional nature of mental attitudes in a natural way. Consider for example a goal like: "I want to go to the cinema, if it is raining". Moreover, the logic can handle the conditional nature of communicative acts. For example, the semantics of a proposal can be expressed by a rule of the form "I want to execute what I proposed, if the proposal is accepted" (see Section 3.3).

The conditional nature of mental attitudes introduces some complexities. For example, FIPA assumes introspection on beliefs and goals (p.31): $B(a, B(a,p)) \equiv B(a,p)$ and $B(a, G(a,p)) \equiv G(a,p)$ require nesting of mental attitudes. Moreover, the content of a communicative act or modal operator, can be a rule itself. The notation $B[.]$ or $G[.]$ takes care of such nested occurrences of modal operators, and rules. For example, $B_a[p \rightarrow G_a q^t]$ becomes $B_a p \rightarrow G_a q^t$ (an agent is correct about its own goals) and $G_a[p \rightarrow G_b q^t]$ is updated to $G_a p \rightarrow G_a G_b q^t$ (the rule is embedded in a goal of a).

Definition 8 *Let M be a modal literal, X a modal operator.*

$$X[M] \equiv M, \text{ if } M = XM' \text{ or } (X = B_i^t \text{ and } M = G_i^t M'),$$
$$XM \text{ otherwise.}$$
$$X[M_1 \rhd M_2] \equiv X[M_1] \rhd X[M_2], \text{ for} \rhd \in \{\rightarrow, \Rightarrow, \rightsquigarrow\}$$

In the rest of the section we will introduce set of rules formalising the preconditions and the effects of speech act. These rules are intended to be general rules for the relative speech acts. As such the rules are assumed to be valid from the begin of the interaction among the agents. Thus all these rules will be valid from a time t_0 coinciding with the origin of $\mathscr{T}$, which in turn corresponds to the start of the computation for an agent interaction. For the sake of readability, given the above assumption, we will omit the time instant (t_0) associated with all these rules.

3.1 Inform

The constitutive rules $R_{inf} = \{i_1, ..., i_{13}\}$ define the meaning of an *inform* communicative act, for a standard type of cooperative dialogue, like information exchange. In Section 3.5 we consider a non-cooperative version of these rules.

In the following rules, a, b are agents, r is a dialogue instance of cooperative type, $r(a)$ and $r(b)$ the role-playing-agents in the dialogue, *inform* is a communicative act type, s is a rule, and $t < t'$ are time points in $\mathscr{T}$. The rules have priority $i_2 \succ i_1$, $i_9 \succ i_8$, $i_{11} \succ i_{10}$ and $i_{13} \succ i_{12}$.

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. Obviously, there are different ways of handling retraction. The solution here is to withdraw the original communicative act by means of a defeater rule. Because of the way the defeasible logic is set-up, that 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 $\quad utter_{a,b,r}(inform, s, t) \Rightarrow_\tau inform_{r(a),r(b)}(s,t)$

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

Note that the agent of actions *utter* and *retract* is the individual agent a and not its role $r(a)$. Rules i_1, i_2 are used to connect individual agents to their roles. Note furthermore that these rules are non-persistent, since the action of uttering only temporally coincides with execution of an *inform* communicative act.

Rules $i_3 - i_5$ represent the preconditions of making an *inform*. Following FIPA Properties 4 and 5, they are interpreted as strict rules. In contrast, the effect should persist non-monotonically. Only rule i_3 is persistent towards the future, since the *inform* possibly changes the beliefs of the hearer b.

The consequents of precondition rules should persist also to the past, since they are observations and not caused by the speech acts. We do not have the space to cope with this issue.

i_3 $\quad inform_{r(a),r(b)}(s,t) \rightarrow_\pi B^t_{r(a)}[s]$

i_4 $\quad inform_{r(a),r(b)}(s,t) \rightarrow_\tau \neg B^{t-1}_{r(b)}[s]$

i_5 $\quad inform_{r(a),r(b)}(s,t) \rightarrow_\tau \neg B^{t-1}_{r(b)}[\neg s]$

Note that the consequent of the rule has the form $B^t_{r(a)}[s]$ as in Definition 8, and not $B^t_{r(a)}(s)$, since s is a rule too.

