A logic of intention with cooperation principles and with assertive speech acts as communication primitives

2 authors, including:

Dominique Longin
Institut de Recherche en Informatique de T…
54 PUBLICATIONS   **512** CITATIONS

A Logic of Intention with Cooperation Principles and with Assertive Speech Acts as Communication Primitives

Andreas Herzig
IRIT
118 route de Narbonne
31062 Toulouse cedex 04, FRANCE
Andreas.Herzig@irit.fr

Dominique Longin [*]
IRIT
118 route de Narbonne
31062 Toulouse cedex 04, FRANCE
Dominique.Longin@irit.fr

ABSTRACT

We give a new logic of intention where (contrarily to Cohen&Levesque's approach) intention is a primitive modal operator having a non-normal possible worlds semantics. We then highlight the relation between intention and belief by a set of axioms. In our logic we formulate principles of cooperation allowing an agent to infer new intentions from his beliefs about other agents' intentions. Finally, building on results of linguistic pragmatics, we show that our cooperation principles allow to infer the effects of a yes-no question "Does A hold?" from that of an associated assertive "I have the intention to know whether A". In the same manner requests can be inferred, which form another important subclass of directives. It is the aim of this work to obtain a minimal logic that can be mechanized in a simple way.

Categories and Subject Descriptors

I.2.3 [**Deduction and Theorem Proving**]: Deduction, belief revision; I.2.4 [**Knowledge Representation Formalisms and Methods**]: Modal logic; I.2.11 [**Distributed Artificial Intelligence**]: Intelligent agents, Multiagent Systems

General Terms

Modal logic, speech act theory

Keywords

Logic of intention, belief, and action, cognitive robotics, cooperation

---

[*]Currently in Laboratoire Travail et Cognition, Maison de la Recherche, Université Toulouse-le-Mirail, 5 allées Antonio-Machado, 31058 Toulouse cedex 01.

1. INTRODUCTION

In this paper, we consider a BDI architecture for agents and want to describe the dynamics of their mental states.

Starting from the 'common belief' that the frame problem has been solved [16], existing solutions have been extended in the last years to handle the dynamics of knowledge [20, 29].

Modifying these solutions in order to handle belief is not easy.[1] We have shown in [11] that the recent proposal of [28] encounters some difficulties.

Our aim being to define the simplest dynamic doxastic logic, we focus on speech acts and try to define a minimal set of speech act types. Nevertheless, agents must be able to perform assertive speech acts, queries and requests.

We use recent results in linguistic pragmatics showing how what is called *indirect speech acts* can be inferred from *literal speech acts* (i.e. the act that has been literally performed), explaining thus how e.g. the literal assertive act "I want to have the salt." expresses the speaker's request "Pass the salt!". We exploit these theoretical results to encode requests within assertive speech acts. We show that in the same way yes-no questions of the form "Does A hold?" can be simulated by the associated assertive "I have the intention to know whether A."

We need for this a well-defined notion of intention and a fine grained relation between intention and belief. We have chosen to use a *primitive* notion of intention having a non-normal modal logic (neither closed under implication nor conjunction). It is opposed to the [7, 17] approaches where intention is built from the more basic concept of goal. Our modal operator of intention has a non-normal possible worlds semantics.

We study the interaction between intention and belief and give a new axiom.

Finally, we use this framework to define some cooperation principles allowing an agent to infer new intentions from his beliefs about other agents' intentions. We analyze and formalize all the principles associated to these notions.

The paper is organized as follows: we introduce first the philosophical motivations of this work (Sect. 2). Then we present the formal framework and in particular the modal operators of belief and intention (Sect. 3), show how belief and intention can be related (Sect. 4), and sketch the possible worlds semantics for the resulting logic (Sect. 5). We give some principles of cooperation (Sect. 6). Finally we

---

[1]'Knowledge' is viewed as true belief.

apply our results to examples (Sect. 7).

2. PHILOSOPHICAL MOTIVATIONS

In this section, we show how speech act theory can be used to simulate other classes of speech acts by assertive speech acts via cooperation principles.

Sometimes, by saying something, we want to say something else. Within *speech act theory* [1, 21] this idea has been exploited by Searle in [22] to construct the *theory of indirect speech acts.*

An indirect speech act is the speech act (indirectly) performed by performing another speech act (the 'direct' speech act). We also call these speech acts "literal speech act" and "non-literal speech act". Thus, the utterance of "Can you pass me the salt?" achieves the direct speech act corresponding to a yes-no question on the capacity of the hearer to pass the salt to the speaker.

