Updating Mental States from Communication

A.F. Dragoni[1], P. Giorgini[2], and L. Serafini[3]

[1]University of Ancona, 60131, Ancona, Italy.dragon@inform.unian.it
[2]DISA, University of Trento, 38100, Trento, Italy. pgiorgini@cs.unitn.it
[3]ITC-IRST, 38050, Trento, Italy, serafini@irst.itc.it

Abstract. In order to perform effective communication agents must be able to foresee the effects of their utterances on the addressee's mental state. In this paper we investigate on the update of the mental state of an hearer agent as a consequence of the utterance performed by a speaker agent. Given an agent communication language with a STRIPS-like semantics, we propose a set of criteria that allow to bind the speaker's mental state to its uttering of a certain sentence. On the basis of these criteria, we give an abductive procedure that the hearer can adopt to partially recognize the speaker's mental state that led to a specific utterance. This procedure can be adopted by the hearer to update its own mental state and its image of the speaker's mental state.

1 Introduction

In multi-agent systems, communication is necessary for the agents to cooperate and coordinate their activities or simply to avoid interfering with one another. If agents are not designed with embedded pre-compiled knowledge about the beliefs, intentions, abilities and perspective of other agents, they need to exchange information to improve their social activities. However, in real application domains, communication might be a limited resource (limited bandwidth, low signal/noise ratio etc.). In such cases, it is very important that, when deciding whether to send a message, agents consider their expected benefits vs. the costs of communication.

BDI agents (namely agents able to have Beliefs, Desires and Intentions) [21–23, 17, 7, 25] are supposed to have a, so called, *mental state* which contains beliefs, desires and intentions about the environment, and about the other agents' beliefs, desires and intentions. The behavior of each agent strongly depends on its mental state. Communication is the main road for agents to affects the behavior of other agents by exchanging information about the environment and their beliefs, desires and intentions.

Communication is also supposed to be intentional, i.e. activated by the speaker's reasoning about its own beliefs, desires and intentions. In other words, it is generally possible to regard utterances as the consequence of the speaker's being in a particular mental state, that provokes its desire to influence the hearer's mental state. This position is very general and is independent of the particular class of speech acts (assertive, commissive, directive, declarative or expressive). If utterances are the effects of mental conditions, it seems natural to suppose that the hearer tries to recognize the speaker's mental state through some form of backward reasoning (e.g., abduction) from the kind and the content of the received communication to the hypothetical mental state that originated it. If this were possible, then we could develop logical theories that partially predict dialogue.

The main goal of this paper is to provide some methods that a hearer can adopt to recognize the speaker's mental state on the basis of its utterances. To do this we propose:

1. a formal representation of mental states based on the theory of contexts [3];
2. a correlation between the mental state of an agent and its utterances, based on the plan-based theory of speech acts [4];
3. a formal characterization of the operations which can be performed by the hearer to update its mental state at the receipt of a message.

The novelty of the paper stands in the fact that, not only do we devise an abductive theory for revising the mental state of an agent, but we also relate this theory to the semantics of agent communication languages.

The paper is structured as follows. In section 2 we present the context framework, which we use to formalize agents' mental states. Section 3 describes a simple scenario used as an explanatory example throughout the paper. Section 4 illustrates how to exploit the plan-based theory of speech acts to correlate the speaker's mental state to its utterances. In section 5 we recall Konolige's definition of causal theories and abduction, and we extend it to multiple contexts. We define three basic operations on mental states: abductive expansion, abductive contraction and abductive revision. Section 6 presents some abductive methods which can be used to update the hearer's image of the speaker's mental state. Finally, sections 7 and 8 present the related work and draw conclusions, respectively.