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

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

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

i_7 $\quad G^t_{r(a)}[B^t_{r(b)}[s]] \wedge B^t_{r(b)} reliable(a) \rightarrow_\pi B^t_{r(b)}[s]$

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 $i_9 \succ i_{10}$, $i_{11} \succ i_{10}$ and $i_{13} \succ i_{12}$.

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

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

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

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

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

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

3.2 Request

Analogously to *inform*, we define the preconditions and effects of a *request* communicative act, used in deliberation dialogues, by rules $R_{req} = \{r_1, \ldots, r_{13}\}$, with priority $r_2 \succ r_1$, $r_7 \succ r_6$, $r_9 \succ r_8$, $r_{11} \succ r_{10}$ and $r_{13} \succ r_{12}$. For space reasons, we do not explicitly model preconditions of actions in AA, so compared to FIPA, we have to simplify the definitions. Again, the cooperative behavior is the default, which is overruled when the agent refuses the *request*, or when there is evidence that the agent is insincere, or non-cooperative.

r_1 $\quad utter_{a,b,r}(request, s, t) \Rightarrow_\tau request_{r(a),r(b)}(s,t)$

r_2 $\quad retract_{a,b,r}(request_{r(a),r(b)}(s,t),\ t') \rightsquigarrow_\tau \neg request_{r(a),r(b)}(s,t)$

r_3 $\quad request_{r(a),r(b)}(s,t) \rightarrow_\pi G^t_{r(a)}[s]$

r_4 $\quad request_{r(a),r(b)}(s,t) \rightarrow_\tau \neg G^{t-1}_{r(b)}[s]$

r_5 $\quad request_{r(a),r(b)}(s,t) \rightarrow_\pi G^t_{r(a)}[G^t_{r(b)}[s]]$

r_6 $\quad G^t_{r(a)}[G^t_{r(b)}[s]] \Rightarrow_\pi G^t_{r(b)}[s]$

r_7 $\quad request_{r(a),r(b)}(s,t) \wedge refuse_{r(b),r(a)}(s,t') \rightsquigarrow_\pi \neg G^t_{r(b)}[s]$

r_8 $\quad request_{r(a),r(b)}(s,t) \Rightarrow_\pi G^t_a[s]$

r_9 $\quad request_{r(a),r(b)}(s,t) \wedge B^t_{r(b)} \neg sincere(a) \rightsquigarrow_\pi \neg G^t_a[s]$

r_{10} $\quad request_{r(a),r(b)}(s,t) \wedge G^t_{r(b)}[s] \Rightarrow_\pi G^t_b[s]$

r_{11} $\quad request_{r(a),r(b)}(s,t) \wedge G^t_{r(b)}[s] \wedge B^t_{r(a)} \neg cooperative(b) \rightsquigarrow_\pi \neg G^t_b[s]$

r_{12} $\quad request_{r(a),r(b)}(s,t) \wedge G^t_{r(a)}[G^t_{r(b)}[s]] \Rightarrow_\pi G^t_a[G^t_b[s]]$

r_{13} $\quad request_{r(a),r(b)}(s,t) \wedge B^t_{r(b)} insincere(a) \rightsquigarrow_\pi \neg G^t_a[G^t_b[s]]$

