



HAL
open science

An analysis of communication and cooperation in a logic of belief, intention, and action

Olivier Gasquet, Andreas Herzig, Dominique Longin

► To cite this version:

Olivier Gasquet, Andreas Herzig, Dominique Longin. An analysis of communication and cooperation in a logic of belief, intention, and action. [Research Report] IRIT/2001-07-R, IRIT : Institut de Recherche en Informatique de Toulouse, France. 2001. hal-03523404

HAL Id: hal-03523404

<https://hal.science/hal-03523404>

Submitted on 12 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An analysis of communication and cooperation in a logic of belief, intention, and action

Olivier Gasquet Andreas Herzig
Dominique Longin

Institut de Recherche en Informatique de Toulouse (IRIT)
Équipe Logic, Interaction, Language and Computation (LILAC)
CNRS – Université Paul Sabatier
118 Route de Narbonne, F-31062 Toulouse cedex 4
{gasquet,herzig,longin}@irit.fr
<http://www.irit.fr/ACTIVITES/LILaC/>

October 16, 2001

RAPPORT TECHNIQUE IRIT/2001-07-R

Content Areas: reasoning about actions, cognitive robotics.

Abstract

In this paper we propose a logic for the dynamics of beliefs and intentions. We suppose that agents are cooperative, and interact via both communicative and physical actions. We build on an existing Dynamic Logic based solution to the frame problem, enabling in particular tableau theorem proving methods. This provides an alternative to recent approaches of Shapiro *et col.*

Contents

Résumé	3
1 Introduction	7
2 Mental states	7
2.1 Belief	7
2.2 Intention	8
2.3 Relations between mental attitudes	8
3 Actions	9
3.1 Dynamic logic	9
3.2 Action laws	10
3.3 Communicative actions	10
4 Rationality principles	11
4.1 Nonperceptive actions	12
4.2 Preservation of mental attitudes	14
5 Cooperation principles	15
5.1 Belief adoption	15
5.2 Intention adoption	16
5.3 Intention generation	16
6 Semantics	17
7 Discussion and related work	18
7.1 Integrating belief and revision	18
7.2 Integrating goals	19
8 Conclusion	19

1 Introduction

Starting from the ‘common belief’ that the frame problem has been solved, existing solutions have been extended in the last years to handle the dynamics of knowledge [Scherl and Levesque, 1993, Thielscher, 2000].

Modifying these solutions in order to handle belief is not easy. In this paper we show that the recent proposal of [Shapiro et al., 1999] encounters some difficulties when applied to communication acts. We propose a logic for belief dynamics based on a Dynamic Logic solution to the frame problem [Castilho et al., 1999]. We integrate a primitive notion of intention having a non-normal modal logic (not closed under implication). Within this logic we give an analysis of cooperation in communication, and compare it to recent Situation Calculus based approaches of Shapiro *et col.* [Shapiro et al., 1997, Shapiro et al., 1998, Shapiro and Lespérance, pear].

The paper is organized as follows: we introduce modal operators of belief and intention (Sect. 2), and of action (Sect. 3). Then we give some principles of rationality (Sect. 4) and cooperation (Sect. 5). We sketch a possible worlds semantics for the resulting logic (Sect. 6). Finally we discuss related work (Sect. 7).

2 Mental states

Based on the philosophical theories of Searle [Searle, 1983] and Bratman [Bratman, 1987], our logic follows the tradition of Cohen & Levesque [Cohen and Levesque, 1990a, Cohen and Levesque, 1990b] and Sadek [Sadek, 1992, Sadek, 2000]. As the latter, we aim at generalizing speech act theory into a theory of communication, and we suppose that the properties of the latter are derivable from (more general) principles of rational interaction.

2.1 Belief

Let $\mathcal{AGT} = \{i, j, \dots\}$ be the set of agents. We associate a modal operator of belief Bel_i to every $i \in \mathcal{AGT}$. The formula $Bel_i A$ is read “agent i believes that A ”. $BelIf_i A$ is an abbreviation of $Bel_i A \vee Bel_i \neg A$ and reads “agent i knows whether A is true or not”.¹ We adopt the modal logic KD45 as the logic of belief. This implies that we suppose that

¹We use the term “knows” here because “ i believes if A ” sounds odd.

agents cannot entertain inconsistent beliefs, that they are aware of their beliefs and of their disbeliefs.

2.2 Intention

Intention is a fundamental mental attitude, because it is at the origin of every voluntary action. We associate a modal operator of intention Int_i to every $i \in \mathcal{AGT}$, and read the formula $Int_i A$ “agent i intends that A ”.

Intention is neither closed under logical truth, nor under logical consequence, conjunction, and material implication. We only postulate:

$$\frac{A \leftrightarrow B}{Int_i A \leftrightarrow Int_i B} \quad (\text{RE}_{Int_i})$$

This is in accordance with [Bratman, 1987, Cohen and Levesque, 1990a, Sadek, 1992], but contrarily to these approaches, intention is primitive here, as in [Rao and Georgeff, 1991] and as in [Konolige and Pollack, 1993] where only closure under logical consequence had been given up. We thus generalize the latter semantics.