Sometimes, this utterance achieves an (indirect) speech act corresponding to a request, where the speaker asks the hearer to pass him the salt.[2]

Legitimately, we may reject this dichotomy between direct and indirect speech acts. We have shown ourselves how results in psycholinguistics can attack the point of view where an indirect speech act is necessarily inferred from the direct speech act: a better point of view would be to associate different sets of effects to the same act according to the utterance context [5].

Neglecting such aspects of "cognitive realism" of the model we can nevertheless benefit from the fact that an assertive speech act allows us to extract its non-literal interpretation via cooperation principles.

It is important to note that we do not want to process indirect speech acts: we just want to exploit mechanisms underlying them in order to encode requests and yes-no queries by assertive speech acts.

In [4], Virbel extended Searle's approach of [22] by showing that indirect speech acts are performed by assertions or questions on three main types of arguments:

(a) the success conditions of the intended direct speech act (i.e. the speech act that we want to perform indirectly);

(b) reasons of doing (or not doing) the intended act;

(c) the planning of the intended act.

For example, let α be the speech act that is achieved by the utterance "Can you pass me the salt?". When that utterance is interpreted indirectly, its achievement produces an indirect speech act α' corresponding to the utterance "Pass me the salt!". The preparatory conditions of α' being that the speaker thinks that the hearer can pass him the salt, α is a yes-no question about the preparatory condition of α'.

Similarly, the sincerity condition of a request such as "Execute action β" being that the speaker wants the hearer to perform action β, an assertion of that sincerity condition ("I want you to execute β.") is a 'form of indirection' that is used to mean "Execute action β!".

Finally, a reason to ask somebody whether p is true being that one wants to know whether p is true, a request such as

"Inform me whether p is true" can be performed by asserting that reason, e.g. by the utterance "I want to know if p is true".

We have thus found a way of performing requests and yes-no questions by means of an assertion. In the sequel of this paper we shall give a formal framework where such non-assertive acts can indeed be inferred from assertive acts.

3. FORMAL FRAMEWORK

Based on the philosophical theories of Searle [23] and Bratman [2], our logic follows the tradition of Cohen & Levesque [7, 8] and Sadek [17, 18]. As the latter approaches, 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.

Our language is of the BDI sort, containing modal operators of belief, intention, and action. It is a first-order multimodal logic with neither equality nor 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 *objective* if it contains no modal operator.

Belief.. Let $\mathcal{AGT} = \{i, j, k, \ldots\}$ 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".[3] We adopt the modal logic KD45 as the logic of belief. This implies that we suppose that agents cannot entertain inconsistent beliefs, and that they are aware of their beliefs and of their disbeliefs. $Bel_{i,j} A$ is read "agents i and j mutually believe that A". Semantically, it is the reflexive transitive closure of union of dynamic accessibility relations.

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 $Int_i A$ as "agent i intends that A".

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

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

This is in accordance with [2, 7, 17], but contrarily to these approaches, intention is primitive here, as in [15, 14]. In the latter two only closure under logical consequence had been given up, and we thus generalize their 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

---

[2]Note this is a possibility, not a necessity. This phenomenon has been stated as "every indirect speech act is *cancellable*" [30].

[3]We use the term "knows" here because "i believes if A" sounds odd.

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.

Speech acts.. In accordance with speech act theory, an act is represented by an illocutionary force and a propositional content [21]. As we only work with speech acts of assertive type, we do not mention the illocutionary force. Thus, an assertive speech act is a tuple of the form:

$$\langle i, j, A \rangle$$

where i is the author, j the addressee, and A is a formula representing the propositional content of the act. For example, $\langle u, s, Blue(\text{sky}) \rangle$ represents the assertive speech act achieved by agent u's utterance towards agent s: "The sky is blue".

We suppose that utterances are public, i.e. although $\langle i, j, A \rangle$ is addressed to j, every other agent k perceives that act.

Action operators.. To each action α there are associated modal operators $After_\alpha$ and $Before_\alpha$. For example, $Bel_i Before_{\langle j,i,p \rangle} Bel_j p$ (the agent i believes that before informing that p, j believed that p, i.e. j is sincere w.r.t. to p).