3.3 Composed actions: propose

In our framework, it is relatively easy to model non-primitive communicative acts, which are defined in terms of other ones. For example, following FIPA [2002], communicative acts *propose* and *accept* are defined in terms of *inform*, using a set of constitutive

$$\frac{+\Delta utter_{a,b,r}(propose,\alpha,t)}{-\Delta\neg propose_{r(a),r(b)}(\alpha,t)}\ p_1$$
$$\frac{-\Delta\neg propose_{r(a),r(b)}(\alpha,t)}{+\delta propose_{r(a),r(b)}(\alpha,t)}\ p_1$$
$$\frac{+\delta propose_{r(a),r(b)}(\alpha,t)}{+\Delta inform_{r(a),r(b)}(G_{r(b)}(\alpha)\rightarrow^t G_{r(a)}(\alpha),t)}\ p_3$$
$$\frac{}{+\Delta B_{r(a)}(G_{r(b)}(\alpha))\rightarrow^t G_{r(a)}(\alpha)}\ i_3$$
$$\frac{}{+\Delta\neg(G_{r(b)}(\alpha)\rightarrow^{t-1} B_{r(b)}G_{r(a)}(\alpha))}\ i_4$$
$$\frac{}{+\Delta\neg(B_{r(b)}\neg(G_{r(b)}(\alpha)\rightarrow^{t-1} G_{r(a)}(\alpha)))}\ i_5$$
$$\frac{}{+\Delta G_{r(a)}(G_{r(b)}(\alpha))\rightarrow^t G_{r(a)}B_{r(b)}G_{r(a)}(\alpha))}\ i_6$$
$$\frac{B_{r(b)}reliable(a) \quad +\Delta G_{r(a)}(G_{r(b)}(\alpha))\rightarrow^t G_{r(a)}B_{r(b)}G_{r(a)}(\alpha))}{+\Delta(G_{r(b)}(\alpha)\rightarrow^t B_{r(b)}G_{r(a)}(\alpha))}\ i_7$$
$$\frac{}{-\Delta\neg(B_aG_{r(b)}(\alpha)\rightarrow^t B_aG_{r(a)}(\alpha))}\ i_8$$
$$\frac{}{+\delta(B_aG_{r(b)}(\alpha)\rightarrow^t B_aG_{r(a)}(\alpha))}\ i_8$$
$$\frac{}{-\Delta(B_b(G_{r(b)}(\alpha))\rightarrow^t B_bG_{r(a)}(\alpha))}\ i_{10}$$
$$\frac{}{+\delta(B_b(G_{r(b)}(\alpha))\rightarrow^t B_bG_{r(a)}(\alpha))}\ i_{10}$$

Figure 1: A proof for theory T.

rules $R_{prop} = \{p_1, p_2, p_3, a_1\}$. In this way, *propose* and *accept* inherit the feasibility preconditions and effect rules of *inform*. In the following, assume that action $\alpha \in AA$.

This kind of definition makes explicit use of the rule-based character of the logic. In rule p_3 a proposal is defined as the announcement of a conditional goal $G_{r(b)}(\alpha) \rightarrow^t G_{r(a)}(\alpha)$. It means that the speaker will adopt goal α, if the hearer indicates also to have α as a goal, i.e., if the hearer accepts the proposal. In rule a_1, an acceptance is defined as an announcement of such an goal.

p_1 $\quad utter_{a,b,r}(propose,\alpha,t) \Rightarrow_\tau propose_{r(a),r(b)}(\alpha,t)$

p_2 $\quad retract_{a,b,r}(propose_{r(a),r(b)}(\alpha,t),\ t') \rightsquigarrow_\tau \neg propose_{r(a),r(b)}(\alpha,t)$

p_3 $\quad propose_{r(a),r(b)}(\alpha,t) \rightarrow_\tau inform_{r(a),r(b)}(G_{r(b)}(\alpha)\rightarrow^t G_{r(a)}(\alpha),\ t)$

a_1 $\quad accept_{r(a),r(b)}(\alpha,t) \rightarrow_\tau inform_{r(a),r(b)}(G_{r(a)}\alpha,t)$

By means of illustration, we construct a proof for the defeasible ACL theory $T = \langle \mathscr{T}, \{B^t_{r(b)}reliable(a), utter_{a,b,r}(propose,\alpha,t)\}, \{a,b\}, \{r_1,r_2\}, \langle r(a)=r_1, r(b)=r_2\rangle, R_{prop}\cup R_{inf}, R_{prop}\cup R_{inf}, \prec\rangle$. The proof is shown in Figure 1.

