Beliefs, Intentions, Speech Acts and Topics

Article · April 2000

Source: CiteSeer

Dominique Longin
Institut de Recherche en Informatique de To…

Beliefs, intentions, speech acts and topics*

Andreas Herzig and Dominique Longin†

Institut de Recherche en Informatique de Toulouse (IRIT)
Université Paul Sabatier,
118 Route de Narbonne, F-31062 Toulouse Cedex 4
`mailto: {herzig,longin}@irit.fr`
`http://www.irit.fr/ACTIVITES/EQ_ALG/`

Rapport IRIT 00-08-R

Key words: representation (belief, intention, action), reasonning (deduction), dynamic epistemic logic, topics.

*This work has been supported by France Telecom - CNET, pôle 'Interactions Intelligentes et Dialogue'.

†Thanks to O. Gasquet, J. Lang and Th. Polacsek for useful discussions.

Abstract

We study the dynamics of belief in cooperative task-oriented man-machine dialogues. We suppose that during the dialogue each participant can mistake a piece of information, forget it or simply change his mind.

We begin with some hypotheses about the participants, and present some approaches. We point out shortcomings of the latter and introduce a new logic of speech acts, beliefs and intentions, where intentions have a non-normal modal logic. We focus on the interactions between these modal operators. Our basic notion is that of a topic: we suppose that we can associate a set of topics to every agent, speech act and formula. This allows to express an agent's competence, belief adoption and preservation.

1 Introduction

Cooperative task-oriented man-machine dialogue, both in natural and artificial languages, is one of the most important challenges for computer science. Participants in such dialogues have one major common goal, viz. to achieve the task under concern. Each of the participants has some information contributing to the achievement of some goals, but none of them can achieve it alone. Cooperativity is a fundamental and useful hypothesis. Informally, an agent is cooperative w.r.t. another one if the former helps the latter to achieve his goals (cf. Grice's cooperation principles, as well as his conversation maxims [24]).

We suppose that each participant is *sincere*. This means that his utterances faithfully mirror his beliefs: e.g. if a participant says 'the sky is blue' then he indeed believes that the sky is blue. Such a hypothesis entails that contradictions between the presuppositions of an utterance and the hearer's beliefs about the speaker cannot be explained in terms of lies.

We finally suppose that utterances are *public*: all the agents perceive the performance of an utterance (but might misinterpret them).

The background of our work is an effective generic real-time cooperative dialogue system which has been specified and developped by the France Telecom R&D Center CNET. This approach consists in first describing the system's behaviour within a logical *theory of rational interaction* [42, 43, 44], and second implementing this theory within an *inference system* called ARTIMIS [40, 46, 41]. For a fixed set of domains, this system is able to accept nearly unconstrained spontaneous language as input, and react in a cooperative way. The activities of the dialogue system are twofold: to consume the speaker's utterances, and to generate appropriate reactions. The reactive part is completely defined in the current state of both the theory and the implementation. On the other hand, the comsumption of an utterance is handled only partially, in particular the belief change part of the consumption process. The difficulties are illustrated in our running example. There are only two agents, the system s and the user u.

$\boldsymbol{s_1}$: *Hello. What do you want?*
$\boldsymbol{u_1}$: *A first class train ticket to Paris, please.*
$\boldsymbol{s_2}$: *150 \$, please.*
$\boldsymbol{u_2}$: *Ahem ... A second-class train ticket, please.*
$\boldsymbol{s_3}$: *100 \$, please.*
$\boldsymbol{u_3}$: *Can I pay the 80 \$ by credit card?*
$\boldsymbol{s_4}$: *The price isn't 80 \$ but 100 \$. Yes, you can pay by credit card.*
$\boldsymbol{u_4}$: ...

Our example illustrates that in a dialogue agents might change their mind, make mistakes, and misinterpret. Such phenomena must be taken into account when modelling the evolution of mental states.

Within such dialogues we study the evolution of the system's beliefs. The system must be able to: (1) *accept* some type of information (e.g. destination and class information in u_1); (2) *deduce* supplementary information not directly contained in the utterance using laws about the world (cf. deduction of the price in s_2, given u_1); (3) *preserve* some information it believed before the utterance (viz. the destination in u_2); (4) sometimes *accept* information contradicting its own beliefs (in particular when the user changes his mind, as in u_2); (5) sometimes *refuse* to take over some information, in particular if the user informs about facts he isn't competent at (cf. u_3 and s_4).

We aim at a semantics having both a complete axiomatization and proof procedure, and an effective implementation. This has motivated several choices, in particular a Sahlqvist-type possible worlds semantics (for which general completeness results exist), and a notion of intention that is primitive (contrarily to the complex constructions in the literature), having a non-normal modal logic (which nevertheless can be reduced to the Sahlqvist framework).

In the next section we introduce our multimodal framework (Sect. 2). It is based on a metalinguistic theory of topics (Sect. 3). Topics permit to integrate appropriate definitions of belief adoption and preservation into our logical theory, which is presented both axiomatically and semantically in Sect. 4 and Sect. 5.

Within this framework one can formulate nonlogical laws (in particular domain laws, laws governing speech acts, and reactive laws). For reasons of space we do not present this nonlogical theory. It can be found in [28], which contains also a less elaborate logical theory. We refine it here by adding a semantical analysis, and introducing a more appropriate notion of intention.

2 Rational interaction and belief change

The most prominent frameworks analyse interactions between rational agents in terms of their mental states. The mental state of an agent is made up of different mental attitudes (belief, intention, ...). This is the basis of so-called *BDI-architectures*. Therein, belief change must be situated within a formal *rational balance theory* and a formal *rational interaction theory* [44, 9, 11]. Utterances are represented by speech acts [3, 50, 49].