We have chosen this solution for three reasons. First, building intention on top of other primitive notions such as goals or desires leads to various sophisticated notions of intention, with subtle differences between them. We have kept here only those properties of intention that are common to all of them, viz. extensionality. Second, as these definitions are rather complex, it is difficult to find complete automated theorem proving methods for them, while our analysis enables more or less standard completeness techniques and proof methods. Third and most importantly, we think that our simplified notion of intention is sufficient at least in many applications.

2.3 Relations between mental attitudes

We think that rather than the interaction between intentions and goals or desires, it is the interaction between intentions and beliefs which is crucial. Most importantly, an agent must abandon his intention to realize A as soon as he believes that A is true. This is expressed by the axiom schema:

$$Int_i A \rightarrow Bel_i \neg A \quad (\text{Realism})$$

This axiom, combined with the consistency of belief, entails consistency of intentions, i.e. $Int_i A \rightarrow \neg Int_i \neg A$.

Several other properties describe the relation between belief and intention. The following are important for us:

$$Bel_i Int_i A \leftrightarrow Int_i A \quad (\text{BelIntend}_1)$$

$$Bel_i \neg Int_i A \leftrightarrow \neg Int_i A \quad (\text{BelIntend}_2)$$

(cf. [Cohen and Levesque, 1990a, Sadek, 1992] for more details).

3 Actions

3.1 Dynamic logic

We use a version of dynamic logic [Harel, 1984]. Actions being done by agents, we suppose that atomic actions are of the form $\langle i, e \rangle$, where $i \in \mathcal{AGT}$ is the author of the event e .

We moreover suppose that for every atomic action $\langle i, e \rangle$ there is also an atomic action $\langle nat, \text{Enable}_- \langle i, e \rangle \rangle$. We suppose that these actions are executed by nature nat (hence they can be called events), and that $\langle nat, \text{Enable}_- \langle i, e \rangle \rangle$ can occur in every state, its effect being to make the executability preconditions of $\langle i, e \rangle$ true.² At first glance this seems to be a somewhat artificially constructed event, but it makes sense when applied to belief states: if an agent learns that an action he believed inexecutable has nevertheless happened, then he enables executability of that action in his possible states (see Sect. 4.1).

We construct complex actions (noted α, α', \dots) using the standard operators “ λ ” (the empty action) “;” (sequential execution) “ \cup ” (non-deterministic choice). We hence neither have the iteration operator $*$, nor the test operator $?$ of dynamic logic.

To each action α there are associated modal operators $After_\alpha$ and $Before_\alpha$. An example of a formula is $Bel_i \neg After_{\langle j, e \rangle} \perp \wedge After_{\langle j, e \rangle} \perp$, expressing that i believes that j can do e , while j is in fact not able to do it. Another example is $\neg Bel_i p \wedge \neg Bel_i \neg p \wedge After_{\langle i, e \rangle} Bel_i p$, expressing that i ignores if p is true or not, and starts to believe p after having executed e . $\neg After_{\langle nat, \text{Enable}_- \langle i, e \rangle \rangle} \perp$ expresses the action law that nature can enable every atomic action.

The operators $Feasible_\alpha$ and $Done_\alpha$ are introduced by stipulating that

²In consequence we must suppose that for every atomic action there is at least one state where it is executable. For the same reason, **Enable** cannot be applied to complex actions. Indeed, consider an atomic action $\langle i, e \rangle$ always having effect p , and with precondition of executability $\neg p$. Then $\langle i, e \rangle; \langle i, e \rangle$ cannot be enforced.

$Feasible_\alpha A$ abbreviates $\neg After_\alpha \neg A$, and $Done_\alpha A$ abbreviates $\neg Before_\alpha \neg A$.³ Complex formulas are denoted by A, B, C, \dots

We respectively note \mathcal{ACT} and \mathcal{FORM} the set of actions and formulas thus defined. We say that a formula of \mathcal{FORM} is *objective* if it contains no modal operator.

We adopt the standard axiomatics for the fragment of dynamic logic corresponding to our language.

3.2 Action laws

For an atomic action $\langle i, e \rangle$, we suppose that action laws are of the following form:

$$\begin{aligned} A &\rightarrow After_{\langle i, e \rangle} C \\ B &\rightarrow Feasible_{\langle i, e \rangle} \top \end{aligned}$$

We note $E(\langle i, e \rangle)$ the set of effect laws for $\langle i, e \rangle$, and $X(\langle i, e \rangle)$ the set of executability laws for $\langle i, e \rangle$.