In the example, a *propose* has been uttered ($+\Delta$) and no evidence can be non-defeasibly gathered ($-\Delta$) against the proposal: rule p_2 is not applicable. So we defeasibly infer ($+\delta$) that *propose* holds (rule p_1). Rule p_3 defeasibly derives that a corresponding *inform* has been generated. From now on, rules concerning *inform* can be applied. Rules $i_3 - i_5$ strictly derive ($+\Delta$) the preconditions of the *inform*. Due to Definition 8, the clause $B_{r(a)}[G_{r(b)}(\alpha)\rightarrow^t G_{r(a)}(\alpha)]$ is first reduced to $B_{r(a)}[G_{r(b)}(\alpha)] \rightarrow^t B_{r(a)}[G_{r(a)}(\alpha)]$ and then to $B_{r(a)}(G_{r(b)}(\alpha)) \rightarrow^t G_{r(a)}(\alpha)$. Finally, the example derives (i_{10}) that the hearer also privately believes the conditional goal that the speaker wants to do α, if the hearer wants that too.

3.4 Abstract communicative acts: inform-if

Like in FIPA, we can also handle abstract communicative acts, like *inform-if*(a,b,p) which is composed of the nondeterministic choice of *inform*(a,b,p) and *inform*$(a,b,\neg p)$. Note that in FIPA *inform-if* is an abstract action which cannot directly be executed:

ii_1 $\quad inform_{r(a),r(b)}(s,t) \rightarrow_\tau inform\text{-}if_{r(a),r(b)}(s,t)$

ii_2 $\quad inform_{r(a),r(b)}(\neg s,t) \rightarrow_\tau inform\text{-}if_{r(a),r(b)}(s,t)$

Thus, we can define *query-if* as a *request* to *inform-if*:

qi_1 $\quad queryif_{r(a),r(b)}(s,t) \rightarrow_\tau request_{r(a),r(b)}(inform\text{-}if_{r(b),r(a)}(s,t'),t)$

To satisfy the *request*, the receiver has to execute either an $inform_{r(b),r(a)}(s,t')$ or an $inform_{r(b),r(a)}(\neg s,t')$.

3.5 Persuasion dialogues

Agent communication languages are usually studied in relation to specific *dialogue types*, such as cooperative information exchange, negotiation or persuasion, see, e.g., the taxonomy of Walton and Krabbe [1995]. Given the non-monotonic character of DL, we can extend FIPA with persuasion, by defining acts like *challenge* and *concede*.

Interestingly, a single communicative act like *inform* may have different semantics, in different types of dialogue. This is due to different background assumptions, for example regarding sincerity, cooperativity, or trust. Thus, in uncooperative dialogue types like persuasion or negotiation, it is possible to reverse the general principle of rules $i_6 - i_{13}$, that cooperative behavior is expected by default, but can be overruled by evidence to the contrary.

Alternatively, we can follow the principle that "silence means consent". In a persuasion dialogue, the hearer is assumed to believe what the speaker said (adapted rule i'_7), unless he explicitly challenges the proposition (additional rule i''_7), thus defeating the conclusion that he believes the content of the *inform*.

i'_7 $\quad G^t_{r(a)}[B^t_{r(b)}[s]] \Rightarrow_\pi B^t_{r(b)}[s]$

i''_7 $\quad inform_{r(a),r(b)}(s,t) \wedge challenge_{r(b),r(a)}(s,t') \rightsquigarrow_\pi \neg B^t_{r(b)}[s]$

