Intention Dynamics -preliminary report-

Andreas Herzig, Dominique Longin

Intention Dynamics
– preliminary report –

Andreas Herzig Dominique Longin

Institut de Recherche en Informatique de Toulouse
Équipe Logic, Interaction, LAnguage and Computation
118 Route de Narbonne, F-31062 Toulouse cedex 4
{herzig,longin}@irit.fr
http://www.irit.fr/ACTIVITES/LILaC/

April 15, 2002

Technical Report – IRIT/2002-12-R

Abstract

The aim of this paper is to describe the evolution of an agent's intentions in cooperative contexts. The role of intentions in BDI architectures being to ensure stable behaviour of agents, they must nevertheless be abandoned when they are fulfilled. The other way round, new intentions must be generated (in particular in order to respect cooperation principles).

We propose a logic of knowledge, intention and action. Building on Scherl & Levesque's solution to the frame problem for knowledge, we give a new solution to the frame problem for intentions in terms of axioms regulating preservation, abandonment, and cooperative generation of intentions.

Keywords: Autonomous Agents, Cognitive Robotics , Reasoning about Actions and Change.

1 INTRODUCTION

Intention is a central mental attitude in agent architectures. For Searle it is one of the key notions in the structure of behaviour [21]. Intention is of primary importance in reasoning about actions [1].

In this paper we analyse the dynamic aspects of the notion of intention. We want to describe how by the occurrence of actions some intentions are generated, others are dropped, and some others are preserved. There are a lot of reasons to generate intentions, but here we are only interested in those related to cooperation.

We nevertheless do not handle here reconsideration of intentions, which we view as a different process.

To formally analyze intentions we need an account of actions. In order to simplify things we make an assumption that is often made in the domain of reasoning about actions: we suppose that the agents' perception of action occurrences is correct and complete. This means that (1) if an agent believes that some action a occurred then a indeed occurred (correctness), and (2) if a occurred then the agent believes a occurred (completeness). The information acquired by agents through actions thus faithfully reflects reality. It is more natural in such a framework to consider *knowledge* instead of *belief*.[1]

In the sequel we propose an epistemic dynamic logic where intention is formalized in a non-normal modal logic (Sect. 2). We give axioms linking intention and knowledge (Sect. 3), and study the different aspects of intention dynamics: abandonment (Sect. 4), generation (Sect. 5), and preservation (Sect. 6).

2 FORMAL FRAMEWORK

Based on the philosophical theories of Searle [20] and Bratman [1], our logic follows the tradition of Cohen & Levesque [5, 6] and Sadek [15]. The only difference with our work is that we use knowledge rather than belief.

Our language contains modal operators of knowledge, intention, and action. It is a first-order multimodal logic without equality and function symbols (although the first-order aspect is not important here), and with a possible worlds semantics in terms of accessibility relations and neighborhood functions for intention.

Atomic formulas are noted p, q, ... or $P(t_1, ..., t_n)$, and $\mathcal{ATM}$ is the set of all atomic formulas. The formulas will be denoted by A, B, ... We say that a formula is *factual* if it contains no modal operator.

Knowledge. Let $\mathcal{AGT} = \{i, j, k, \ldots\}$ be the set of agents. We associate a modal operator of knowledge K_i to every $i \in \mathcal{AGT}$. The formula $K_i A$ is read "agent i knows that A". $Kif_i A$ is an abbreviation of $K_i A \vee K_i \neg A$, and reads "agent i knows whether A is true or

[1] There are approaches that do without such hypotheses, but the price to pay is a more complex formalism [12, 7, 25]: for example one must apply belief revision when an agent realizes that his beliefs are incorrect.

not". We adopt the modal logic S5 as the logic of knowledge. An important property of this logic is the (T) axiom for K : $K_iA \rightarrow A$.