What are the action laws for the event $\langle nat, \text{Enable}_- \langle i, e \rangle \rangle$? Given that $\langle nat, \text{Enable}_- \langle i, e \rangle \rangle$ makes $\langle i, e \rangle$ executable, the set of executability laws for $\langle i, e \rangle$ determines the unique effect law for $\langle nat, \text{Enable}_- \langle i, e \rangle \rangle$:

$$After_{\langle nat, \text{Enable}_- \langle i, e \rangle \rangle} \left(\bigvee_{B_i \rightarrow Feasible_{\langle i, e \rangle} \top \in X(\langle i, e \rangle)} B_i \right)$$

And $\langle nat, \text{Enable}_- \langle i, e \rangle \rangle$ is always executable:

$$Feasible_{\langle nat, \text{Enable}_- \langle i, e \rangle \rangle} \top$$

Just as similar approaches [Scherl and Levesque, 1993, Shapiro et al., 1999, Thielscher, 2000], we suppose that the laws governing the actions are known by all agents. We can therefore consider the elements of $E(\langle i, e \rangle)$ and $X(\langle i, e \rangle)$ to be global axioms in the standard sense of modal logic.

3.3 Communicative actions

Following Austin, Searle and Vanderveken, we suppose that our language contains particular actions of communication. In order to simplify the exposition we suppose that assertive acts take the form $\langle i, \text{informs } A \rangle$,

³ $After_\alpha$ and $Before_\alpha$ correspond to the dynamic logic operators $[\alpha]$ and $[\alpha^{-1}]$, and $Feasible_\alpha$ and $Done_\alpha$ correspond to $\langle \alpha \rangle$ and $\langle \alpha^{-1} \rangle$.

where A is a formula. Hence we do not mention the addressee. This will be partly justified by the fact that actions are public. (It is even more justified in the case where there are only two agents.)

We suppose agents are sincere, i.e. that an assertive act $\langle i, \text{informs } A \rangle$ can be performed only if its author believes its propositional contents A . Hence we have the axiom:

$$\text{Before}_{\langle i, \text{informs } A \rangle} \text{Bel}_i A \quad (\text{Sin}_{\text{CInform}})$$

We also suppose that the sincerity condition of an act is not modified by its execution, expressed by the axiom

$$\text{After}_{\langle i, \text{informs } A \rangle} \text{Bel}_i A \quad (\text{PresSin}_{\text{CInform}})$$

We moreover suppose that communicative actions are deterministic:

$$\neg \text{After}_{\langle i, \text{informs } A \rangle} \neg B \rightarrow \text{After}_{\langle i, \text{informs } A \rangle} B \quad (\text{Det}_{\text{Inform}})$$

Finally, assertions do not change the physical world. This is expressed by the influence-based axiom

$$C \rightarrow \text{After}_{\langle i, \text{informs } A \rangle} C$$

if C is an objective formula ($\text{Pres}_{\text{Inform}}$)

Cooperativity justifies the hypothesis that the other types of communicative acts can be deduced from assertive acts. More precisely, we consider that an assertive act is in fact a directive if its propositional contents is a speaker intention on a *Bellf* or a *Done* operator. For example, we express the closed question $\langle i, \text{queryYN } A \rangle$ (i asks if A is true) by the act $\langle i, \text{informs } \text{Int}_i \text{Bellf}_i A \rangle$: the hearer j being cooperative, it will treat this act as if i had explicitly asked i to perform either $\langle i, \text{informs } A \rangle$ or $\langle i, \text{informs } \neg A \rangle$. We express the directives $\langle i, \text{request } \langle j, e \rangle \rangle$ (i asks j to do e) by the act $\langle i, \text{informs } \text{Int}_i \text{Done}_{\langle j, e \rangle} \top \rangle$. Our cooperation principles of Sect. 5 will guarantee this.

4 Rationality principles

On the axiomatic level, are there any principles beyond the simple juxtaposition of the axioms for belief and those for actions? It seems difficult to answer that question without analyzing actions in a finer manner, and making some restrictions.

But first of all, and with many similar approaches such as [Scherl and Levesque, 1993, Shapiro et al., 1999, Thielscher, 2000] we

suppose that the occurrence of an action is public, and thus perceived by every agent:

$$Bel_i Done_\alpha \top \leftrightarrow Done_\alpha \top \quad (\text{Public})$$

(Public) stipulates that the perception of action occurrences is correct and complete.

4.1 Nonperceptive actions

As done more or less explicitly in similar approaches [Scherl and Levesque, 1993, Shapiro et al., 1999], we consider that from the point of view of agent i , every atomic action α can be decomposed into the sequence $(\alpha^{i\neq}; \alpha^{i\leftrightarrow})$, where $\alpha^{i\neq}$ is nonperceptive and $\alpha^{i\leftrightarrow}$ is perceptive:

- as i cannot observe the effects of $\alpha^{i\neq}$, he predicts them in an a priori way, according to his mental states and the action laws (' i mentally executes α ');
- in a second step, i executes $\alpha^{i\leftrightarrow}$ and thus perceives which of the possible effects of α have obtained.