Cohen&Levesque [9, 11] have laid the bases of such a general theory of rational interaction. Their theory allows the agent to either reject the input (if the speaker is believed to be unsincere), or change his beliefs and adopt it. But in the latter case, the theory suffers from the well-known frame problem [34]. Hence in our example the system does not preserve the destination through u_2 (and also through u_3).

Perrault [35] tries to solve the frame problem in a framework based on Reiter's default logic [39]. His theory suffers from some problems pointed out in [2]. In particular, Perrault's agents never question old beliefs, and can only expand their mental state (in the sense of AGM revision). Hence they cannot take into account u_2, where the user switches from a first-class ticket to a second-class ticket.

Appelt&Konolige [2] advocate the use of hierarchic autoepistemic logic (HAEL). Basically, what one gains there is that defaults can be stratified. This can be used to fine-tune default application and thus avoid unwanted extensions. Their theory can represent our example correctly. Nevertheless, apart from the relatively complex HAEL technology, it seems that Appelt and Konolige's belief adoption criterion encounters problems similar to Perrault's 'self-convincing' utterances they had highlighted: indeed, suppose the hearer has no opinion about p. If the speaker informs the hearer that p then under otherwise favourable circumstances then the hearer adopts p. But if the speaker informs the hearer that *the hearer believes* p (or that he believes the hearer believes p) then it is clearly at odds with our intuitions that the hearer should accept such an assertion about his mental state. The only way to avoid that seems to be to shift the hearer's ignorance about p to the prioritary level 0 of the HAEL hierarchy. But in this case the acceptance of the assertion p would be blocked as well, which is counterintuitive.

In [42, 45], Sadek introduces the notion of *belief reconstruction*. It is defined in terms of four axioms (memory, admission, consumption and persistence). In [43] he introduces original formal characterisations of speech acts in the spirit of Searle [50]. Sadek distinguishes several effects: the *indirect effect* of a speech act is the preservation of its preconditions; its *intentional effect* corresponds to the gricean point of view of communication [25], and its *rational effect* is the expected effect of the act. Although his axioms can handle our example, due to their autoepistemic flavour they do not give us a constructive formal definition of deduction.

Rao&Georgeff [36, 37] have proposed theories and architectures for rational agents. Recently, they have defined a tableau proof procedure for their logic [38]. In a manner similar to STRIPS, actions and plans are represented by their preconditions together with add- and delete-lists. The latter lists are restricted to sets of atomic formulas. In such a framework, one can *a priori* neither represent nondeterministic actions nor actions with indirect effects (obtained through integrity constraints). Hence one cannot use laws about the world to e.g. derive the price if the user informs about his destination and class, as needed in s_2. Even more importantly, actions can only have effects that are factual. This excludes the handling of speech acts, whose effects are typically represented by means of nested intensional operators such as intentions to bring about belief.

There exist important formal analyses of belief change outside the theories of rational interaction that we have reviewed up to now. Most prominent are AGM belief revision and KM update operations [1, 30]. Which of these different operations could apply to our example dialogue? The official explanation [30, 23] says that when the hearer realizes that the speaker has changed his mind then he should *update* his beliefs, and when the hearer realizes that he had misunderstood an information from the speaker he should *revise* his mental state.

It appears that in many dialogue examples the system lacks information to decide whether the incoming information corresponds to real world change or not.[1] In our example, after u_2 the system might be unable to distinguish between the case where the user has changed his mind and the case where he has misunderstood utterance u_1. (See [29] for a detailed critique of the KM framework.)

Moreover, both revision and update have several common properties that are not suitable in dialogues. In particular, the over-informing nature of some information is neglected, expressed by the postulate $(K \circ A) \leftrightarrow K$ if $K \rightarrow A$. Finally, no distinction is made between different levels of belief (factual, introspective, alternating, etc.).

In our approach, building on previous work in [15, 28], we implement belief change by topic-based belief adoption and preservation. But first of all we introduce the language.

The modal language. As most of the authors, we work in a multi-modal framework, with modal operators for belief, mutual belief, intention and action. Our language is that of first order multi-modal logic without equality and without function symbols. Let AGT be the set of agents. For $i \in AGT$, $Bel_i A$ is read 'the agent i believes that A', and $Intend_i A$ is read 'the agent i intends that A'.

Speech acts [3, 50] are represented by tuples of the form $\langle \mathsf{FORCE}_{i,j}\ A\rangle$ where FORCE is the illocutionary force of the act, $i, j \in AGT$, and A is the propositional content of the act. For example $\langle \mathsf{Inform}_{u,s}\, Dest(\text{Paris})\rangle$ represents a declarative utterance of the user informing the system that the destination of his ticket is Paris. In Sect. 4 we shall add acts that are contextualized by agents.

Let ACT be the set of all speech acts. With every $\alpha \in ACT$ we associate a modal operator $Done_\alpha$. $Done_\alpha A$ is read 'α has just been performed, before which A was true'.[2] $Done_\alpha \top$ is read 'α has just been performed'.

For example, $Bel_s(Dest(\text{Paris}) \wedge Class(\text{1st}) \rightarrow Price(150\ \$))$ is a formula. (It is also a domain law, allowing the system to deduce the price from information about destination and class.) Other meaningful examples are the formulas $Done_{\langle \mathsf{Inform}_{u,s}\ p\rangle} Bel_u p$ and $Done_{\langle \mathsf{Inform}_{u,s}\ p\rangle} Bel_u(\neg Bel_s p \wedge \neg Bel_s \neg p)$. They express the sincerity of u, and are part of the nonlogical theory. Atomic formulas are denoted by $p, q, \ldots$, and Atm is the set of all atomic formulas. Formulas are denoted by $A, B, \ldots$.

[1] We consider that the real world includes everything that is external to the agent (including the speaker's intentions).