2 Mental States

Each agent is supposed to be characterized by a *mental state*. We regard mental state as a structure based on two primitive mental attitudes: *beliefs* and *desires*.[1]

Following [3, 10–12], we use propositional contexts to formalize agents' mental states. A context is defined as a set of formulae closed under a set of inference rules (a theory). Suppose to have a set of agents I. For each agent $i \in I$, its set of beliefs and intentions are represented by the contexts B_i and I_i, respectively. A formula ϕ in the context B_i (denoted by the pair $B_i : \phi$) represents the fact that i believes ϕ and, analogously, a formula ϕ in the context I_i ($I_i : \phi$) represents the fact that i has the intention to bring about ϕ. In general beliefs and intentions are not expressed in the same language. Although contexts support this possibility, for the sake of presentation we consider the simpler case in which the languages for beliefs and intentions coincides. We call this language L_i. L_i contain formulas to represent the environment and, being the agents part of the environment, formulas that express the fact that an agent has a certain belief or intention. Therefore L_i contains the following atomic formulas, for any agent j:

$B_j\phi$ is an atomic proposition of L_i, with the intuitive meaning that agent j believes ϕ;
$I_j\phi$ is an atomic proposition of L_i, with the intuitive meaning that agent j intends to bring about ϕ.

Formulas of the form $B_j\phi$ and $I_j\phi$ are called BDI atoms [1]. Reasoning capabilities of an agent i on its beliefs and intentions are represented in the contexts I_i and B_j, by two sets of inference rules R_{B_i} and R_{I_i}, respectively. Examples of reasoning capabilities could be any set of logical inference rules. Reasoning capabilities are supposed to be general purpose reasoning

[1] Intuitively, intentions represent what the agent desires to be true (or false) and also it believes it could be true (or false). The "could" means that the agent is able to act in order to change the external world and/or the other agents mental states to reach the desired state of affairs. Of course, this opens many critical questions which are far aside the scope of this dissertation. However, although we distinguished between the two, we still continue the tradition of calling "intention" what should be better defined "desire".

machineries which does not contains special inference rules for BDI atoms. For the sake of this paper we suppose that any inference machinery is the set of rules for propositional logic. In B_i and I_i, BDI atoms are considered as any other atomic formulas; this implies, for instance, that $B_j\phi$ and $B_j(\phi \vee \phi)$ are completely independent beliefs of i. On the other hand, if i ascribes to the agent j enough reasoning capabilities, then either i believes that j believes both ϕ and $\phi \vee \phi$, or i believes that j does believe neither ϕ nor $\phi \vee \phi$. The relation between BDI atoms in B_i and I_i, therefore, depends on the reasoning capabilities that i ascribes to j. The beliefs, the intentions, and the reasoning capabilities that i ascribes to another agent j are explicitly modeled by means of a mental state, called i's *image of j's mental state*. In particular, i's beliefs about the beliefs and the intentions of j, are represented by the contexts B_iB_j and B_iI_j respectively. The same representation is used to formalize i's intentions regarding j's beliefs and intentions, that is the contexts I_iB_j and I_iI_j. Analogously, i's beliefs about j's beliefs about another agent k are formalized by the pair of contexts $B_iB_jB_k$ and $B_iB_jI_k$. This iteration can go on infinitely, but in many cases it is enough to consider a finite amount of iterations.

The intuitive interpretation of a formula depends on the context. For instance, as already said, the formula ϕ in the context B_i, denoted by $B_i : \phi$, expresses the fact that agent i believes ϕ. The same formula in the context $B_iB_jI_i$, denoted by $B_iB_jI_i : \phi$, expresses the fact that i believes that j believes that i intends ϕ. On the other hand, different formulas in different contexts can represent the same fact. For instance, the formulas ϕ in the context B_iI_j and the formula $I_j\phi$ in the context B_i have the same meaning. The effect of this "meaning overlapping" is that, contexts cannot be considered as isolated theories, and that the set of theorems of a context might affect the set of theorems in another context. The interaction between contexts is formalized by *bridge rules*. In particular, we use the following bridge rules between contexts in a mental states and contexts in images of mental states.

$$\frac{\alpha : B_i\phi}{\alpha B_i : \phi}\ \mathcal{R}_{dn.B} \qquad \frac{\alpha B_i : \phi}{\alpha : B_i\phi}\ \mathcal{R}_{up.B} \qquad \frac{\alpha : I_i\phi}{\alpha I_i : \phi}\ \mathcal{R}_{dn.I} \qquad \frac{\alpha I_i : \phi}{\alpha : I_i\phi}\ \mathcal{R}_{up.I}$$