(Note that $\alpha^{i\neq}$ and $\alpha^{i\leftrightarrow}$ might be the empty action.) For example, the action of tossing a coin can be decomposed in $\langle i, \text{Toss} \rangle^{i\neq}; \langle i, \text{Toss} \rangle^{i\leftrightarrow}$. $\langle i, \text{Toss} \rangle^{i\neq}$ corresponds to i 's action of tossing the coin without observing the result ('eyes shut'). As tossing a coin is a nondeterministic action, the effect of $\langle i, \text{Toss} \rangle^{i\neq}$ is that i considers both heads and tails as possible outcomes, i.e. we have $\neg \text{After}_{\langle i, \text{Toss} \rangle^{i\neq}} Bel_i heads$. $\langle i, \text{Toss} \rangle^{i\leftrightarrow}$ is i 's action of opening his eyes and checking whether the coin fell heads or tails. Hence we should have $\text{After}_{\langle i, \text{Toss} \rangle^{i\leftrightarrow}} Bel_i heads$. As actions are public, the effect of tossing will be the same on the other agents' mental states.

Now we can link beliefs and actions in a stronger way. The following axioms require that α is atomic and nonperceptive.

$$\text{Feasible}_\alpha Bel_i A \rightarrow \text{After}_\alpha Bel_i A \quad (\text{NonPerc}_1)$$

$$\begin{aligned} (\neg \text{After}_\alpha \perp \wedge \neg Bel_i \text{After}_\alpha \perp) \rightarrow \\ (Bel_i \text{After}_\alpha A \leftrightarrow \text{After}_\alpha Bel_i A) \end{aligned} \quad (\text{NonPerc}_2)$$

(NonPerc₁) expresses what may be called *epistemic determinism* of non-perceptive actions. (NonPerc₂) is the regression axiom for knowledge of

[Scherl and Levesque, 1993]. It expresses e.g. for $\alpha = \langle i, \text{Toss} \rangle^{j \neq i}$ and $A = \neg \text{Bel}_i \text{heads}$ that i 's uncertainty about the result of tossing (coming from its nondeterminism) is preserved through the mental execution of its nonperceptive part.⁴ Note that the plausibility of (NonPerc₂) heavily relies on our hypothesis that actions are public.

Our axioms do not account for situations where i believes that α is inexecutable, and learns that it has nevertheless been executed. Such situations are indeed problematic, because i cannot just mentally execute α and collect the resulting states, but must first change his beliefs about the preconditions of α .

We propose to implement this process by the mental execution of $\langle \text{nat}, \text{Enable}_- \alpha \rangle$, where α is an atomic nonperceptive action:

$$\begin{aligned} & (\text{Bel}_i \text{After}_\alpha \perp \wedge \neg \text{After}_\alpha \perp) \rightarrow \\ & \quad (\text{After}_\alpha \text{Bel}_i A \leftrightarrow \\ & \quad \quad \text{Bel}_i \text{After}_{\langle \text{nat}, \text{Enable}_- \alpha \rangle} \text{After}_\alpha A) \end{aligned} \quad (\text{NonPerc}_3)$$

Note that this is a revision (as opposed to updates), in the sense that $\langle \text{nat}, \text{Enable}_- \alpha \rangle$ leaves the real world unchanged, while changing the mental state of i in order to take into account the new piece of information $\text{Feasible}_\alpha \top$.

Assertions as nonperceptive actions. How do communicative actions fit into our picture?

The assertion $\langle i, \text{informs } A \rangle$ is nonperceptive for i (apart from learning its execution), because i does not learn anything his beliefs would not have allowed him to infer beforehand. This corresponds to the fact that instantiating α by $\langle i, \text{informs } A \rangle$ in our axioms of nonperception we obtain an intuitively acceptable formula.

It is more surprising that instantiating α by $\langle j, \text{informs } A \rangle$ our axioms for nonperception are also acceptable. This is related to the fact that $\langle i, \text{informs } A \rangle$ is deterministic.

In other words, all communicative actions are not perceptive, and their effects on beliefs is characterized by axioms (NonPerc₁), (NonPerc₂), and (NonPerc₃).

⁴The only case where the \rightarrow direction of (NonPerc₂) cannot be accepted is when α erases all or part of the memory of i . (This is the case e.g. if i is a robot, and α the action of taking off the batteries of i). The \leftarrow direction of (NonPerc₂) is unintuitive only if i knows that α adds unjustified information to its memory. This is the case e.g. when i is hypnotized or takes drugs. We exclude here such extreme cases.

4.2 Preservation of mental attitudes

Given the regression axioms we can reuse non-epistemic solutions to the frame problem. Just as Scherl and Levesque have applied Reiter’s solution [Scherl and Levesque, 1993] we use the solution of [Castilho et al., 1999].