Intention. Intention is a fundamental mental attitude, because it is at the origin of every voluntary action. We associate a modal operator of intention I_i to every $i \in \mathcal{AGT}$, and read the formula I_iA as "agent i intends that A". Intention is neither closed under logical truth, nor under logical consequence, conjunction, and material implication. We only postulate:

$$\frac{A \leftrightarrow B}{I_iA \leftrightarrow I_iB} \qquad (\text{RE}_{I_i})$$

This is in accordance with [1, 5, 15], but contrarily to these approaches, intention is primitive here, as in [14, 11, 4]. In [14, 11] only closure under logical consequence had been given up, and we thus generalize their semantics. This conforms with the majority of approaches in the agent community. Such a choice also enables more or less standard completeness techniques and proof methods.

Dynamic Epistemic Logic. We use a simple version of PDL [10] to speak about actions. To each action a there is associated a modal operator $After_a$. An example of a formula is $K_i\neg After_a\bot$, expressing that the agent knows that a can be executed. The operator $Feasible_a$ is introduced as an abbreviation: $Feasible_aA \stackrel{\text{def.}}{=} \neg After_a\neg A$.

We adopt the standard axiomatics of PDL, which for our fragment is nothing but the multimodal logic K. ($After_a$ corresponds to the Dynamic Logic operator $[a]$, and $Feasible_a$ to $\langle a\rangle$.)

Semantics. All the K_i operators are normal modal operators. For all of our axioms characterizing them, the famous modular completeness result due to Sahlqvist [16] applies, and we get for free a possible worlds semantics for our logic based on accessibility relations.

The modal operators I_i are non-normal. Their logic is that of a classical modal logic, having a neighborhood semantics [3]. These models can be combined with the models in terms of accessibility relations, and completeness of the resulting multi-modal logic can be proven in a fairly standard way for most of the axioms.

In [8] it is shown that non-normal modal operators can be translated to normal modal logics: I_iA 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.

3 INTENTION AND KNOWLEDGE

An agent's knowledge is supposed to partly reflect the state of the world. Thus, the interaction between intention and knowledge is crucial in order to link intentions to reality.

The most important constraint for a rational agent is formalized by the following axiom, that can also be found in [5, 14, 15]:

$$I_i A \rightarrow K_i \neg A \qquad (\text{Rel}_{\text{IK}}1)$$

It expresses that an agent can only hold the intention to achieve a property if he knows that that property is false.

Theorem 1 *The following properties can be derived from (Rel$_{\text{IK}}$1):*

$$I_i A \rightarrow \neg I_i \neg A \qquad (1)$$
$$I_i A \rightarrow \neg A \qquad (2)$$
$$K_i A \rightarrow \neg I_i A \qquad (3)$$
$$\neg K_i A \rightarrow \neg I_i \neg K_i A \qquad (4)$$

Proof 1 *(1) follows from (Rel$_{\text{IK}}$1) and the (T) axiom of the modal logic for knowledge.*

(Rel$_{\text{IK}}$1) combined with the (T) axiom of the modal logic for knowledge entails (2).

By (Rel$_{\text{IK}}$1) and the axiom (D) of the modal logic for knowledge, we have $I_i A \rightarrow \neg K_i A$. From this (3) follows by contraposition.

(4) can be shown from axiom (5) for knowledge together with (3): an instance of that theorem is $K_i \neg K_i A \rightarrow \neg I_i \neg K_i A$. Then axiom (5) being $\neg K_i A \rightarrow K_i \neg K_i A$, we thus obtain $\neg K_i A \rightarrow \neg I_i \neg K_i A$. □

Axiom (Rel$_{\text{IK}}$1) has been criticized in the literature, because it strongly links intention to knowledge. For example, if an agent does not know whether the light is off in a room, he will not be able to intend to switch it off. Formally, if p denotes "the light is off", then $\neg Kif_i p \wedge I_i p$ is a contradiction (by definition of Kif_i and by (Rel$_{\text{IK}}$1)). (That aspect is illustrated even more radically by (2), which says that if an agent has the intention that A then A is false.)

Generally, a "rational behavior" is to consider that the agent should go to the room, and if the light is already off, drop his intention to switch the light off (because his intention is already satisfied). Thus, we would be tempted to weaken (Rel$_{\text{IK}}$1) to: $I_i A \rightarrow \neg K_i A$. Then an agent could ignore whether p is true, and at the same time intend that p.