$\mathcal{R}_{dn.B}$ and $\mathcal{R}_{dn.I}$ are called *Reflection down* and $\mathcal{R}_{up.B}$ and $\mathcal{R}_{up.I}$ *Reflection up*. Reflection up and reflection down are often used in combination. For instance, reflection down allows an agent i to convert a formula $B_j\phi$ into a simpler format (by eliminating B_i) in order to perform reasoning about ϕ in its image of j's mental state. Reflection up are used by i to lift up, in its mental state, the result of such a reasoning This reasoning pattern allows i to prove relations among BDI atoms. For instance, to prove $B_j\phi \supset B_j(\phi \vee \phi)$ in the contexts of its beliefs, i can perform the following deduction:

$$\dfrac{\dfrac{\dfrac{\dfrac{B_i : B_j\phi}{B_iB_j : \phi}\ \mathcal{R}_{dn.B}}{B_iB_j : \phi \vee \psi}\ \vee\mathrm{I}}{B_i : B_j(\phi \vee \psi)}\ \mathcal{R}_{up.B}}{B_i : B_j\phi \supset B_j(\phi \vee \psi)}\ \supset\mathrm{I}$$

The beliefs and the intentions of an agent are not independent. The relation between the beliefs and the intentions of an agent can also be represented by bridge rules from the context of its beliefs to that of its intentions. For instance, the bridge rule:

$$\frac{B_i : raining}{I_i : bring\_umbrella}\ B2I$$

formalizes the fact that, if agent i believes that *it is raining*, then i intends *to bring an umbrella*.

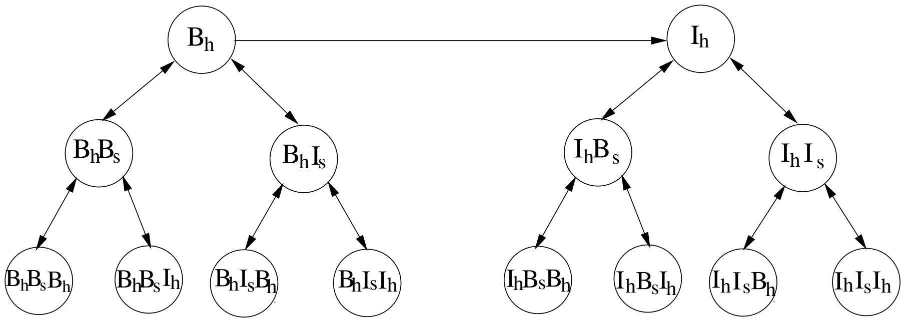

Fig. 1. Contexts for agent h

Since we are interested in formalizing the effects of an utterance performed by a speaker on the beliefs and intentions of a hearer, we consider the two agents s and h denoting the speaker and hearer, respectively. Furthermore, we focus only on the effects of the utterance on the hearer's mental state. We therefore consider only the contexts for the hearer's beliefs and intentions (namely B_h and I_h), the contexts for the hearer's beliefs and intentions regarding the speaker's beliefs and intentions (namely B_hB_s, B_hI_s, I_hB_s, and I_hI_s), and the contexts for the hearer's beliefs and intentions regarding the speaker's beliefs and intentions regarding the hearer's beliefs and intentions. Of course, this nesting could be extended indefinitely (for more details see [8, 9, 13]), but three levels (as depicted in figure 1, where circles represent contexts and arrows represent bridge rules) are sufficient to illustrate the abductive methods for the inference of mental states from communicative acts.

The logical systems that formalize the reasoning with a set of contexts connected by bridge rules are called multi-context systems [12].[2] In the following we do not give a formal definition of multi context system, we rather apply the general definition to our special case.