Which beliefs of the hearer can be preserved after the performance of an action? Our key concept here is that of the *influence of an action on beliefs*. If there exists a relation of influence between the action and a belief, this belief cannot be preserved.⁵ Thus, $\langle i, e \rangle \overset{p}{\rightsquigarrow} q$ means “the action α influences the truth value of q ”.⁶ We suppose that $\overset{p}{\rightsquigarrow}$ is in metalanguage. We note $i \not\overset{p}{\rightsquigarrow} A$ when for every atom p occurring in A , $i \overset{p}{\rightsquigarrow} p$ does not hold.

The preservation of mental attitudes not influenced by an action is formalized by the influence-based axiom

$$\begin{array}{l} A \rightarrow \text{After}_\alpha A \\ \text{if } \alpha \not\overset{p}{\rightsquigarrow} A \text{ and } A \text{ is objective} \end{array} \quad (\text{Preserv}_1)$$

This expresses that if α has ‘nothing to do’ with A then A is preserved. The restriction on the form of A is necessary to e.g. avoid $\text{Feasible}_\beta \top \rightarrow \text{After}_\alpha \text{Feasible}_\beta \top$ (which is not necessarily the case).

An agent i ’s competence at a formula expresses credibility of the i ’s believes. Competence can be viewed as a metalinguistic relation between an agent and a formula. We suppose given a relation $i \overset{c}{\rightsquigarrow} p$ expressing that i is competent at the atomic formula p . Then we note $i \overset{c}{\rightsquigarrow} A$ when $i \overset{c}{\rightsquigarrow} p$ holds for every p in A . We note $i \not\overset{c}{\rightsquigarrow} A$ when for every p in A , $i \overset{c}{\rightsquigarrow} p$ does not hold, and we say “ i is incompetent at A ”. Note that $\not\overset{c}{\rightsquigarrow}$ is strictly included in the set complement of $\overset{c}{\rightsquigarrow}$. Note that this allows an agent to preserve his intentions even if he believes that they cannot be satisfied (in which case, according to Cohen & Levesque, he should drop it).

The following axiom expresses that incompetent utterances of i can

⁵The concept of influence (or dependence) of an action is close to the notions that have recently been studied in the field of reasoning about actions in order to solve the frame problem, e.g. Sandewall’s [Sandewall, 1994] occlusion, Thielscher’s [Thielscher, 1995] influence relation, or Giunchiglia *et al.*’s [Giunchiglia et al., 1997] possibly changes operators.

⁶We could also adopt the solution of Herzig & Longin [Herzig and Longin, 2000] where the dependence relations are defined by the primitive notion of topic.

at most change the beliefs of the other agents about those of i .

$$\begin{aligned} B &\rightarrow \text{After}_{\langle i, \text{informs } A \rangle} B \\ \text{if } i \not\rightsquigarrow A \text{ and no } Bel_i \text{ or } Int_i \text{ is in } A \end{aligned} \quad (\text{Preserv}_2)$$

In particular, an agent j preserves his belief $\neg A$ if i asserts that A .

The next axiom says that if an agent intends that A then he abandons that intention only if he learns that A :

$$\begin{aligned} Int_i A &\rightarrow \\ \text{After}_{\langle j, e \rangle} (Bel_i \neg A &\rightarrow Int_i A) \end{aligned} \quad (\text{Preserv}_3)$$

Our agents are not fanatic: the contraposition $\neg Bel_i \neg A \rightarrow \neg Int_i A$ of axiom (Realism) makes that $Int_i A$ is abandoned as soon as A gets possible for i .

These axioms are enough to guarantee the preservation of non-influenced intentions:

$$\begin{aligned} Int_i A &\rightarrow \text{After}_{\langle j, e \rangle} Int_i A \\ \text{if } \langle j, e \rangle \not\rightsquigarrow A \text{ and } A \text{ is objective} \end{aligned} \quad (1)$$

This follows from (Realism), (Preserv₁), and (Preserv₃).

From (Realism), (Adopt_{Bel2}), and (Preserv₃) one can prove using the consistency of beliefs

$$Int_i \neg A \rightarrow \text{After}_{\langle j, \text{informs } A \rangle} Int_i \neg A \quad (2)$$

From (Realism), (Preserv₂), and (Preserv₃) one can prove

$$\begin{aligned} Int_i A &\rightarrow \text{After}_{\langle j, \text{informs } A \rangle} Int_i A \\ \text{if } j \not\rightsquigarrow A \text{ and no } Bel_j \text{ or } Int_j \text{ is in } A \end{aligned} \quad (3)$$

5 Cooperation principles

The cooperative behavior of an agent i with respect to another agent j can be analyzed in terms of belief adoption (Sect. 5.1), intention adoption (Sect. 5.2), and intention generation (Sect. 5.3).