[2] $Done_\alpha A$ is just as $\langle \alpha^{-1}\rangle A$ of dynamic logic [27].

3 Topic-based belief change and adoption

The competence of agents and the influence of speech acts. In our approach, unlike to Sadek's we always accept[3] speech acts, but we proceed in two steps: the hearer accepts the indirect and intentional effects, but only adopts the speaker's beliefs if he believes the speaker to be competent at these beliefs. Thus, *the speaker's competence* is our criterion to determine which part of the input is accepted by the hearer. For example, s accepts the input about the new class (after u_2) but rejects the input about the price (after u_3), the reason being that he considers u to be competent at classes but not at ticket prices.

Which beliefs of the hearer can be preserved after the performance of a speech act? Our key concept here is that of the *influence of a speech act on beliefs*. If there exists a relation of influence between the speech act and a belief, this belief cannot be preserved in the new mental state. In our example, the old transport class cannot be preserved through u_2, because the act of informing about classes influences the hearer's beliefs about classes. On the other hand, the destination is not influenced by u_2, and can thus be preserved.

The concept of influence (or dependence) of an act 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 [48] occlusion, Thielscher's [51] influence relation, or Giunchiglia *et al.*'s [21] possibly changes operators. In particular, such a concept has been used by one of the authors in [6].

All this presupposes that we are able to determine the competence of an agent and the influence of a speech act. Contrarily to the above approaches in reasoning about actions, the foundation for both notions will be provided here by the concept of *topic*: we start from the idea that with every agent, speech act, and formula, some set of topics can be associated. Thus, an agent is competent about a formula A if and only if the set of topics associated with A is a subset of the set of topics associated with this agent – the set of topics the agent is competent in. And a formula A is preserved after the performance of a speech act α if there is no topic that occurs both in the set of topics associated to A and in the set of topics associated to α.

Topics are a natural and intuitively appealing concept which will allows us to fine-tune the consumption of speech acts. This notion is important in linguistics [4, 20, 52, 53]. From a logical point of view, Epstein [13, p. 68] defines the *subject matter of a proposition A*. He shows that we can then define two propositions as being related if they have some subject matter in common. Our $\mathfrak{subject}$ function can be seen as an extension of this function to a multi-modal language.[4] In the rest of the section we give our metalinguistic theory of topics.

[3]'Accepting' an act means that we admit that it has been performed.

[4]Other studies of the notion of topic exist in the philosophical logic literature, in particular [32, 22, 12].

Themes and topics. A *theme* is what something is about. We suppose that the set of themes is nonempty. In our running example, the themes are destinations, classes, and prices.

For $i \in AGT$, ma_i is called an *atomic context*. ma_i stands for 'the mental attitude of agent i'. A *context* is a possibly empty sequence of atomic contexts, noted $ma_{i_1}\mathbf{:}ma_{i_2}\mathbf{:}\ldots ma_{i_n}$. The empty context is noted λ. A theme t together with a context c makes up a topic of information, denoted by $c\mathbf{:}t$. For example, $ma_u\mathbf{:}price$ is a topic consisting in the user's mental attitude at prices, and $ma_s\mathbf{:}ma_u\mathbf{:}price$ is a topic consisting in the system's mental attitude at the user's mental attitude at prices.

By convention, $\lambda\mathbf{:}c = c\mathbf{:}\lambda = c$, and $\lambda\mathbf{:}t = t$. Furthermore, principles of introspection motivate the identity $ma_i\mathbf{:}ma_i = ma_i$.

Given a set of themes and a set of agents we note $\mathbb{T}$ the associated set of topics. $\mathbb{T}_n$ is the set of topics whose contexts have length at most n. Hence $\mathbb{T}_0$ is the set of themes. In this paper, for technical reasons (and also reasons of representational economy) we shall suppose that the length of each context is at most 2. Hence we restrict $\mathbb{T}$ to $\mathbb{T}_2$.

The subject of a formula. We call *subject of a formula* a set of topics the formula is about. This notion is formalized by a function $\mathfrak{subject}(A) \subseteq \mathbb{T}$. In our running example $\mathfrak{subject}(Class(\text{1st})) = \{class\}$, $\mathfrak{subject}(Dest(\text{Paris})) = \{dest\}$, and $\mathfrak{subject}(Bel_s Bel_u Price(80\ \$) \wedge Bel_s Price(100\ \$)) = \{ma_s\mathbf{:}ma_u\mathbf{:}price, ma_s\mathbf{:}price\}$. We give the following axioms:

$$\mathfrak{subject}(p) \subseteq \mathbb{T}_0 \text{ and } \mathfrak{subject}(p) \neq \emptyset \text{ if } p \in Atm; \tag{1}$$
$$\mathfrak{subject}(\top) = \emptyset; \tag{2}$$
$$\mathfrak{subject}(\neg A) = \mathfrak{subject}(A); \tag{3}$$
$$\mathfrak{subject}(A \wedge B) = \mathfrak{subject}(A) \cup \mathfrak{subject}(B); \tag{4}$$
$$\mathfrak{subject}(Bel_i A) = \{ma_i\mathbf{:}c\mathbf{:}t \mid c\mathbf{:}t \in \mathfrak{subject}(A)\}; \tag{5}$$
$$\mathfrak{subject}(Intend_i A) = \mathfrak{subject}(Bel_i A); \tag{6}$$
$$\mathfrak{subject}(Done_{\langle \mathsf{FORCE}_{i,j}\ A'\rangle} A) = \mathfrak{subject}(A) \cup \mathfrak{subject}(A'); \tag{7}$$
$$\mathfrak{subject}(\forall x A) = \mathfrak{subject}(A); \tag{8}$$
$$\mathfrak{subject}(A[t/x]) \subseteq \mathfrak{subject}(A). \tag{9}$$

(2) says that the truth is about nothing. Note that in (5) c might be the empty context. (5) entails that $\mathfrak{subject}(Bel_i A) = \mathfrak{subject}(Bel_i \ldots Bel_i A)$. (6) expresses that atomic contexts abstract from the belief or intention character of mental attitudes. (7) entails e.g. that $\mathfrak{subject}(Done_{\langle \mathsf{Inform}_{u,s}\ Price(100\ \$)\rangle} Bel_s Price(150\ \$)) = \{price, ma_s\mathbf{:}price\}$. (8) and (9) express that open formulas are considered to be universally quantified. Finally, note that our subject function is not extensional: logically equivalent formulas may have different topics (cf. the discussion in Sect. 6).