The operators $Feasible_\alpha$ and $Done_\alpha$ are introduced by stipulating that $Feasible_\alpha A$ abbreviates $\neg After_\alpha \neg A$, and $Done_\alpha A$ abbreviates $\neg Before_\alpha \neg A$.[4]

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

4. RELATIONS BETWEEN MENTAL ATTITUDES

We think that rather than the interaction between intentions and goals or desires (as studied by Cohen & Levesque and Sadek), it is the interaction between intentions and beliefs which is crucial. Most importantly, an agent must abandon his intention to achieve A as soon as he believes that A is true [7, 15, 17]. This is expressed by:

$$Int_i A \rightarrow Bel_i \neg A \qquad \text{(Rel}_{\text{IntBel}}1)$$

Combined with the (D) axiom expressing consistency of belief it entails consistency of intentions, i.e. $Int_i A \rightarrow \neg Int_i \neg A$.

Theorem 1 $Bel_i A \rightarrow \neg Int_i A$.

PROOF. By (Rel$_{\text{IntBel}}1$) and the axiom (D) of the modal logic for belief, we have $Int_i A \rightarrow \neg Bel_i A$. From this the theorem follows by contraposition. $\square$

Theorem 2 $\neg Bel_i A \rightarrow \neg Int_i \neg Bel_i A$

PROOF. This can be shown from axiom (5) for belief together with theorem 1: an instance of that theorem is $Bel_i \neg Bel_i A \rightarrow \neg Int_i \neg Bel_i A$. Then axiom (5) being $\neg Bel_i A \rightarrow Bel_i \neg Bel_i A$, we thus obtain $\neg Bel_i A \rightarrow \neg Int_i \neg Bel_i A$. $\square$

---

[4] $After_\alpha$ 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 $\langle \alpha^{-1} \rangle$.

Axiom (Rel$_{\text{IntBel}}1$) has been criticized in the literature, because it describes intention as strongly related to belief. For example, if an agent does not know if 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 $\neg BelIf_i p \wedge Int_i p$ is a contradiction (by definition of $BelIf_i$ and by (Rel$_{\text{IntBel}}1$)). 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$_{\text{IntBel}}1$) to: $Int_i A \rightarrow \neg Bel_i A$ Then an agent could not know whether p is true, and at the same time he could intend that p (e.g. see [24]). But we must keep in mind that according to Sadek [17, p. 120], 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 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$_{\text{IntBel}}1$) 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 $Int_i A$ and $Int_i Bel_i A$ can be formally captured.

As far as we know, the only work addressing this problem is [17], where a new mental attitude *want* is proposed (also named *potential intention*). This mental attitude abbreviates $Bel_i A \vee Int_i Bel_i A$. It follows from $Bel_i A \rightarrow (Bel_i A \vee Int_i Bel_i A)$ that if an agent believes A, then he wants A.

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

$$Bel_i Bel_i A \rightarrow \neg Int_i Bel_i A$$

In others words, an agent cannot want to believe A, if he believes that he already believes A.

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

$$(Int_i Bel_i A \wedge Bel_i \neg A) \rightarrow Int_i A \qquad \text{(Rel}_{\text{IntBel}}2)$$

This axiom is read "if an agent believes A is false and intends to believe A then he will intend A". Suppose i intends to believe A, but actually he believes $\neg A$; 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 believe 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 $(Int_i Bel_i A \wedge \neg Int_i A) \rightarrow \neg BelIf_i A$. Hence if i intends to believe A without intending A, then he ignores whether A.

Finally, our third constitutive property of the rational balance between intention and belief is the following axiom (that is derived in [17]):

$$Int_i A \rightarrow Int_i Bel_i A \qquad \text{(Rel}_{\text{IntBel}}3)$$

This means that an agent cannot intend A without intending to believe A. The converse should not be valid: we can intend to believe A without intending A. Suppose e.g. you ignore whether the light is off or not, and you intend to believe that it is off. In this case you are prepared to act in order to acquire that belief, 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{IntBel}}$1) and (Rel$_{\text{IntBel}}$3) that (Rel$_{\text{IntBel}}$2) is an equivalence: $Int_i A \leftrightarrow Int_i Bel_i A \wedge Bel_i \neg A$.

Finally, there are two essential properties related to the agent's introspection capacity (cf. [17]):

$$Bel_i Int_i A \leftrightarrow Int_i A \qquad \text{(Rel}_{\text{IntBel}}4)$$
$$Bel_i \neg Int_i A \leftrightarrow \neg Int_i A \qquad \text{(Rel}_{\text{IntBel}}5)$$

These two axioms mean that intentions (respectively, non-intentions) of an agent are sound and complete with respect to his believed intentions (respectively, non-intentions).