Let MC be the multi-context system composed of the contexts shown in figure 1 connected by the bridge rules $\mathcal{R}_{up.}$ and $\mathcal{R}_{dn.}$, and by a set of bridge rules $B2I$ from B_h to I_h. Deductions in MC are trees of wffs starting from a finite number of formulas (either axioms or assumptions), possibly belonging to distinct contexts, and applying a finite number of inference and bridge rules. A wff $\alpha : \phi$ is derivable from a set of axioms AX in MC (in symbols, $\mathsf{AX} \vdash_{\mathsf{MC}} \alpha : \phi$), if there is a deduction that ends with $\alpha : \phi$ and whose axioms are in AX. For a detailed description on the proof theory of MC we refer the reader to [12]. Given a set of axioms AX, for each context α, let $\alpha^* = \{\phi \mid \mathsf{AX} \vdash_{\mathsf{MC}} \alpha : \phi\}$.

The *mental state* of the agent h is defined as the pair of sets containing h's beliefs and h's intentions which are derivable in MC. In symbols $ms(h) = \langle B_h^*, I_h^* \rangle$. Analogously *$h$'s image of s' mental state* is composed of the set of h's beliefs on s' beliefs and the set of h's beliefs on s' intentions. In symbols $ms(h, s) = \langle B_hB_s^*, B_hI_s^* \rangle$. In general we define $ms(h, s, h, \ldots s, h) \stackrel{\triangle}{=} \langle B_hB_sB_h \ldots B_sB_h^*, B_hB_sB_h \ldots B_sI_h^* \rangle$, which stands for s's image of h's image of s's image of ... of h's image of s's mental state.

According to the definition above, in MC hearer's mental states and hearer's images of mental states are completely determined by the set of axioms AX. Therefore, the effects of the receipt of a message from the hearer, on its mental state and on its images of mental states, can be represented by a suitable change of the set of axioms AX into a new set of axioms AX'. For each context α, we define the set $\alpha'^* = \{\phi \mid \mathsf{AX}' \vdash_{\mathsf{MC}} \alpha : \phi\}$. Analogously, we can provide

[2] In [12], multi-context systems are called multi-language systems to stress the fact that they allow for multiple distinct languages

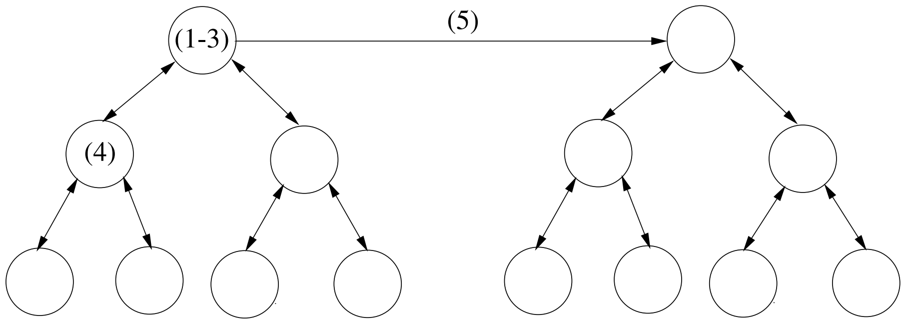

Fig. 2. Initial h's mental state

the definition of the updated (due to the utterance) mental state and images of mental states $ms'(\ldots)$.

3 Working Example

Let MC be the multi-context system composed of the contexts shown in figure 1 connected by the bridge rules $\mathcal{R}_{up.}$ and $\mathcal{R}_{dn.}$ and by the following bridge rulefrom B_h to I_h:

$$\frac{B_h : \mathit{temp\_higher\_20^\circ}}{I_h : \mathit{conditioning\_on}}\ B2I \tag{1}$$

with the following meaning: If the hearer believes that the temperature is higher than 20 degrees, then it has the intentions to switch the conditioning on. Let us consider the situation, represented in figure 2, where h has the beliefs represented by axioms (2)–(4) in the context B_h.

$$B_s \mathit{temp\_higher\_20^\circ} \supset \mathit{temp\_higher\_20^\circ} \tag{2}$$

If s believes that the temperature is higher than 20 degrees, then the temperature is higher than 20 degrees.