It follows from our axioms that the subject of an arbitrary formula is completely determined by the subjects of its atomic formulas. This is representationally interesting. The same motivation has led us to restrict topics to $\mathbb{T}_2$. Due to that restriction we suppose that

$$ma_i\mathbf{:}ma_j\mathbf{:}c\mathbf{:}t = ma_i\mathbf{:}ma_j\mathbf{:}t$$

The corresponding subject function can be obtained by first reducing $\mathfrak{subject}(A)$ by the above equations $\lambda\mathbf{:}t = t$ etc., and then throwing out topics not in $\mathbb{T}_2$.

The competence of an agent. The user is competent at destinations and classes, but not at prices.[5] We call *competence of an agent* the set of topics i is competent at, i.e. $\mathfrak{competence}(i) \subseteq \mathbb{T}$. We assume here that $\mathfrak{competence}(i) \subseteq \mathbb{T}_0$, i.e. we do not consider competence of agents at his own or other agents' mental state.[6]

The scope of an act. In the formalization of speech acts the illocutionary force determines a set of formula schemes (the preconditions and the effects of the act) instantiated by the propositional content. The scope of a speech act determines which mental attitudes of an agent are questioned by this act. The performance of a speech act always influences some mental attitudes of the hearer. The scope of a speech act is a set of topics, i.e. $\mathfrak{scope}(\alpha) \subseteq \mathbb{T}$. As said above, we suppose in this paper $\mathfrak{scope}(\alpha) \subseteq \mathbb{T}_2$.

Roughly speaking, the themes of a speech act are determined by its propositional content, while the contexts are determined by its illocutionary force. Thus, contexts tell us which mental attitudes might change.

An act always influences the hearer's beliefs about the speaker's attitude towards the propositional content. Formally, for every illocutionary force FORCE:

$$\mathfrak{scope}(\langle \mathsf{FORCE}_{i,j}\ A\rangle) \supseteq \{ma_j\mathbf{:}ma_i\mathbf{:}t \mid t \in \mathfrak{subject}(A)\} \tag{10}$$

Consider e.g. the speech act where the user informs the system about the ticket price. This speech act influences the system's belief about the user's attitude towards prices. Hence $ma_s\mathbf{:}ma_u\mathbf{:}price \in \mathfrak{scope}(\langle \mathsf{Inform}_{u,s}\ Price(150\ \$)\rangle)$.

Interactions. Is there a relationship between these functions? We propose the following axiom for acts of the informative type.

$$\begin{array}{c}\text{If } \alpha = \langle \mathsf{Inform}_{i,j}\ A\rangle \text{ and } t \in \mathfrak{subject}(A) \cap \mathfrak{competence}(i) \\ \text{then } t \in \mathfrak{scope}(\alpha) \text{ and } ma_j\mathbf{:}t \in \mathfrak{scope}(\alpha).\end{array} \tag{11}$$

[5] Note that the competence of the speaker might depend on the hearer: two agents might disagree on the competence of a third agent at theme t. Hence competence should be a 2-argument function. As we only have two participants in our examples, we have dropped the second argument for the sake of simplicity.

[6] Note that the competence of i at his own beliefs and intentions will be ensured anyway lateron by standard axioms of introspection (cf. Sect. 5).

Suppose e.g. the user informs the system about his destination. As the user is competent at destinations, this influences the system's factual beliefs about the destination. And also about prices: a new destination means a new price. Hence $\mathfrak{scope}(\langle \mathsf{Inform}_{u,s} Dest(\mathrm{Paris})\rangle)$ contains *dest*, *price*, ma_s*:dest* and ma_s*:price*. Postulates for other illucutionary forces are in [33].

In the next two sections we show how topics allows us to change beliefs by means of two principles: adoption and preservation.

4 Semantics

Semantics is in terms of *possible worlds models* $\langle W, \mathcal{S}, D, V\rangle$, where (1) W is a set of worlds; (2) D is the domain; (3) V is a valuation mapping variable and constant symbols to elements of D, and associating to each possible world $w \in W$ an interpretation V_w of predicate symbols; (4) $\mathcal{S}$ is a collection of structures on W, consisting of

- relations $\mathcal{D}_\alpha$ and $\mathcal{B}_i$ for every $\alpha \in ACT$ and every $i \in AGT$, viewed as functions mapping W to 2^W, and
- functions $\mathcal{I}_i$ for every $i \in AGT$, mapping W to 2^{2^W}.

$\mathcal{D}_\alpha$ and $\mathcal{B}_i$ are accessibility relations as usual. $\mathcal{D}_\alpha(w)$ represents the possible results of act α. We speak of $\mathcal{B}_i(w)$ as the belief state of agent i. $\mathcal{I}_i$ are neighbourhood functions [8, Chap. 7]. Every set of worlds $U \in \mathcal{I}_i(w)$ corresponds to an intention of i.