In addition to challenges, we can add explicit concessions (Walton and Krabbe [1995]). If an agent concedes to p, it does not necessarily mean that he now believes p, but it means that he does no longer believe the opposite. For example, it blocks the agent from performing an *inform* that $\neg s$ later in the interaction.

c_1 $\quad concede_{r(a),r(b)}(s,t) \rightarrow_\pi \neg B^t_{r(a)}[\neg s]$

So for persuasion, the rules R_{inf} of Section 3.1 are altered as follows: $R_{persuasion} = (R_{inf}\setminus\{i_7\}) \cup \{i'_7, i''_7, c_1\}$, where $i''_7 \succ i'_7$.

4 Related work

Recently, some other papers went in the same direction of redefining FIPA semantics: e.g., Nickles *et al.* [2006]; Verdicchio and Colombetti [2006]; Gaudou *et al.* [2006a]; Boella *et al.* [2006a,b]. Like us, most of them distinguish between public and private mental attitudes. There are various differences.

Like in Boella *et al.* [2006a,b], the distinguishing feature of our approach is that the public mental attitudes attributed to agents during the dialogue are associated with roles. However, we use roles to redefine the FIPA semantics in a non-monotonic framework based on DL which allows us to extend FIPA to persuasion dialogue. We distinguish interactive roles, such as speaker, (over)hearer and addressee. Clearly, different constitutive rules apply to speaker and hearer. Further, we could add rules so that the effects of implicit acknowledgement differ between the addressee of a message, and a mere overhearer (Gaudou *et al.* [2006b]). Because social roles are associated with dialogue types, with specific set of dialogue rules, roles allow us to reason about assumptions in different kinds of dialogues. E.g., sincerity could be assumed in cooperative dialogues, such as information exchange, but not in non-cooperative dialogues, such as persuasion or negotiation. Ostensible beliefs and the grounding operator distinguish only interactive roles or different groups of agents.

The importance of roles is recognized in multiagent systems and their function ranges from attributing responsibilities to assigning powers to agents in organizations.Other solutions, instead, need to add to dialogue new theoretical concepts which are not always completely clear or diverge from existing work. In particular, Gaudou *et al.* [2006a] use an explicit grounding operator, which only partially overlaps with the tradition of grounding in theories of natural language dialogue. Opinions Nickles *et al.* [2006] are introduced specifically for modelling dialogue, but with no relation with persuasion and argumentation. Finally, commitments in Verdicchio and Colombetti [2006] overlap with obligations.

Moreover, the approaches relate to the well known FIPA semantics in different degrees: Gaudou *et al.* [2006b] and Nickles *et al.* [2006] try to stay close to the original semantics, as we do, while Verdicchio and Colombetti [2006] substitute it entirely with a new semantics, which, among other things, does not consider preconditions of actions. Gaudou *et al.* [2006a] and Nickles *et al.* [2006] use modal logic, and Verdicchio and Colombetti [2006] use CTL. They thus use more common frameworks, but they do not consider the computational properties of their proposals. DL, instead, may provide us with a proof theory which is linear (cf. Governatori *et al.* [2006a]). Moreover, most of the other approaches do not consider the persistence of preconditions and effects, an essential point when dealing with actions and their effects. Time introduces most of the complexities in our formal system, but time is crucial for agent communication, because speech acts are uttered one after the other, but their effects on mental attitudes are persistent.

5 Summary

We introduce a defeasible logic for role-based agent communication and have shown how to formalize the essential features of FIPA semantics on public mental states, including composed and abstract actions. Moreover, we have shown how it can be extended to reason about different assumptions about agents in cooperative dialogue, like information exchange, and non-cooperative dialogue, like persuasion and negotiation. We use the logic to study the non-monotonic and rule-based character of role based agent communication.