5.1 Belief adoption

In our approach, actions are public and speakers are sincere. Therefore we have $\text{After}_{\langle j, \text{informs } A \rangle} Bel_i Bel_j A$ (see Sect. 4.2). Under which conditions

does i adopt A , i.e. when do we have $After_{\langle j, \text{informs } A \rangle} Bel_i A$? We consider that belief amounts to knowledge in the case of competence:

$$\begin{array}{l} Bel_i A \rightarrow A \\ \text{if } i \overset{c}{\rightsquigarrow} A \text{ and } A \text{ is objective} \end{array} \quad (\text{Adopt}_{\text{Bel1}})$$

Hence j adopts i 's beliefs if i is competent at these beliefs. Note that a similar notion of competence is already in the Cohen & Levesque framework.

Another cooperative principle is the following: if what i says is consistent with j 's beliefs then j starts to believe it:

$$\neg Bel_j \neg A \rightarrow After_{\langle i, \text{informs } A \rangle} Bel_j A \quad (\text{Adopt}_{\text{Bel2}})$$

Note that i might be incompetent at A . The same principle is used in [Shapiro et al., 1997].

5.2 Intention adoption

If an agent i believes that j intends A and A is consistent with i 's intentions, then according to his beliefs about A , i either adopts the intention that A , or the intention to believe that A , or the intention to make j believe that A . The latter corresponds to intention generation (v.i.).

$$\begin{array}{l} Bel_i Int_j A \wedge \neg Bel_i \neg A \wedge \\ \neg Int_i \neg A \rightarrow Int_i Bel_i A \end{array} \quad (\text{Adopt}_{\text{Int1}})$$

$$\begin{array}{l} Bel_i Int_j A \wedge Bel_i \neg A \wedge \\ \neg Int_i \neg A \rightarrow Int_i A \end{array} \quad (\text{Adopt}_{\text{Int2}})$$

Note that these axioms are consistent with the sincerity axiom ($\text{Sinc}_{\text{Inform}}$).

We stress that in the second case ($\text{Adopt}_{\text{Int2}}$), as soon as i believes A he should automatically generate the intention to make j believe that A . This will be done by the next axiom.

5.3 Intention generation

$$Bel_i Int_j A \wedge Bel_i A \rightarrow Int_i Bel_j A \quad (\text{Gen}_{\text{Int1}})$$

This axiom expresses the principle that agents generate intentions in order to satisfy other agents' goals. Several other axioms contribute to

this principle:

$$Bel_i(A \wedge Bel_j \neg A \rightarrow Int_i Bel_j A) \quad (\text{Gen}_{Int2})$$

$$Bel_i(A \wedge Done_{\langle j, \text{informs } A \rangle} Bel_i \neg A \rightarrow Int_i Bel_j Bel_i A) \quad (\text{Gen}_{Int3})$$

$$Bel_i(Done_\alpha(Done_\gamma \top \wedge Bel_i Done_\beta \top) \rightarrow Int_i Bel_j Bel_i Done_\alpha Done_\gamma \top) \quad (\text{Gen}_{Int4})$$

The last axiom expresses that i cannot sincerely intend to make j believe A without believing A himself.

$$Int_i Bel_j A \rightarrow Bel_i A \vee Int_i Bel_i A \quad (\text{Sinc}_i)$$

6 Semantics

As we have seen, except the operators Int_i , we only have normal modal operators. For all of our axioms characterizing them, the famous modular completeness result due to Sahlqvist [Sahlqvist, 1975] applies, and we get for free a possible worlds semantics for our logic based on accessibility relations.

The modal operators Int_i are non-normal. Their logic is that of a classical modal logic, having a neighborhood semantics [Chellas, 1980]. These models can be combined with the accessibility relation models, and completeness of the resulting multi-modal logic can be proven in a fairly standard way for most of the axioms.

Theorem proving. In [Fariñas del Cerro and Herzig, 1995, Gasquet and Herzig, 1996] it is shown that intention operators can be translated to normal modal logics: $Int_i A$ becomes $\neg \Box_{i,1} \neg (\Box_{i,2} A \wedge \Box_{i,3} \neg A)$, where $\Box_{i,1}$, $\Box_{i,2}$ and $\Box_{i,3}$ are normal modal operators.

We currently investigate tableau theorem proving algorithms for our logic, and we have already implemented part of the logic. In [Castilho et al., 1997] the theoretical basis of the `Lotrec` generic tableau prover (which is still under development) was presented. As soon as semantical completeness is ensured, `Lotrec` offers an easy way of implementing sound and complete tableau method for our logic. There still remains to address the termination issue (and of course the decidability).