The satisfaction relation $\Vdash$ is defined as usual. A useful abbreviation is $[\![A]\!] = \{w \in W : w \Vdash A\}$, called the extension of the formula A. Then

- $w \Vdash P(t_1, \ldots, t_n)$ if $\langle V_w(t_1), \ldots, V_w(t_n)\rangle \in V_w(P)$;
- $w \Vdash Done_\alpha A$ if $\mathcal{D}_\alpha(w) \cap [\![A]\!] \neq \emptyset$;
- $w \Vdash Bel_i A$ if $\mathcal{B}_i(w) \subseteq [\![A]\!]$;
- $w \Vdash Intend_i A$ if $[\![A]\!] \in \mathcal{I}_i(w)$.

Our notion of intention is neither closed under logical consequence nor under conjunction, in accordance with Bratman's, Cohen&Levesque's and Sadek's analyses of intention. Contrarily to these approaches, intention is primitive here, as in [36] and [31]. We thus generalize the semantics in [31], where only closure under logical consequence had been given up.[7]

Models must satisfy several constraints in order to account for the properties that we want the mental attitudes to have, as well as for the interaction between these attitudes. We present them in the rest of the section.

[7] 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

Standard constraints.

- $\mathcal{B}_i$ is serial, transitive and euclidean;
- $\mathcal{D}_\alpha$ is deterministic (i.e. $\mathcal{D}_\alpha(w)$ is either a singleton or the empty set);
- if $v \in \mathcal{B}_i(w)$ then $\mathcal{I}_i(v) = \mathcal{I}_i(w)$;
- if $U \in \mathcal{I}_i(w)$ then $U \cap \mathcal{B}_i(w) = \emptyset$.

The last two constraints say that intentions are introspective, and that if i entertains some intention then he believes that the contrary is currently the case [44]. For the sequel we need the following definition.

$$Atm_W(w, T) = \{p \in Atm \mid w \Vdash p \text{ and } \mathfrak{subject}(p) \subseteq T\}$$

Thus $Atm_W(w, \mathfrak{competence}(i))$ is the part of the real world w at which i is competent, and $Atm_W(w, \mathbb{T}_0 \setminus \mathfrak{scope}(\alpha))$ is the part of w that is independent of α.

Belief adoption constraint. Belief should amount to knowledge in the case of competence.

- For every $w \in W$ and every agent i there is some $v \in \mathcal{B}_i(w)$ such that $Atm_W(w, \mathfrak{competence}(i)) = Atm_W(v, \mathfrak{competence}(i))$.

This means that in the belief state of i there is a 'witness world' mirroring the part of the actual world i is competent at.

Preservation constraints. What is the effect of an act α on the actual world? If an atom is independent of α then its truth value should be preserved.

- If $w' \in \mathcal{D}_\alpha(w)$ then $Atm_W(w, \mathbb{T}_0 \setminus \mathfrak{scope}(\alpha)) = Atm_W(w', \mathbb{T}_0 \setminus \mathfrak{scope}(\alpha))$.

What is the effect of α on beliefs? In accordance with [30, 23] we suppose that when an agent i learns that α has been executed then i updates his belief state 'worldwise', mapping each $v \in \mathcal{B}_i(w)$ to a new set of worlds $\mathcal{D}_{i:\alpha}(v)$:[8]

- If $w' \in \mathcal{D}_\alpha(w)$ then $\mathcal{B}_i(w') = \bigcup_{v \in \mathcal{B}_i(w)} \mathcal{D}_{i:\alpha}(v)$

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. Indeed, it seems that rather than the interaction of intentions with goals or desires it is their interaction with belief which is crucial (e.g. in order to abandon intentions).

[8] One might think at first glance that i mentally executes just the same act α in every possible world. This cannot be the case; indeed, one can think of speech acts as always modifying the belief state of the agents, while leaving the physical world unchanged. Therefore we need a different act $i{:}\alpha$ depending on i and α, reflecting the effect of α on i's belief state.

We must thus introduce contextual acts into our language. We extend our definition of acts by the recursive

$$i{:}\alpha \text{ is an act if } i \text{ is an agent and } \alpha \text{ is an act}$$

We speak of $i{:}\alpha$ as the mental act associated to α. Technically these acts can be compared to Skolem functions, which are a device to obtain completeness: the same will be the case here.

How are α and $i{:}\alpha$ related? $i{:}\alpha$ being the mental image of α, its scope is obtained by stripping off ma_i from $\mathfrak{scope}(\alpha)$, i.e.

$$\mathfrak{scope}(i{:}\alpha) = \{c{:}t \mid ma_i{:}c{:}t \in \mathfrak{scope}(\alpha)\}$$

Hence our scope function is completely defined by the non-contextual part of the domain. In particular we have $\mathfrak{scope}(i{:}i{:}\alpha) = \mathfrak{scope}(i{:}\alpha)$. As in this paper we have restricted the context of a theme t to length at most 2, we suppose consequently that

$$\mathfrak{scope}(i{:}j{:}k{:}\alpha) = \mathfrak{scope}(i{:}j{:}\alpha)$$

What is the effect of acts on intentions? We do not consider here the generation of new intentions, because we consider that this can be done in a separate subsequent step. We focus on the preservation of intentions via independence.[9] To that end, given a set of worlds $U \subseteq W$ we define

$$\mathfrak{subject}(U) = \bigcup_{\{p \in Atm \mid p \text{ occurs in the set of minimal models of } U\}} \mathfrak{subject}(p)$$

This allows us to compute the subject of the extension $[\![A]\!]$ of a formula A. Now we are ready to define preservation of intentions.

- If $w' \in \mathcal{D}_\alpha(w)$, $U \in \mathcal{I}_i(w)$, and $\mathfrak{scope}(i{:}\alpha) \cap \mathfrak{subject}(U) = \emptyset$ then $U \in \mathcal{I}_i(w')$

Hence intentions are preserved if their subject is not in the scope of the mental act $i{:}\alpha$ associated to α.