But we must keep in mind that according to Sadek [15], intention is a mental attitude that commits us (in a persistent manner) to *achieve* a goal. Hence there are in fact two intentions here: (1) in a first step, there is the *intention to know if* the light is on or off; (2) in a second step, there is the intention to switch the light off if it is on. Generally, it might be said that it is not rational to seek to achieve a goal which may already hold (although here are cases where, by caution or temporal constraints, we perform an action whose goal might already hold). Thus, generally, before intending to switch off the light, we check whether the light is on. The idea underlying (Rel$_{\text{IK}}$1) is that each time the agent is in doubt whether it is necessary to generate an intention (as in the previous example), he should first intend to know the state of the world. And only if this state does not satisfy this property, he will then intend to achieve it.

In the rest of this section we investigate how the interplay between I_iA and I_iK_iA can be formally captured.

As far as we know, the only work addressing this problem is [15], where a new mental attitude *want* is proposed (also named *potential intention*). This mental attitude abbreviates $K_iA \vee I_iK_iA$. It follows from the tautologie $K_iA \rightarrow (K_iA \vee I_iK_iA)$ that if an agent knows A, then he wants A.

Instead of a want operator we focus on the intention to know. Here, "to intend to know" refers to an introspective mechanism. Thus, an instance of (3) of theorem 1 is:

$$K_iK_iA \rightarrow \neg I_iK_iA$$

In others words, an agent cannot want to know A, if he is aware that he already knows A.

We propose to add to (Rel$_{\text{IK}}$1) a second principle as formalized by the following axiom:

$$(I_iK_iA \wedge K_i\neg A) \rightarrow I_iA \qquad \text{(Rel}_{\text{IK}}\text{2)}$$

This axiom is read "if an agent knows A is false and intends to know A then he will intend A". Suppose i intends to know A, but actually he knows $\neg A$; then i should be prepared to act in order to change the world, which justifies the intention that A. The other way round, if i ignores whether A is true or not, then the intention to know that A can be held without holding the intention to *act* in order to bring about A.

There seems to be no similar axiom in the literature. It allows us to prove that $(I_iK_iA \wedge \neg I_iA) \rightarrow \neg Kif_iA$. Hence if i intends to know A without intending A, then he ignores whether A.[2]

Finally, our third constitutive property of the rational balance between intention and knowledge is the following axiom (that is derived in [15] from properties of the more basic notion of goal):

$$I_iA \rightarrow I_iK_iA \qquad \text{(Rel}_{\text{IK}}\text{3)}$$

This means that an agent cannot intend A without intending to know A.

The converse should not be valid: we can intend to know A without intending A. Suppose e.g. you ignore whether the light in your neighborhood room is off or not, and you intend to know that it is off. In this case you are prepared to act in order to acquire that knowledge, typically by a sensing action (checking that it is indeed off), but you are not necessarily prepared to switch it off: the latter intention might be generated in a second stage when realizing that the light is on.

Note that it follows from (Rel$_{\text{IK}}$1) and (Rel$_{\text{IK}}$3) that (Rel$_{\text{IK}}$2) is an equivalence:

$$(I_iK_iA \wedge K_i\neg A) \leftrightarrow I_iA$$

Finally, there are two essential properties related to an agent's introspection capacity (cf. [15]):

$$I_iA \rightarrow K_iI_iA \qquad \text{(Rel}_{\text{IK}}\text{4)}$$

$$\neg I_iA \rightarrow K_i\neg I_iA \qquad \text{(Rel}_{\text{IK}}\text{5)}$$

[2] From (Rel$_{\text{IK}}$2) it follows that $(I_iK_iA \wedge \neg I_iA) \rightarrow \neg K_i\neg A$. From (Rel$_{\text{IK}}$1) it follows that $I_iK_iA \rightarrow K_i\neg K_iA$, whose consequent is equivalent to $\neg K_iA$ in S5.