Remark 1 *We have neglected here a property of intention of [7], viz. that an agent i cannot entertain the intention that A if he believes that A is always false. We have omitted this here in order not to introduce into our logic a temporal operator.*

5. 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 [19] 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 [6]. 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.

In [9, 10] it is shown that non-normal modal 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 [3] 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. The termination issue still remains to be addressed (and with it decidability).

6. COOPERATION PRINCIPLES

Generally speaking, to be cooperative w.r.t. an agent j means to contribute to the satisfaction of j's goals. While being a quite popular definition nowadays, it is nevertheless superficial. Ideally, the contribution should be balanced against a lot of other aspects, such as social rules and the cognitive capacities of the agent one is supposed to help. For example, to listen to j without interrupting him is a rule of social cooperation (one thus helps j to *better* satisfy his goals, viz. to speak), while not to answer more to his question than he can memorize is a rule of cognitive cooperation (related to the Gricean maxim of quantity [12]).

To be cooperative w.r.t. j also means to try to understand and satisfy j's ultimate goals (cf. [17] for that aspect). As we have shown in [5] accepting non-literal utterances of j can be seen as a form of cooperation, as well as adopting beliefs and intentions of j, and generating intentions with the aim of (indirectly) allowing j to satisfy his intentions.

Having in mind our aim of defining a minimal logic of a rational agent, we do not take into account here cognitive capacities (such as limited reasoning and introspection) and social rules. (Our axioms are nevertheless a priori consistent with a more refined analysis.) We thus restrict ourselves to two principles: *belief adoption* (an agent adopts the beliefs of another agent); *intention generation* (an agent generates intentions, in particular in order to aswer to the questions that have been put to him, and to correct erroneous beliefs of other agents).

If one wants to completely describe the mental state of an agent after such a belief adoption and intention generation process, one has to supplement these principles by principles of belief preservation, as studied in cognitive robotics [20, 28]. We have studied such principles before [13, 11], and do not go into the details here.

6.1 Belief adoption

When an agent i adopts a belief of another agent j he starts to believe himself what he believes j believes. Adoption must be constrained in some way in order to avoid to take over just everything another agent has uttered. We here formulate that condition in terms of *competence*: an agent adopts j's belief if he believes that j is competent at that belief. This notion has also been used by Cohen & Levesque [7] and Sadek [17]. Formally, in order to describe the competence of an agent at a formula we use a *relation of dependence* [13]: $i \overset{c}{\leadsto} p$ means that i is competent at p.

This enables us to formulate the following axiom:

$$\begin{array}{c} Bel_i A \to A \\ \text{if } i \overset{c}{\leadsto} A \text{ and } A \text{ is objective} \end{array} \qquad \text{(Adopt}_{\text{Bel}}1)$$

(Remember a formula is objective if it contains no modal operator Bel or Int.) Note that $\overset{c}{\leadsto}$ is a metalinguistic notion.

Remark 2 *As this is a logical axiom, and as our logic of belief is the logic modal KD45, by the inference rule of necessitation we obtain from (Adopt$_{\text{Bel}}$1) that an agent's competence is mutual knowledge: axiom entails that if $i \overset{c}{\leadsto} A$ then $Bel_{k_1} \ldots Bel_{k_n}(Bel_i A \to A)$ for every $\{k_1, \ldots, k_n\} \subseteq \mathcal{AGT}$.*

Another principle of belief adoption that supplements the above is the following: if j asserts A and A does not contradict k's beliefs, then k adopts A:

$$\neg Bel_k \neg A \to After_{\langle i,j,A \rangle} Bel_k A \qquad \text{(Adopt}_{\text{Bel}}2)$$

(Remember that k observes i's act because we have supposed that actions are public.) Note that k adopts A even if i is not competent at A. A similar principle has been proposed in [26].

6.2 Intention generation

Intention generation completes our principles of cooperation. Suppose $Bel_i Int_j A$. Intuitively, the difficulty is to take

into account the preceding axiom (Rel$_\text{IntBel}$1) in an appropriate way: i should only generate the intention to bring about A when i believes that A is currently false. We formalize this in the sequel.