[9] It is already a 'built-in' feature that sometimes intentions are abandoned because of new beliefs. Suppose e.g. $w \Vdash Intend_i A$, i.e. $[\![A]\!] \in \mathcal{I}_i(w)$. Let $w' \in \mathcal{D}_\alpha(w)$ and suppose there exists $v' \in \mathcal{B}_i(w')$ such that $v' \Vdash A$. Now we cannot have $[\![A]\!] \in \mathcal{I}_i(w')$, because the constraint linking $\mathcal{B}_i$ and $\mathcal{I}_i$ requires that $[\![A]\!] \cap \mathcal{B}_i(w) = \emptyset$.

According to [10], yet another condition triggering the abandon of intentions is the belief that it can never be satisfied. We do not treat this here, because it requires a temporal operator that we haven't introduced here in order to simplify the exposition. Such a modal operator is integrated into our framework in [33].

5 Axiomatics

Standard axioms. Just as in [10, 35, 42], with each belief operator Bel_i we associate the (normal) modal logic KD45 [26]. All $Done_\alpha$ operators obey the principles of the normal modal logic K, plus the converse of the modal axiom D, i.e. we suppose speech acts to be deterministic. $Intend_i$ operators validate neither the K-axiom nor necessitation. They have the classical modal logic RE [8, Chap. 7], whose only principle is $\dfrac{A_1 \leftrightarrow A_2}{Intend_i A_1 \leftrightarrow Intend_i A_2}$. Interactions between the different mental attitudes (in the spirit of [10, 44]) are expressed by:

$$Intend_i A \to Bel_i Intend_i A \tag{12}$$

$$\neg Intend_i A \to Bel_i \neg Intend_i A \tag{13}$$

$$Intend_i A \to Bel_i \neg A \tag{14}$$

$$(Done_\alpha \top \wedge \neg Bel_i \neg Done_{i:\alpha} A) \leftrightarrow (Done_\alpha \neg Bel_i \neg A) \tag{15}$$

(12) and (13) express that intentions are introspective. They entail that $Bel_i Intend_i A \leftrightarrow Intend_i A$ and $Bel_i \neg Intend_i A \leftrightarrow \neg Intend_i A$. (14) makes that intentions are abandoned if A is compatible with the beliefs of the agent. (15) means that if A was compatible with i's beliefs before α then it remains compatible with i's beliefs that A was true before the mental execution of α. The rest of the axiom schemas is topic-based.

Belief adoption axiom.

$$Bel_i A \to A \text{ if } \mathfrak{subject}(A) \subseteq \mathfrak{competence}(i)$$

The schema expresses that if i believes that A and is competent at A, then A is true.

For example $Bel_u Dest(\text{Paris}) \to Dest(\text{Paris})$ because $\mathfrak{subject}(Dest(\text{Paris})) \subseteq \mathfrak{competence}(u)$. From that the formula $Bel_s Bel_u Dest(\text{Paris}) \to Bel_s Dest(\text{Paris})$ can be proved by using the standard modal necessitation and K-principles for Bel_s. On the contrary, $Bel_u Price(80\ \$) \to Price(80\ \$)$ is not an instance of our axiom schema, because $\mathfrak{subject}(Price(80\ \$)) \not\subseteq \mathfrak{competence}(u)$.

Preservation axioms.

$$A \to \neg Done_\alpha \neg A \text{ if } \begin{cases} \mathfrak{scope}(\alpha) \cap \mathfrak{subject}(A) = \emptyset \\ \text{and } A \text{ contains no } Done_\beta \text{ operator.} \end{cases}$$

$$Intend_i A \to \neg Done_\alpha \neg Intend_i A \text{ if } \begin{cases} \mathfrak{scope}(\alpha) \cap \mathfrak{subject}(Intend_i A) \cap \mathbb{T}_1 = \emptyset \\ \text{and } A \text{ contains no } Done_\beta \text{ operator.} \end{cases}$$

The restriction on the form of A is necessary because our reading of $Done_\beta$ is that β has just been performed (and not at some arbitrary time point in the past).

The second intention preservation axiom is a strengthening of the first axiom, because independence of α and A is restricted to $\mathbb{T}_1$.

Completeness and theorem proving. We prove completeness in two steps. First of all we use a translation mapping the non-normal $Intend_i$ operators to normal modal operators: $Intend_i A$ becomes $\Diamond_{i,1}(\Box_{i,2}A \wedge \Box_{i,3}\neg A)$. This is a simple technique which has already been successfully employed in order to mechanize the standard non-normal logics [18, 14, 19, 17]. Then completeness of the resulting multi-modal logic can be proven in a fairly standard way. Indeed, all the semantical conditions are in a particular class that has been investigated in mathematical logic, and for which general completeness results exist [47, 7, 16]. The only difference here is that the preservation and adoption conditions depend on topics. It has been shown in [6] that nevertheless the standard Henkin proof technique applies straightforwardly.

Our current research focusses on the implementation of a tableau theorem prover for our logic. In previous work we have extended the standard tableaux method in order to deal with dependence information in reasoning about actions [5, 6]. The extension of our approach to the present topic-based framework is straightforward.

6 Discussion

We have sketched a theory of change in the context of dialogues. It is based on the notion of topic of information, which is exploited through topic-based axioms of belief adoption and preservation. Perrault and Appelt&Konolige have argued that defaults are crucial elements in a theory of speech acts. In a sense, what we do is to transfer that task to the metalinguistic relations $\mathfrak{competence}$ and $\mathfrak{scope}$. This permits to stay with a monotonic framework.