Non-monotonic reasoning occurs in reasoning about the persistence in time of the effects of speech acts. Moreover, 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 communicating agents which hold only by default and can always be revised by new information. Moreover, non-monotonic reasoning can be used for challenges and concessions. E.g., an inform is accepted – i.e., its content becomes part of the public beliefs of the addressee – unless it is challenged. Finally, non-monotonic reasoning can be used to deal with persuasion and negotiation, where old arguments can be retracted or overridden by stronger ones.

References

G. Antoniou, D. Billington, G. Governatori, M.J. Maher, and A. Rock. A family of defeasible reasoning logics and its implementation. In *Procs. of ECAI'00*, pages 459–463, 2000.

G. Antoniou, D. Billington, G. Governatori, and M.J. Maher. Embedding defeasible logic into logic programming. *Theory and Practice of Logic Programming*, 2006.

G. Boella, R. Damiano, J. Hulstijn, and L. van der Torre. Public mental attitudes for ACL semantics. In *Agent Communication (AC'06)*, volume LNAI 3859, pages 30–44. Springer Verlag, 2006.

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. of AAMAS'06*, pages 688–690. 2006.

P. Bresciani, P. Giorgini, F. Giunchiglia, J. Mylopoulos, and A. Perini. TROPOS: An agent-oriented software development methodology. *Autonomous Agents and Multi-Agent Systems*, 8(3):203–236, 2004.

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

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

B. Gaudou, A. Herzig, and D. Longin. A logical framework for grounding-based dialogue analysis. In *Procs. of LCMAS'05*, pages 117–137. 2006. ENTCS 157(4).

B. Gaudou, A. Herzig, D. Longin, and M. Nickles. A new semantics for the fipa agent communication language based on social attitudes. In *Procs. of ECAI'06*, pages 245–249. 2006.

G. Governatori, M.J. Maher, D. Billington, and G. Antoniou. Argumentation semantics for defeasible logics. *Journal of Logic and Computation*, 14(5):675–702, 2004.

G. Governatori, A. Rotolo, and G. Sartor. Temporalised normative positions in defeasible logic. In *Procs. of ICAIL'05*, pages 25–34. ACM Press, 2005.

G. Governatori, V. Nair, and A. Rotolo. The cost of social agents. In *Procs. of AAMAS06*, pages 513 – 520. ACM, 2006.

G. Governatori, V. Padmanabhan, and Rotolo. A. Rule-based agents in temporalised defeasible logic. In *Procs. of PRICAI'06*, volume LNAI 4099, pages 31–40. Springer, 2006.

P. McBurney and S. Parsons. Locutions for argumentation in agent interaction protocols. In *Agent Communication (AC'04)*, volume LNAI 3396, pages 209–225. Springer Verlag, 2004.

M. Nickles, F. Fischer, and G. Weiss. Communication attitudes: A formal approach to ostensible intentions, and individual and group opinions. In *Procs. of LCMAS'05*, pages 95–115. 2006. ENTCS 157(4).

J. Pitt and A. Mamdani. Some remarks on the semantics of FIPA's agent communication language. *Autonomous Agents and Multi-Agent Systems*, 2(4):333–356, 1999.

R. Riveret, G. Governatori, and A. Rotolo. Argumentation semantics for temporal defeasible logic. In *Third European Starting AI Researcher Symposium (STAIRS 2006)*, pages 267–268, Amsterdam, 2006. IOS Press.

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

I. Song and G. Governatori. Nested rules in defeasible logic. In *Procs. RuleML 2006*, volume LNAI 3791, pages 204–208. Springer, 2005.

M. Verdicchio and M. Colombetti. From message exchanges to communicative acts to commitments. In *Procs. of LCMAS'05*, pages 75–94. 2006. ENTCS 157(4).

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

M.J. Wooldridge, N.R. Jennings, and D. Kinny. The Gaia methodology for agent-oriented analysis and design. *Autonomous Agents and Multi-Agent Systems*, 3(3):285–312, 2000.

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