These implications are in fact equivalences: this can be obtained via the (T) axiom of the modal logic for knowledge. The two equivalences express that intentions (respectively, non-intentions) of an agent are sound and complete with respect to his known intentions (respectively, non-intentions).

4 ABANDONMENT OF INTENTIONS

Intentions should contribute to an agent's stability, and should in general be maintained as long as they are not fulfilled. But the agent should not be fanatic, and should be prepared to drop intentions under some circumstances. It is generally admitted that there are principally three reasons to do that [5]:

1. the intention is satisfied;
2. the agent learns that the intention can never be satisfied;
3. the intention must be abandoned for some other reason.

The first case (1) is encoded in theorem (3), which stipulates that if an agent knows that A then he does not have the intention that A.

The idea behind case (3) can be captured by axioms of the form $A' \rightarrow I_i A$. In other words, if some fact A' holds then agent i no longer has the intention that A. For example if i has the intention to open a safe in order to steal the money contained in it, then as soon as he learns that the safe is empty i has no more reason to entertain that intention. A' here is the fact that the (superseding) intention of stealing the money is abandoned. Generally, A' is some property which renders superfluous the intention to bring about A.

Letting $\Diamond$ represent the "possible" operator of alethic logic, the case (2) can be formalized as: $I_i A \rightarrow K_i \Diamond A$ (if i has the intention that A then i knows that A is possible). The latter formula is entails $K_i \neg \Diamond A \rightarrow \neg I_i A$ (if i knows that A is impossible then he doesn't have the intention that A). We have neglected here case (2), in order to avoid introducing a temporal operator into our logic.

5 INTENTION GENERATION VIA COOPERATIVE PRINCIPLES

Agent have a lot of motivations that might generate intentions, such as desire for money, pleasure, beauty, ... Just anything might make us act. In the sequel we focus on the generation of intentions via some cooperation principles again, the latter might be numerous and varying. Note that our axioms here are an open list, that might be augmented by other axioms in order to capture other aspects of intention generation.

5.1 Intention generation

Suppose i is cooperative, and suppose i has learned that j has the intention that A, i.e. $K_i I_j A$. To which extent should i accomodate j's intention? Intuitively, the difficulty is to take into account the preceding axiom (Rel$_{IK}$1) in an appropriate way: i should only generate the intention to bring about A when i knows that A is currently false. We formalize this in the sequel.

When i doesn't know that A is currently true ($\neg K_i A$) then i does not necessarily entertain the intention that A be true.[3] If we rewrite this we obtain our central axiom:

$$(K_i I_j A \wedge \neg K_i A \wedge \neg I_i K_i \neg A) \rightarrow I_i K_i A \qquad \text{(Gen}_I\text{1)}$$

Remark 1 *As we have said, we did not include in our axioms that an agent i cannot entertain the intention that A if he knows that A is always false. This means that i must abandon his intention $I_i A$ as soon as i starts to know that A will always be false, and the other way round $I_i A$ cannot be generated if i knows that A will always be false. We can constrain axiom (Gen$_I$1) in order to guarantee this.*

Remark 2 *(Gen$_I$1) is too strong if there are more than two agents. Indeed, suppose that i cooperates with both j and k, and that i thinks j and k have contradictory intentions: $K_i I_j A \wedge K_i I_k \neg A$. Suppose moreover that $\neg Kif_i A \wedge \neg I_i \neg A \wedge \neg I_i A$ (i.e. i doesn't bother at all about A). Then by (Gen$_I$1) i generates the intentions $I_i K_i A$ and $I_i K_i \neg A$. But this is inconsistent according to (Rel$_{IK}$1).*

A way of taking into account such possible inconsistencies is to weaken (Gen$_I$1) by adding to the premisses the condition that j's intention must be consistent with the intentions i attributes to the other agents:

$$(K_i I_j A \wedge \neg K_i A \wedge C) \rightarrow I_i K_i A$$

where C is a formula of the form $\neg K_i I_{k_1} \neg A \wedge \ldots \wedge \neg K_i\ I_{k_n} \neg A$. Such a condition C might also take into account priorities and preferences of i w.r.t. the intentions of his fellow agents.