7 Discussion and related work

In the last years existing solutions to the frame problem in the Situation Calculus and in the Fluent Calculus have been extended to handle the dynamics of knowledge. All these approaches divide actions in perceptive and nonperceptive actions, and use semantic versions of the axiom (NonPerc₂) to construct the beliefs of an agent after an action from his previous beliefs. Scherl and Levesque [Scherl and Levesque, 1993] were the first to integrate what they called a regression axiom for knowledge into the Situation Calculus. For nonperceptive actions it writes in our language $\neg After_\alpha \perp \rightarrow (After_\alpha Knows_i A \leftrightarrow Knows_i After_\alpha A)$

Subsequent work has tried to integrate more of Cohen and Levesque’s theory of rational interaction [Cohen and Levesque, 1990a], in particular the notions of belief and goal. We here concentrate on these extensions, and give a detailed critique.

7.1 Integrating belief and revision

Scherl and Levesque’s regression axiom for knowledge must be weakened if one tries to apply their approach to belief (as opposed to knowledge). This is what we did in (NonPerc₂): $(\neg After_\alpha \perp \wedge \neg Bel_i After_\alpha \perp) \rightarrow (After_\alpha Bel_i A \leftrightarrow Bel_i After_\alpha A)$ This does not tell us anything about the beliefs of i in the state where α has been executed, and i believed before that α was inexecutable. This requires a revision of i ’s beliefs.

In a recent proposal, Shapiro et col. add to the Scherl and Levesque framework a revision-like operation based on plausibility orderings [Shapiro et al., 1999]. They define $Bel_i A$ as truth of A in the most plausible states among the states possible for i . If a sensing action eliminates the most plausible states from the states possible for i , then previously less plausible states become the most plausible ones.

While being intuitively appealing, such a solution has several drawbacks. (1) As the authors note, it seems to be restricted to nondeterministic actions. (2) “The specification of [the plausibility ordering] over the initial situation is the responsibility of the axiomatizer of the domain”. This is particularly demanding because (3) in order to guarantee that after α the set of states for i is nonempty, the authors require the set of states possible for i to “contain enough situations initially”. This restricts the agent’s ‘doxastic freedom’ in an important way. Worse, it is not sufficient if there are three or more communicating agents and more than one conversation turn.

Consider the following example: agent k is competent at atom p , and

j is not. Agent i is completely ignorant in the initial state: this can be expected to guarantee that his possible states contain enough situations. Suppose all of these situations are equally plausible for i . Then (under adequate hypotheses of cooperation) we can expect that if j asserts p , then i adopts p , i.e. $After_{(j, \text{informs } p)} Bel_i p$. Moreover, as all states were equally plausible, p holds in every world possible for i . Therefore when subsequently k asserts $\neg p$, i will unavoidably move to an empty set of possible states.

7.2 Integrating goals

In a series of papers, Shapiro et col. have added the notion of goal to the Situation Calculus. The proposals are all based on the notion of knowledge (and not belief), public actions and differ in the regression axiom for goals. As the authors themselves note, those in [Shapiro et al., 1997, Shapiro and Lespérance, pear] lead to so-called fanatic agents, who never abandon their goals (even when they learn that they became true). In [Shapiro et al., 1998] every goal A comes with a canceling condition B associated to it. Once i has adopted A , he can abandon A when he learns that B is true. Nevertheless, other agents are still free to communicate goals with canceling condition \top , which can never be abandoned.

It seems to us that the difficulties are inherent to the choice of defining the goals after an action by a successor state axiom. The latter requires expressing the resulting goals explicitly as a function of the previous mental state and the new information. This is not modular enough, in the sense that all the cognitive processes that are involved when i achieves a rational balance among his mental attitudes must be taken into account in that axiom. To witness, the three versions of the successor state axiom for goals differ according to the underlying hypotheses concerning trust and sincerity.

8 Conclusion

We have defined a modal logic of belief, intention and action. We have supposed that actions are public and that agents are sincere. Under these hypotheses we have formalized principles of rational interaction and cooperation (in particular supposing sincerity). We have given axioms for the preservation of mental attitudes (based on the notion of influence of an action), and for their adoption (based on the notion of competence of an agent). We have then focussed on the analysis informative acts,

and have shown that due to cooperativity they can support requests and yes-no-queries.

We have restricted our attention to propositional logic. Nevertheless, our solution to the Frame Problem can be extended to first-order logic [Herzig and Longin, 2000].