We have supposed that the set of topics associated with a formula is determined by those of the atomic formulas occurring in it. This is certainly a debatable choice. It was mainly motivated by representational economy. Notwithstanding, the way we use the $\mathfrak{subject}$ function is sound: suppose e.g. $\mathfrak{subject}(p) = \{t\}$, $\mathfrak{subject}(q) = \{t'\}$, and $\mathfrak{scope}(\alpha) = \{t'\}$. Hence p and $p \wedge (q \vee \neg q)$ do not have the same subject, and $Done_\alpha p \to p$ is an instance of the preservation axiom, while $Done_\alpha(p \wedge (q \vee \neg q)) \to (p \wedge (q \vee \neg q))$ is not. Nevertheless, the latter formula can be deduced from the former by standard modal logic principle as $p \leftrightarrow p \wedge (q \vee \neg q)$ we have $Done_\alpha p \leftrightarrow Done_\alpha(p \wedge (q \vee \neg q))$. Hence $Done_\alpha p \to p$ is equivalent to $Done_\alpha(p \wedge (q \vee \neg q)) \to (p \wedge (q \vee \neg q))$.

We did not formulate such strong compositionality axioms for the $\mathfrak{scope}$ function. The reason is that a speech act might influence more than the topics of its propositional contents. For example, the scope of $\langle \mathsf{Inform}_{u,s}\, Class(1\mathrm{st})\rangle$ contains not only ma_u*:*ma_s*:class* but also ma_u*:*ma_s*:price*. Our hypothesis here is that the scope of a speech act is determined by the subject of its propositional contents together with the integrity constraints e.g. linking destinations, classes, and prices. This is subject of future research.

References

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50(2):510–530, June 1985.

[2] Douglas Appelt and Kurt Konolige. A nonmonotonic logic for reasoning about speech acts and belief revision. In M. Reinfrank, J. de Kleer, M.L. Ginsberg, and E. Sandewall, editors, *Proceedings of Second International Workshop on Non-Monotonic Reasoning*, volume 346 of *LNAI*, pages 164–175. Springer-Verlag, June 1989.

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

[4] Daniel Büring. Topic. In Peter Bosch and Rob van der Sandt, editors, *The Focus Book*. Cambridge University Press, 1995.

[5] Marcos Alexandre Castilho, Olivier Gasquet, and Andreas Herzig. Modal tableaux for reasoning about actions and plans. In Sam Steel and Rachid Alami, editors, *European Conference on Planning (ECP'97)*, number 1348 in LNAI. Springer-Verlag, 1997.

[6] Marcos Alexandre Castilho, Olivier Gasquet, and Andreas Herzig. Formalizing action and change in modal logic I: The frame problem. *Journal of Logic and Computation*, 1999. to appear.

[7] Laurent Catach. *Les logiques multi-modales*. PhD thesis, Université Paris VI, France, 1989.

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

[9] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence Journal*, 42(2–3):213–261, 1990.

[10] Philip R. Cohen and Hector J. Levesque. Persistence, intentions, and commitment. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, chapter 3, pages 33–69. MIT Press, Cambridge, MA, 1990.

[11] Philip R. Cohen and Hector J. Levesque. Rational interaction as the basis for communication. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, chapter 12, pages 221–255. MIT Press, Cambridge, MA, 1990.

[12] Robert Demolombe and Andrew J.I. Jones. On sentences of the kind "sentence 'p' is about topic t": some steps towards a formal-logical analysis. In Hans Jürgen Ohlbach and Uwe Reyle, editors, *Essays in Honor of Dov Gabbay*. Kluwer, 1998.

[13] R. L. Epstein. *The Semantic Foundations of Logic Volume 1: Propositional Logic*. Kluwer Academic Publishers, 1990.

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

[15] Luis Fariñas del Cerro, Andreas Herzig, Dominique Longin, and Omar Rifi. Belief reconstruction in cooperative dialogues. In Fausto Giunchiglia, editor, *Proceedings of 8th International Conference on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA'98)*, volume 1480 of *LNAI*, pages 254–266, Bulgaria, September 1998. Springer-Verlag.

[16] Olivier Gasquet. *Déduction automatique en logique multi-modale par traduction*. PhD thesis, Université Paul Sabatier, Toulouse, July 1994.

[17] Olivier Gasquet. Completeness results in neighbourhood semantics for mutimodal monotonic and regular logics. *Journal of the IGPL*, 4(3):419–428, 1996.

[18] Olivier Gasquet and Andreas Herzig. Automated deduction for non-normal modal logics. In Andrews I. Jones and Marek Sergot, editors, *2nd International Workshop on Deontic Logic in Computer Science (DEON'94)*, pages 136–144. TANO, Oslo, January 1994.

[19] Olivier Gasquet and Andreas Herzig. From classical to normal modal logics. In Heinrich Wansing, editor, *Proof Theory of Modal Logics*, number 2 in Applied Logic Series, pages 293–311. Kluwer Academic Publishers, 1996.

[20] J. Ginzburg. Resolving questions i,ii. *Linguistics and Philosophy*, 18:567–609, 1995.

[21] E. Giunchiglia, G. N. Kartha, and V. Lifschitz. Representing action: indeterminacy and ramifications. *Artificial Intelligence Journal*, 95:409–438, 1997.

[22] N. Goodman. About. *Mind*, LXX(277), 1961.

[23] Gösta Grahne. Updates and counterfactuals. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 269–276. Morgan Kaufmann Publishers, 1991. extended version to appear in the J. of Logic and Computation.

[24] H. Paul Grice. Logic and conversation. In J. P. Cole and J. L. Morgan, editors, *Syntaxe and Semantics : Speech acts*. Academic Press, New York, 1975. Also available in [25, Chapter 2].

[25] H. Paul Grice. *Studies in the way of words*. Harvard University Press, USA, 3rd edition, 1989.