Another way of weakening (Gen$_I$1) is to stipulate that i cannot stay without taking position as soon as he learns that j and k have inconsistent intentions. This can be formalized by a principle such as

$$K_i I_j A \rightarrow (I_i A \vee I_i \neg A)$$

5.2 Intention generation: derived principles

In the rest of the section we discuss two other important principles, and we show that they can be derived from our central axiom.

[3] Indeed, if moreover $\neg K_i \neg A$ then i doesn't know whether A is currently true or not, and it cannot be the case that $I_i A$ because $\neg K_i \neg A$ implies $\neg I_i A$. The only thing that can be guaranteed here is that i adopts *the intention to know* that A (cf. our discussion about (Rel$_{IK}$1) in Sect. 3).

First of all, note that by theorem (4) the second premiss $\neg K_i A$ of our central axiom (Gen$_I$1) ensures that i will not generate the intention to know A if i already has a contradictory intention.

In accordance with (Rel$_{IK}$2) and(Gen$_I$1) we have now that if an agent i knows that an agent j has the intention that A be true, and i does not have the intention that A be false, then i adopts the intention that A be true. By theorem (3), if agent i knows that A is false then he cannot have the intention that A be false. Putting this together we obtain:

Theorem 2 *Axiom (Gen$_I$1) implies*

$$(K_i I_j A \wedge K_i \neg A) \rightarrow I_i A \qquad \text{(Gen}_I\text{2)}$$

Hence our first principle (Gen$_I$2) says that if i knows that the world must necessarily change, then j's intention is directly adopted. This axiom accounts for a particular form of intention generation that can be called intention adoption.

Proof 2 *The hypothesis is $K_i I_j A \wedge K_i \neg A$. On the one hand, $K_i \neg A \rightarrow \neg K_i A$ with axiom (D), and we thus obtain the second hypothesis of (Gen$_I$1).*
To establish the third hypothesis of (Gen$_I$1) we proceed as follows: first, we derive $I_i K_i \neg A \rightarrow K_i \neg K_i \neg A$ with (Rel$_{IK}$1). Then $K_i \neg K_i \neg A \rightarrow \neg K_i \neg A$ with the modal axiom (5) of negative introspection. We thus obtain $I_i K_i \neg A \rightarrow \neg K_i \neg A$, and by contraposition $K_i \neg A \rightarrow \neg I_i K_i \neg A$.
In consequence the hypotheses of axiom (Gen$_I$2) imply those of axiom (Gen$_I$1). The latter allows us to obtain $I_i K_i A$:

$$(K_i I_j A \wedge K_i \neg A) \rightarrow I_i K_i A$$

Now $(I_i K_i A \wedge K_i \neg A) \rightarrow I_i A$ with (Rel$_{IK}$2), entailing (Gen$_I$2). □

The second principle of intention generation stipulates that if agent i knows that agent j has the intention that A, and i knows that A is currently true, then i will generate the intention that j knows A.

Theorem 3 *Axiom (Gen$_I$1) implies*

$$K_i I_j A \rightarrow I_i K_j A \qquad \text{(Gen}_I\text{3)}$$

Proof 3 *(Gen$_I$1) allows to derive (Gen$_I$2), of which $(K_i I_j K_j A \wedge K_i \neg K_j A) \rightarrow I_i K_j A$ are an instance.*
As $I_j K_j A$ implies $K_j \neg K_j A$ by (Rel$_{IK}$1), we have $K_i I_j K_j A \rightarrow K_i K_j \neg K_j A$ by the principles of modal logic K. As $K_j \neg K_j A$ is equivalent to $\neg K_j A$ in S5, we obtain $K_i I_j K_j A \rightarrow K_i \neg K_j A$.
Thus we obtain $K_i I_j K_j A \rightarrow I_i K_j A$ from (Gen$_I$2).
On the other hand, as $I_j A$ implies $I_j K_j A$ by (Rel$_{IK}$3), we have $K_i I_j A \rightarrow K_i I_j K_j A$ by the principles of modal logic K.
Finally transitivity of $\rightarrow$ allows us to conclude that $K_i I_j A \rightarrow I_i K_j A$. □