References

- [Bratman, 1987] Bratman, M. E. (1987). *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, MA.
- [Castilho et al., 1997] Castilho, M. A., Fariñas del Cerro, L., Gasquet, O., and Herzig, A. (1997). Modal tableaux with propagation rules and structural rules. *Fundamenta Informaticae*, 32(3/4):281–297.
- [Castilho et al., 1999] Castilho, M. A., Gasquet, O., and Herzig, A. (1999). Formalizing action and change in modal logic I: the frame problem. *Journal of Logic and Computation*, 9(5):701–735.
- [Chellas, 1980] Chellas, B. (1980). *Modal logic: An introduction*. Cambridge University Press.
- [Cohen and Levesque, 1990a] Cohen, P. R. and Levesque, H. J. (1990a). Intention is choice with commitment. *Artificial Intelligence Journal*, 42(2–3).
- [Cohen and Levesque, 1990b] Cohen, P. R. and Levesque, H. J. (1990b). Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press.
- [Fariñas del Cerro and Herzig, 1995] Fariñas del Cerro, L. and Herzig, A. (1995). Modal deduction with applications in epistemic and temporal logic. In Gabbay, D., Hogger, C. J., and Robinson, J. A., editors, *Handbook of Logic and Artificial Intelligence*, volume 4 - Epistemic and Temporal Reasoning, pages 499–594. Oxford University Press.
- [Gasquet and Herzig, 1996] Gasquet, O. and Herzig, A. (1996). From classical to normal modal logics. In Wansing, H., editor, *Proof Theory of Modal Logics*, number 2 in Applied Logic Series. Kluwer Academic Publishers.
- [Giunchiglia et al., 1997] Giunchiglia, E., Kartha, G. N., and Lifschitz, V. (1997). Representing action: indeterminacy and ramifications. *Artificial Intelligence Journal*, 95.
- [Harel, 1984] Harel, D. (1984). Dynamic logic. In Gabbay, D. M. and Günthner, F., editors, *Handbook of Philosophical Logic*, volume II, pages 497–604. D. Reidel, Dordrecht.

- [Herzig and Longin, 2000] Herzig, A. and Longin, D. (2000). Belief dynamics in cooperative dialogues. *J. of Semantics*, 17(2). vol. published in 2001.
- [Konolige and Pollack, 1993] Konolige, K. and Pollack, M. E. (1993). A representationalist theory of intention. In *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*. Morgan Kaufmann Publishers.
- [Rao and Georgeff, 1991] Rao, A. S. and Georgeff, M. P. (1991). Modeling rational agents within a BDI-architecture. In Allen, J. A., Fikes, R., and Sandewall, E., editors, *Proc. Second Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484. Morgan Kaufmann Publishers.
- [Sadek, 1992] Sadek, M. D. (1992). A study in the logic of intention. In Nebel, B., Rich, C., and Swartout, W., editors, *Proc. Third Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 462–473. Morgan Kaufmann Publishers.
- [Sadek, 2000] Sadek, M. D. (2000). Dialogue acts are rational plans. In Taylor, M., Néel, F., and Bouwhuis, D., editors, *The structure of multimodal dialogue*, pages 167–188, Philadelphia/Amsterdam. John Benjamins publishing company. From ESCA/ETRW, Workshop on The Structure of Multimodal Dialogue (Venaco II), 1991.
- [Sahlqvist, 1975] Sahlqvist, H. (1975). Completeness and correspondence in the first and second order semantics for modal logics. In Kanger, S., editor, *Proc. 3rd Scandinavian Logic Symposium*, volume 82 of *Studies in Logic*.
- [Sandewall, 1994] Sandewall, E. (1994). *Features and Fluents*. Oxford University Press.
- [Scherl and Levesque, 1993] Scherl, R. and Levesque, H. J. (1993). The frame problem and knowledge producing actions. In *Proc. Nat. Conf. on AI (AAAI'93)*, pages 689–695. AAAI Press.
- [Searle, 1983] Searle, J. R. (1983). *Intentionality: An essay in the philosophy of mind*. Cambridge University Press.
- [Shapiro and Lespérance, pear] Shapiro, S. and Lespérance, Y. (To appear). Modeling multiagent systems with the cognitive agents specification language - a feature interaction resolution application. In

- Castelfranchi, C. and Lespérance, Y., editors, *Intelligent Agents Vol. VII - Proc. 2000 Workshop on Agent Theories, Architectures, and Languages (ATAL-2000)*. Springer-Verlag.
- [Shapiro et al., 1997] Shapiro, S., Lespérance, Y., and Levesque, H. J. (1997). Specifying communicative multi-agent systems with ConGolog. In *Working notes of the AAAI fall symposium on Communicative Action in Humans and Machines*, pages 75–82. AAAI Press.
- [Shapiro et al., 1998] Shapiro, S., Lespérance, Y., and Levesque, H. J. (1998). Specifying communicative multi-agent systems. In Wobcke, W., Pagnucco, M., and Zhang, C., editors, *Agents and Multi-Agent Systems - Formalisms, Methodologies, and Applications*, pages 1–14. Springer-Verlag, LNAI 1441.
- [Shapiro et al., 1999] Shapiro, S., Pagnucco, M., Lespérance, Y., and Levesque, H. J. (1999). Iterated belief change in the situation calculus. In *Proc. IJCAI'99*, pages 527–538.
- [Thielscher, 1995] Thielscher, M. (1995). Computing ramifications by postprocessing. In *Proc. 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, pages 1994–2000, Montreal, Canada.
- [Thielscher, 2000] Thielscher, M. (2000). Representing the knowledge of a robot. In Cohn, A., Giunchiglia, F., and Selman, B., editors, *Proc. KR'00*, pages 109–120. Morgan Kaufmann.