When i doesn't believe that A is currently true ($\neg Bel_i A$) then i does not necessarily entertain the intention that A be true. Indeed, if moreover $\neg Bel_i \neg A$ then i doesn't know whether A is currently true or not, and it cannot be the case that $Int_i A$ because $\neg Bel_i \neg A$ implies $\neg Int_i A$. The only thing that can be guaranteed here is that i adopts *the intention to believe* that A (cf. our discussion about (Rel$_\text{IntBel}$1) in Sect. 4). If we rewrite this we obtain our central axiom:

$$(Bel_i Int_j A \wedge \neg Bel_i A \wedge$$
$$\neg Int_i Bel_i \neg A) \rightarrow Int_i Bel_i A \qquad \text{(Gen}_\text{Int}\text{1)}$$

Remark 3 *As we have said, we did not include in our axioms that an agent i cannot entertain the intention that A if he believes that A is always false. This means that i must abandon his intention $Int_i A$ as soon as i starts to believe that A will always be false, and the other way round $Int_i A$ cannot be generated if i believes that A will always be false. We can constrain axiom (Gen$_\text{Int}$1) in order to guarantee this.*

Remark 4 *(Gen$_\text{Int}$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: $Bel_i Int_j A \wedge Bel_i Int_k \neg A$. Suppose moreover that $\neg BelIf_i A \wedge \neg Int_i \neg A \wedge \neg Int_i A$ (i.e. i doesn't bother at all about A). Then by (Gen$_\text{Int}$1) i generates the intentions $Int_i Bel_i A$ and $Int_i Bel_i \neg A$. But this is inconsistent according to (Rel$_\text{IntBel}$1).*

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

$$(Bel_i Int_j A \wedge \neg Bel_i A \wedge C) \rightarrow Int_i Bel_i A$$

where C is a formula of the form $\neg Bel_i Int_{k_1} \neg A \wedge \ldots \wedge \neg Bel_i Int_{k_n} \neg A$. 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$_\text{Int}$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

$$Bel_i Int_j A \rightarrow (Int_i A \vee Int_i \neg A)$$

6.3 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.

First of all, note that by theorem 2 the second premiss $\neg Bel_i A$ of our central axiom (Gen$_\text{Int}$1) ensures that i will not generate the intention to believe A if i already has a contradictory intention.

In accordance with (Rel$_\text{IntBel}$2) and (Gen$_\text{Int}$1), if an agent i believes 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 1, if agent i believes that A is false then he cannot have the intention that A be false. Putting this together we obtain:

Theorem 3 *Axiom (Gen$_\text{Int}$1) implies*

$$(Bel_i Int_j A \wedge Bel_i \neg A) \rightarrow Int_i A \qquad \text{(Gen}_\text{Int}\text{2)}$$

Hence our first principle (Gen$_\text{Int}$2) says that if i believes that the world must necessarily change (in the aim to satisfy the j's intention), then j's intention is directly adopted.

PROOF. The hypothesis is $Bel_i Int_j A \wedge Bel_i \neg A$. On the one hand, $Bel_i \neg A \rightarrow \neg Bel_i A$ with axiom (D), and we thus obtain the second hypothesis of (Gen$_\text{Int}$1).
To establish the third hypothesis of (Gen$_\text{Int}$1) we proceed as follows: first, we derive $Int_i Bel_i \neg A \rightarrow Bel_i Bel_i \neg A$ with (Rel$_\text{IntBel}$1). Then $Bel_i Bel_i \neg A \rightarrow \neg Bel_i \neg A$ with axiom (5). We thus obtain $Int_i Bel_i \neg A \rightarrow \neg Bel_i \neg A$, and by contraposition $Bel_i \neg A \rightarrow \neg Int_i Bel_i \neg A$.
In consequence the hypotheses of axiom (Gen$_\text{Int}$2) imply those of axiom (Gen$_\text{Int}$1). The latter allows us to obtain $Int_i Bel_i A$:

$$(Bel_i Int_j A \wedge Bel_i \neg A) \rightarrow Int_i Bel_i A$$

Now $(Int_i Bel_i A \wedge Bel_i \neg A) \rightarrow Int_i A$ with (Rel$_\text{IntBel}$2), entailing (Gen$_\text{Int}$2). □

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

Theorem 4 *Axiom (Gen$_\text{Int}$1) implies*

$$Bel_i Int_j A \rightarrow Int_i Bel_j A \qquad \text{(Gen}_\text{Int}\text{3)}$$