Remark 3 *Conditions $K_i I_j A$ and $K_i A$ of (Gen$_I$3) cannot be simultaneously true. Indeed, $K_i I_j A \rightarrow K_i K_j \neg A$ by (Rel$_{IK}$1) on the one hand, and on the other hand $K_i K_j \neg A \rightarrow K_i \neg A$ is a theorem of the S5 logic for knowledge. It follows that $K_i A$ cannot be the case because $K_i A$ and $K_i \neg A$ are inconsistent.*

To sum it up, our central axiom allows us to derive natural and powerful principles of cooperation.

6 PRESERVATION OF MENTAL ATTITUDES

Just as for intention generation, intention preservation is intimately linked to knowledge preservation. In the sequel we show that the solution to the frame problem for knowledge induces a solution to the frame problem for intention via an axiom linking action, knowledge, and intention.

6.1 Preserving knowledge

Semantically a (non-deterministic) action is a relation R_a between possible situations – alias possible worlds –, where $w' \in R_a(w)$ means that w' is a possible result of a when applied in w. We view the belief state of an agent in a given situation w as a set of possible worlds $R_{K_i}(w)$, and $v \in R_{K_i}(w)$ means that the situation v is compatible with the agent's beliefs. The occurrence of an action makes the current situation w evolve to a new situation $w' \in R_a(w)$. What can we say about $R_{K_i}(w')$, i.e. the agent's belief state at w'?

Following Moore [13] and Scherl and Levesque [18], the agent's belief state $R_{K_i}(w')$ in w' results from applying action a to all possible worlds in $R_{K_i}(w)$ ('mentally executing a'), and collecting the resulting situations:

$$R_{K_i}(w') = \bigcup_{v \in R_{K_i}(w)} R_a(v)$$

Syntactically, we obtain a generalization of the successor state axiom for knowledge of [18] to non-deterministic actions:

$$\neg After_a \bot \rightarrow (Feasible_a K_i A \leftrightarrow K_i After_a A) \qquad \text{(SSA)}$$

It says that the agent cannot observe anything after a is performed: indeed, for any formula A, if he cannot predict before a is performed that A will hold after a is performed, then he will not know A after a is performed.

6.2 Preserving facts

Given the successor state axiom we can reuse non-epistemic solutions to the frame problem. Just as Scherl and Levesque have applied Reiter's solution [19] we use the solution of [2].

Which knowledge 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 atomic facts*. If there exists a relation of influence between the action and an atomic fact, this fact cannot be preserved.[4] Thus, $a \rightsquigarrow q$ means "the action a influences the truth value of q". We suppose that $\rightsquigarrow$ is in metalanguage. We note $a \not\rightsquigarrow A$ when for every atom p occurring in A, $a \rightsquigarrow p$ does not hold.

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

$$A \rightarrow After_{\alpha} A \qquad \text{if} \quad \alpha \not\rightsquigarrow A \text{ and } A \text{ is factual} \qquad \text{(Preserv}_1\text{)}$$

This expresses that if a has 'nothing to do' with A then A is preserved. The restriction that A must be factual is necessary to e.g. avoid $Feasible_b \top \rightarrow After_a Feasible_b \top$ (which is not necessarily the case).

6.3 Preserving intentions

The next axiom guarantees preservation of intentions as long as they are not achieved.

$$I_i A \rightarrow After_a (K_i \neg A \rightarrow I_i A) \qquad \text{(Preserv}_2\text{)}$$

The axiom says that if an agent intends that A then he abandons that intention only when he learns that A. Note that as said in section 4, this is slightly too strong because even if an intention A is not accomplished it should be dropped when the agent learns that A is impossible.