[26] Joseph Y. Halpern and Yoram Moses. A guide to the modal logics of knowledge and belief: preliminary draft. In *Proceedings of IJCAI'85*, pages 480–490, Los Angeles, 1985.

[27] David Harel. Dynamic logic. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume II. D. Reidel Publishing Company, 1984.

[28] Andreas Herzig and Dominique Longin. Belief dynamics in cooperative dialogues. In N. van Leusen, R. van Rooy, and H. Zeevat, editors, *Proceedings of Amstelogue*, Amsterdam, May 1999. To appear, available from web site.

[29] Andreas Herzig and Omar Rifi. Propositional belief base update and minimal change. *Artificial Intelligence Journal*, 115(1):107–138, December 1999.

[30] Hirofumi Katsuno and Alberto O. Mendelzon. On the difference between updating a knowledge base and revising it. In Peter Gärdenfors, editor, *Belief Revision*, number 29 in Cambridge Tracts in Theorical Computer Science, pages 183–203. Cambridge University Press, 1992.

[31] Kurt Konolige and Martha E. Pollack. A representationalist theory of intention. In *Proceedings of the 13rd International Joint Conference on Artificial Intelligence (IJCAI'93)*, pages 390–395, Chambery, France, 1993.

[32] D.K. Lewis. General semantics. In D. Davidson and G. Harman, editors, *Semantics of natural language*. D. Reidel Publishing Company, 1972.

[33] Dominique Longin. *Interaction rationnelle et évolution des croyances dans le dialogue : une logique basée sur la notion de topique*. PhD thesis, Université Paul Sabatier, Institut de Recherche en Informatique de Toulouse (IRIT), Toulouse, France, November 1999.

[34] J. McCarthy and P. Hayes. Some philisophical problems from the standpoint of artificial intelligence. In B. Meltzer and D. Michie, editors, *Machine Intelligence*, volume 4, pages 463–502. Edinburgh University Press, 1969.

[35] C. Raymond Perrault. An application of default logic to speech act theory. In Philip R. Cohen, Jerry Morgan, and Martha E. Pollack, editors, *Intentions in Communication*, chapter 9, pages 161–185. MIT Press, Cambridge, MA, 1990.

[36] Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a bdi-architecture. In J. A. Allen, R. Fikes, and E. Sandewall, editors, *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, pages 473–484, San Mateo, CA, 1991. Morgan Kaufmann Publishers.

[37] Anand S. Rao and Michael P. Georgeff. An abstract architecture for rational agents. In Bernhard Nebel, Charles Rich, and William Swartout, editors, *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 439–449, Cambridge, Massachusetts, 1992.

[38] Anand S. Rao and Michael P. Georgeff. Decision procedures for bdi logics. *Journal of Logic and Computation*, 8(3):293–343, June 1998. Special Issue: Computational and Logical Aspects of Multiagent Systems, Oxford univerty Press.

[39] Ray Reiter. A logic for default reasoning. *Artificial Intelligence Journal*, 13:81–132, 1980.

[40] David Sadek and Philippe Bretier. A rational agent as the kernel of a cooperative spoken dialogue system: Implementing a logical theory of interaction. In J. P. Müller, M. J. Wooldridge, and N. R. Jennings, editors, *Proceedings of the ECAI'96, Workshop on Agent Theories, Architectures, and Languages*, LNAI. Springer-Verlag, 1997.

[41] David Sadek, Philippe Bretier, and Franck Panaget. Artimis: Natural dialogue meets rational agency. In Martha E. Pollack, editor, *Proceedings of the 15th International Joint Conference on Artificial Intelligence (IJCAI'97)*, volume 2, pages 1030–1035. Morgan Kaufmann Publishers, August 1997.

[42] M. D. Sadek. *Attitudes mentales et interaction rationnelle : vers une théorie formelle de la communication*. PhD thesis in Computer Sciences, Université de Rennes I, Rennes, France, June 1991.

[43] M. D. Sadek. Dialogue acts are rational plans. In *Proceedings of ESCA/ETRW, Workshop on The Structure of Multimodal Dialogue (Venaco II)*, pages 1–29, Maratea, Italie, September 1991.

[44] M. D. Sadek. A study in the logic of intention. In Bernhard Nebel, Charles Rich, and William Swartout, editors, *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning (KR'92)*, pages 462–473, Cambridge, Massachusetts, October 1992.

[45] M. D. Sadek. Towards a theory of belief reconstruction: Application to communication. *Speech Communication Journal'94, special issue on Spoken Dialogue*, 15(3–4):251–263, 1994.

[46] M. D. Sadek, A. Ferrieux, A. Cozannet, P. Bretier, F. Panaget, and J. Simonin. Effective human-computer cooperative spoken dialogue: The AGS demonstrator. In *Proceedings of ICSLP'96 International Conference on Spoken Language Processing*, Philadelphia, USA, October 1996.

[47] H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logics. In S. Kanger, editor, *Proceedings of 3rd Scandinavian Logic Symposium 1973*, volume 82 of *Studies in Logic*, North-Holland, 1975.

[48] Erik Sandewall. *Features and Fluents*. Oxford University Press, 1994.

[49] J. R. Searle and D. Vanderveken. *Foundation of illocutionary logic*. Cambridge University Press, 1985.

[50] John R. Searle. *Speech acts: An essay in the philosophy of language*. Cambridge University Press, New York, 1969.

[51] Michael Thielscher. Computing ramifications by postprocessing. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI'95)*, pages 1994–2000, Montreal, Canada, 1995.

[52] J. van Kuppevelt. *Topic en Comment*. PhD thesis, University of Nijmegen, 1991.

[53] J. van Kuppevelt. Discourse structure, topicality and questioning. *Linguistics*, 31:109–147, 1995.