PROOF. (Gen$_\text{Int}$1) allows to derive (Gen$_\text{Int}$2), of which $(Bel_i Int_j Bel_j A \wedge Bel_i \neg Bel_j A) \rightarrow Int_i Bel_j A$ are an instance. As $Int_j Bel_j A$ implies $Bel_j \neg Bel_j A$ by (Rel$_\text{IntBel}$1), we have $Bel_i Int_j Bel_j A \rightarrow Bel_i Bel_j \neg Bel_j A$ by the principles of modal logic K. As $Bel_j \neg Bel_j A$ is equivalent to $\neg Bel_j A$ in KD45, we obtain $Bel_i Int_j Bel_j A \rightarrow Bel_i \neg Bel_j A$.
Thus we obtain $Bel_i Int_j Bel_j A \rightarrow Int_i Bel_j A$ from (Gen$_\text{Int}$2).
On the other hand, as $Int_j A$ implies $Int_j Bel_j A$ by (Rel$_\text{IntBel}$3), we have $Bel_i Int_j A \rightarrow Bel_i Int_j Bel_j A$ by the principles of modal logic K.
Finally transitivity of $\rightarrow$ allows us to conclude that $Bel_i Int_j A \rightarrow Int_i Bel_j A$. □

Remark 5 *Conditions $Bel_i Int_j A$ and $Bel_i A$ of (Gen$_\text{Int}$3) cannot be simultaneously true if j is competent at A. Indeed, $Bel_i Int_j A \rightarrow Bel_i Bel_j \neg A$ by (Rel$_\text{IntBel}$1), and if j was competent at A then we would have $Bel_i Bel_j \neg A \rightarrow Bel_i \neg A$, which cannot be the case because $Bel_i A$ and $Bel_i \neg A$ are inconsistent.*

To sum it up, our central axiom allows us to derive natural and powerful principles of cooperation. In the next section we shall show that they can be applied successfully to derive yes-no questions and requests from assertions.

7. INFERRING NON-ASSERTIVE ACTS

In this section, we illustrate by two examples how the effects of yes-no questions and request can be obtained from assertive speech acts via our cooperative principles.

7.1 The effects of assertion speech acts

In [18], Sadek describes three types of effects, that we present here in a slightly simplified version.

- The *rational effect* corresponds to the effect of the act on the addressee as expected by the speaker.

- The *intentional effect* is the speaker's intention to produce the rational effect on the addressee (this effect is related to the gricean point of view of communication).

- The *indirect effect*[5] corresponds to persistence (through the performance of the act) of the feasibility preconditions.

The *rational effect* is not directly produced by the addressee, but obtains only if the speech act's satisfaction conditions hold (see [21]). For example, if $\alpha = \langle i, j, p \rangle$ has just been performed, its rational effect is $Bel_j p$.

Note that in [18] the formalization of effects is more fined-grained than here. We aimed at good compromise between soundness of representation and its complexity, and we therefore simplified his model).

The *intentional effect* is based on the rational effect. It describes that the speaker wants the hearer to believe that the speaker intends to produce an effect. Thus, the intentional effect resulting from the performance of $\alpha = \langle i, j, p \rangle$ is: $Int_i Bel_j Int_i Bel_j p$.

Finally, the *indirect effect* is related to the preservation of the capacity precondition and of the relevance precondition. For example, the capacity precondition of $\alpha = \langle i, j, p \rangle$ is $Bel_i p$; the relevance precondition of α is $\neg Bel_i BelIf_j p$.[6]

We suppose here that action laws are part of the common beliefs. Thus, any observer of the performance of a speech act believes that the indirect effect and the intentional effect have occurred: if k observes the performance of α, then $Bel_k Int_i Bel_j \ Int_i Bel_j p \wedge Bel_k \ Bel_i p \wedge Bel_k \neg Bel_i BelIf_j p$ holds. If the addressee j (who is a particular observer of the act) comes to believe p (by inferring it from the effects of the act, his –other– beliefs and his rationality and cooperation rules), then the rational effect obtains, too. In this case the speech act is said to be satisfied.

7.2 The case of yes-no questions

Let $\alpha = \langle u, s, Int_u BelIf_u A \rangle$ be the speech act that has just been performed. Suppose that α corresponds to the utterance "u says he wants to know if A holds". As we have shown previously (Sect. 2), this speech act represents a form of indirection that can be interpreted as a yes-no question. The effects of α on the agent s are as follows (see Sect. 7.1):

1. $Bel_s Int_u Bel_s Int_u Bel_s Int_u BelIf_u A$

2. $Bel_s Bel_u Int_u BelIf_u A$

3. $Bel_s \neg Bel_u BelIf_s Int_u BelIf_u A$

These effects respectively correspond to the rational (1) and to the indirect effect (sincerity (2) and relevance (3)).