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

Theorem 4

$$\begin{array}{c} I_i A \rightarrow After_a I_i A \\ \textit{if } a \not\rightsquigarrow A \textit{ and } A \textit{ is factual} \end{array} \qquad (5)$$

Proof 4 *This follows from (Rel$_{IK}$1), (Preserv$_1$), (SSA), and (Preserv$_2$).*

Suppose A is factual and $a \not\rightsquigarrow A$. First, by (Rel$_{IK}$1) $I_i A \rightarrow K_i \neg A$.
Second, as $a \not\rightsquigarrow A$ we have $\neg A \rightarrow After_a \neg A$ by (Preserv$_1$).
The latter implies $K_i \neg A \rightarrow K_i After_a \neg A$ by standard modal principles (viz. necessitation and axiom (K)).
From that it follows with (SSA) that $K_i After_a \neg A \rightarrow After_a K_i \neg A$.
Putting all this together it follows that $I_i A \rightarrow After_a K_i \neg A$. From this and axiom (Preserv$_2$) it finally follows that $I_i A \rightarrow After_a I_i A$. □

[4] The 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 [17] occlusion, Thielscher's [26] influence relation, or Giunchiglia *et al.*'s [9] possibly changes operators.

7 DISCUSSION

We have presented a minimal logic for cooperative interaction. It is based on a primitive notion of intention satisfying the principle that the intention that A implies the belief that A is currently false. We have completed the principles that have been put forward in the literature by a new one, viz. that if A is ignored then intending to know A amounts to intending A. Our formal framework is thus relatively simple, and facilitates completeness results and theorem proving.

In a series of papers, Shapiro et col. have added the notion of goal to the Situation Calculus. All the different proposals are based on the notion of knowledge (and not belief) and public actions, and differ in the successor state axioms for goals. As the authors themselves note, the successor state axioms in [23, 22] lead to so-called fanatic agents, who never abandon their goals (even when they learn that they became true).

In [24] this is avoided by stipulating that every goal A comes with a cancelling condition B associated to it. As goals are adopted via cooperative communication, the cancelling condition is determined by the author of the message (and not by the receiver). Even if 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 cancelling 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 in the different papers differ according to the underlying hypotheses concerning trust and sincerity.

8 CONCLUSION

We have analyzed the dynamics of intentions in terms of intention abandonment, generation, and preservation. We have shown how intentions are abandoned by a principle linking intentions and knowledge. We have studied intention generation in a cooperative setting in terms of a set of axioms. Most importantly, we have argued that it is a difficult task to define intention preservation via a successor state axiom, and we have proposed an alternative that is more flexible and modular.

References

[1] Michael E. Bratman, *Intention, Plans, and Practical Reason*, Harvard University Press, Cambridge, MA, 1987.

[2] Marcos A. Castilho, Olivier Gasquet, and Andreas Herzig, 'Formalizing action and change in modal logic I: the frame problem', *Journal of Logic and Computation*, **9**(5), 701–735, (1999).

[3] B. F. Chellas, *Modal Logic: an introduction*, Cambridge University Press, 1980.

[4] Xiaopping Chen and Guiquan Liu, 'A Logic of Intention', in *Proc. 16th Int. Joint Conf. on Artificial Intelligence (IJCAI'99)*. Morgan Kaufmann Publishers, (1999).

[5] Philip R. Cohen and Hector J. Levesque, 'Intention is choice with commitment', *Artificial Intelligence Journal*, **42**(2–3), (1990).

[6] Philip R. Cohen and Hector J. Levesque, 'Rational interaction as the basis for communication', in *Intentions in Communication*, eds., Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, MIT Press, (1990).

[7] Robert Demolombe and Maria del Pilar Pozos Parra, 'Formalisation de l'évolution de croyances dans le Calcul des Situations', in *Proc. Journées Francophones Modèles Formels de l'Interaction (MFI'01)*, eds., B. Chaib-draa and P. Enjalbert, volume 2, pp. 205–217, (2001).

[8] Luis Fariñas del Cerro and Andreas Herzig, 'Modal deduction with applications in epistemic and temporal logic', in *Handbook of Logic and Artificial Intelligence*, eds., Dov Gabbay, Chris J. Hogger, and J. A. Robinson, volume 4 - Epistemic and Temporal Reasoning, 499–594, Oxford University Press, (1995).