---

[5] "Indirect" must be understood as "the side effect of the act", and not as the effect of an indirect speech act.

[6] Sadek describes communicative acts that are not necessarily speech acts. As we treat here only speech acts, the capacity precondition can be viewed as the sincerity precondition.

If we suppose every agent is competent at his mental attitudes (as it is manifested by the axioms (Rel_IntBel4) and (Rel_IntBel5)), then

$$Bel_s Int_u BelIf_u A$$

is a consequence of (2) via (Rel_IntBel4). The principle (Rel_IntBel1) (with the standard principles of the logic KD45 for belief) entails

$$Bel_s Bel_u \neg BelIf_u A$$

By the same principles, $Bel_u \neg BelIf_u A$ is equivalent to $\neg BelIf_u A$, and then we get

$$Bel_s \neg BelIf_u A$$

Finally, giving $Bel_s Int_u BelIf_u A$ and $Bel_s \neg BelIf_u A$, (Gen_Int2) allows us to conclude $Int_s BelIf_u A$. Thus, the agent s satisfies the initial intention of the agent u, which was that s adopts the intention that u knows if A holds.

7.3 The case of requests

Let $\alpha = \langle u, s, Int_u Done_\beta \top \rangle$ be the speech act that has just been performed, where s is the author of β. Suppose it corresponds to the utterance "I want you to perform β". This speech act is an indirection that can be interpreted as a request (see Sect. 2). The effects of α on s are as follows (see Sect. 7.1):

1. $Bel_s Int_u Bel_s Int_u Bel_s Int_u Done_\beta \top$

2. $Bel_s Bel_u Int_u Done_\beta \top$

3. $Bel_s \neg Bel_u BelIf_s Int_u Done_\beta \top$

These effects respectively correspond to the intentional effect (1) and to the indirect effect ((2) and (3)).

If we suppose that every agent is competent at his own mental attitudes, then

$$Bel_s Int_u Done_\beta \top$$

is a logical consequence of (2). If we suppose that s believes that he has not just performed β, i.e.

$$Bel_s \neg Done_\beta \top$$

then he will intend to perform β (via (Gen_Int2)). To sum it up, s satisfies the initial intention of u, which was that s performs β.

Remark 6 *If β is an action that must be performed by u (and not by s), the corresponding utterance of u would be of the form "I want to perform β". We might interpret this utterance as an indirect speech act. Then u would ask s (in the allusive mode) to perform the action in his place. This would require a principle of the type "if an agent i believes that an agent j intends to perform some action, then the agent i will intend to perform this action". Thus, we can always add axioms in order to take into account more fined-grained language phenomena.*

Remark 7 *In our example, we have supposed that the agent s is aware that he had not already performed β. If we suppose now that s believes he has already performed β, he will intend*

that the agent u be aware of that (via (Gen$_{Int}$3).[7]) According to the reaction of u ("I did not [hear | understand | remember | ...]"), the agent s may perform β again (in this case, a new intention should be generated, because the first one has already been satisfied).

Finally, we could suppose s does not remember if he has already performed β. The intention generated by (Gen$_{Int}$1) should then be related to a research of s in his memory, with the aim of knowing if he has already performed β. According to the answer, he will generate an intention either via (Gen$_{Int}$2) or via (Gen$_{Int}$3).

This last case formally shows the point of view developed in Sect. 4 on the problem of switching the light in a room where we do not know if the light is on or off. In this sense, this example illustrates that the axiom (Rel$_{IntBel}$1) does not rely intention and belief in a too strong way.

8. DISCUSSION AND CONCLUSION

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.

The only type of speech acts in our logic are assertions. We have shown how requests and yes-no questions can be inferred in this framework from particular assertions, in a way similar to the inference of indirect speech acts. Inference is via cooperation principles, the most important of which are original. We have thus shown that our minimal logic allows nevertheless to reason about communication in a cooperative environment.

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. 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 [26, 25] lead to so-called fanatic agents, who never abandon their goals (even when they learn that they became true). In [27] every goal A comes with a cancelling 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 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.

9. REFERENCES

[7]Because of our semantics of the $Done_\beta$ operator, $Bel_s Done_\beta \top$ will not hold, the last act that has been performed being α (the assertive speech act of u).

[1] J. L. Austin. *How To Do Things With Words.* Oxford University Press, 1962.

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

[3] M. A. Castilho, L. Fariñas del Cerro, O. Gasquet, and A. Herzig. Modal tableaux with propagation rules and structural rules. *Fundamenta Informaticae,* 32(3/4):281–297, 1997.