[9] E. Giunchiglia, G. N. Kartha, and V. Lifschitz, 'Representing action: indeterminacy and ramifications', *Artificial Intelligence Journal*, **95**, (1997).

[10] David Harel, 'Dynamic logic', in *Handbook of Philosophical Logic*, eds., D. Gabbay and F. Guenthner, volume II, D. Reidel Publishing Company, (1984).

[11] Kurt Konolige and Martha E. Pollack, 'A representationalist theory of intention', in *Proc. 13th Int. Joint Conf. on Artificial Intelligence (IJCAI'93)*. Morgan Kaufmann Publishers, (1993).

[12] Yves Lespérance, Hector J. Levesque, and Raymond Reiter, 'A Situation Calculus approach to modeling and programming agents', in *Foundations and theories of Rational Agents*, eds., A. Rao and M. Wooldridge. Kluwer Academic Publishers, (1999).

[13] Robert C. Moore, 'A formal theory of knowledge and action', in *Formal Theories of the Commonsense World*, eds., J.R. Hobbs and R.C. Moore, 319–358, Ablex, Norwood, NJ, (1985).

[14] Anand S. Rao and Michael P. Georgeff, 'Modeling rational agents within a BDI-architecture', in *Proc. Second Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'91)*, eds., J. A. Allen, R. Fikes, and E. Sandewall, pp. 473–484. Morgan Kaufmann Publishers, (1991).

[15] M. D. Sadek, 'A study in the logic of intention', in *Proc. Third Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, eds., Bernhard Nebel, Charles Rich, and William Swartout, pp. 462–473. Morgan Kaufmann Publishers, (1992).

[16] H. Sahlqvist, 'Completeness and correspondence in the first and second order semantics for modal logics', in *Proc. 3rd Scandinavian Logic Symposium*, ed., S. Kanger, volume 82 of *Studies in Logic*, (1975).

[17] E. Sandewall, 'The range of applicability of some nonmonotonic logics for strict inertia', *J. of Logic and Computation*, **4**(5), 581–615, (1994).

[18] Richard Scherl and Hector J. Levesque, 'The frame problem and knowledge producing actions', in *Proc. Nat. Conf. on AI (AAAI'93)*, pp. 689–695. AAAI Press, (1993).

[19] Richard Scherl and Hector J. Levesque, 'The frame problem and knowledge producing actions', in *Proc. Nat. Conf. on AI (AAAI'93)*, pp. 689–695. AAAI Press, (1993).

[20] J. R. Searle, *Intentionality: An essay in the philosophy of mind*, Cambridge University Press, 1983.

[21] John R. Searle, *Minds; Brains and Science*, British Broadcasting Corporation, 1984.

[22] S. Shapiro and Yves Lespérance, 'Modeling multiagent systems with the cognitive agents specification language - a feature interaction resolution application', in *Intelligent Agents Vol. VII - Proc. 2000 Workshop on Agent Theories, Architectures, and Languages (ATAL-2000)*, eds., C. Castelfranchi and Y. Lespérance. Springer-Verlag, (2000).

[23] S. Shapiro, Yves Lespérance, and Hector J. Levesque, 'Specifying communicative multi-agent systems with ConGolog', in *Working notes of the AAAI fall symposium on Communicative Action in Humans and Machines*, pp. 75–82. AAAI Press, (1997).

[24] S. Shapiro, Yves Lespérance, and Hector J. Levesque, 'Specifying communicative multi-agent systems', in *Agents and Multi-Agent Systems - Formalisms, Methodologies, and Applications*, eds., W. Wobcke, M. Pagnucco, and C. Zhang, pp. 1–14. Springer-Verlag, LNAI 1441, (1998).

[25] S. Shapiro, M. Pagnucco, Y. Lespérance, and H. J. Levesque, 'Iterated belief change in the situation calculus', in *Proc. Seventh Int. Conf. on Principles of Knowledge Representation and Reasoning (KR 2000)*, pp. 527–538, (2000).

[26] Michael Thielscher, 'Computing ramifications by postprocessing', in *Proc. 14th Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, pp. 1994–2000, Montreal, Canada, (1995).