[4] M. Champagne, R. Faure, A. Herzig, D. Longin, C. Luc, J.-L. Nespoulous, and J. Virbel. Formalisation logique de la communication non littérale à la lumière d'aperçus pragmatiques et neuropsycholinguistiques. In B. Chaib-draa and P. Enjalbert, editors, *Proc. Journées Francophones Modèles Formels de l'Interaction (MFI'01)*, volume 1, pages 31–47, 2001. 17 pages.

[5] M. Champagne and D. Longin. Non literal communication: From pragmatical to logical and psycholinguistical aspects. In *Int. Colloc. on Cognitive Sciences (ICCS'01)*, 2001. 23 pages.

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

[7] P. R. Cohen and H. J. Levesque. Intention is choice with commitment. *Artificial Intelligence Journal,* 42(2–3), 1990.

[8] P. R. Cohen and H. J. Levesque. Rational interaction as the basis for communication. In P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication.* MIT Press, 1990.

[9] L. Fariñas del Cerro and A. Herzig. Modal deduction with applications in epistemic and temporal logic. In D. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic and Artificial Intelligence*, volume 4 - Epistemic and Temporal Reasoning, pages 499–594. Oxford University Press, 1995.

[10] O. Gasquet and A. Herzig. From classical to normal modal logics. In H. Wansing, editor, *Proof Theory of Modal Logics*, number 2 in Applied Logic Series. Kluwer Academic Publishers, 1996.

[11] O. Gasquet, A. Herzig, and D. Longin. An analysis of communication in a logic of belief, intention and action. Technical Report IRIT/2001-07-R, Institut de Recherche en Informatique de Toulouse, Mar. 2001. 22 pages. Available in www.irit.fr/ACTIVITES/LILaC/.

[12] H. P. Grice. Logic and conversation. In J. Cole and J. Morgan, editors, *Syntaxe and Semantics: Speech acts*, volume 3 : *Speech Acts.* Academic Press, 1975.

[13] A. Herzig and D. Longin. Belief dynamics in cooperative dialogues. *Journal of Semantics,* 17(2), 2000. 20 pages.

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

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

[16] R. Reiter. The frame problem in the Situation

Calculus: A simple solution (sometimes) and a completeness result for goal regression. In V. Lifschitz, editor, *Artificial Intelligence and Mathematical Theory of Computation: Papers in Honor of John McCarthy*, pages 359–380. Academic Press, San Diego, CA, 1991.

[17] M. D. Sadek. A study in the logic of intention. In B. Nebel, C. Rich, and W. Swartout, editors, *Proc. Third Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 462–473. Morgan Kaufmann Publishers, 1992.

[18] M. D. Sadek. Dialogue acts are rational plans. In M. Taylor, F. Nel, and D. Bouwhuis, editors, *The structure of mutimodal dialogue*, pages 167–188, Philadelphia/Amsterdam, 2000. John Benjamins publishing company. From ESCA/ETRW, Workshop on The Structure of Multimodal Dialogue (Venaco II), 1991.

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

[20] R. Scherl and H. J. Levesque. The frame problem and knowledge producing actions. In *Proc. Nat. Conf. on AI (AAAI'93)*, pages 689–695. AAAI Press, 1993.

[21] J. R. Searle. *Speech acts: An essay in the philosophy of language.* Cambridge University Press, 1969.

[22] J. R. Searle. *Expression and Meaning.* Cambridge University Press, 1979.

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

[24] C. Seguin. *De l'action l'intention : vers une caractrisation formelle des agents.* PhD Thesis, Université Paul Sabatier, Toulouse, France, Mar. 1992.

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

[26] S. Shapiro, Y. Lespérance, and H. J. Levesque. 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, 1997.

[27] S. Shapiro, Y. Lespérance, and H. J. Levesque. Specifying communicative multi-agent systems. In W. Wobcke, M. Pagnucco, and C. Zhang, editors, *Agents and Multi-Agent Systems - Formalisms, Methodologies, and Applications*, pages 1–14. Springer-Verlag, LNAI 1441, 1998.

[28] 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)*, pages 527–538, 2000.

[29] M. Thielscher. Representing the knowledge of a robot. In A. Cohn, F. Giunchiglia, and B. Selman, editors, *Proc. KR'00*, pages 109–120. Morgan Kaufmann, 2000.

[30] D. Vanderveken. Formal pragmatics of non literal meaning. *Linguistische Berichte*, 1997.