$$\mathit{conditioning\_on} \supset \neg \mathit{temp\_higher\_20^\circ} \tag{3}$$

If the conditioning is on, then the temperature is lower than 20 degrees.

$$B_s(\mathit{temp\_higher\_20^\circ} \wedge \mathit{conditioning\_on}) \supset I_s \mathit{stop\_working} \tag{4}$$

If s believes that the temperature is higher than 20 degrees and that the conditioning is on, then s intends to stop working. Let suppose also that h ascribes to s the beliefs represented by axiom (5) in the context $B_h B_s$.

$$\neg \mathit{conditioning\_on} \supset \mathit{temp\_higher\_20^\circ} \tag{5}$$

If the conditioning is off, then the temperature is higher than 20 degrees.

With the set of axioms

$$\mathsf{AX} = \{B_h : (2), B_h : (3), B_h : (4), B_h B_s : (5)\}$$

h infers that, if s believes that the conditioning is off, then h will adopt the intention to switch it on. In symbols:

$$\mathsf{AX}, B_h : B_s \neg \mathit{conditioning\_on} \vdash_{\mathsf{MC}} I_h : \mathit{conditioning\_on}$$

The corresponding deduction in MC is shown in figure 3.

$$\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{B_h : B_s \neg \mathit{conditioning\_on}}{B_h B_s : \neg \mathit{conditioning\_on}}\ \mathcal{R}_{dn.B} \qquad B_h B_s : (5)}{B_h B_s : \mathit{temp\_higher\_20^\circ}}\ \supset\!\mathrm{E}}{B_h : B_s \mathit{temp\_higher\_20^\circ}}\ \mathcal{R}_{up.B} \qquad B_h : (2)}{B_h : \mathit{temp\_higher\_20^\circ}}\ \supset\!\mathrm{E}}{I_h : \mathit{conditioning\_on}}\ B2I$$

Fig. 3. Deduction of $I_h : \mathit{conditioning\_on}$, from $B_h : B_s \mathit{temp\_higher\_20^\circ} \supset \mathit{temp\_higher\_20^\circ}$, with axiom $B_h B_s : \neg \mathit{conditioning\_on} \supset \mathit{temp\_higher\_20^\circ}$.

4 Plan-Based Model of Speech Acts

The plan-based vision of *speech acts* [4], which treats them as actions and represents them as STRIPS-like operators, offers us an intuitive way to correlate the speaker's mental state to its utterances. The "trick" is in the modeling of the speech acts' preconditions. To illustrate this idea we use a simplified and revised version of the INFORM operator which is the prototypical member of the assertive speech act class [24].

In a plan-based theory of speech acts, INFORM(s, h, ϕ) is generally defined to be an action whose main effect on hearer's mental state is that the hearer believes that the speaker believes the propositional content ϕ, and its prerequisite is that the speaker believes ϕ (sincerity).

Speech act	*Preconditions*	*Main effects*
$\mathrm{INFORM}(s, h, \phi)$	$\phi \in B_s^*$	$\phi \in B_h B_s^*$

The structure of this simple operator is closely related to the one by Cohen and Perrault [4]. We envisage, however, a larger range of effects described in the following:

1. The effects of the INFORM operator given by Cohen and Perrault [4] are the main effects here. The complete effects of a speech act on the hearer's mental state are beyond the speaker's control. We think that part of these *perlocutionary effects* are the result of some kind of abductive reasoning performed by the hearer from the received communication and from his actual mental state (which is in general different from the speaker's image of the hearer's mental state).
2. The precondition of the INFORM operator does not include the speaker's goal to perform such a speech act. As we see later, the speaker's actual intentions which leads to the execution of the speech act, are ascribed by the hearer to the speaker, again with some kind of abductive reasoning.

The basic assumption in this paper is that there is a causal relationship between an agent's mental state and its possible uttering a sentence. We may say that s plans an $\mathrm{INFORM}(s, h, \phi)$ because of the facts that:

I1. *s has the intention* to bring h in a mental state where a formula ψ (which might differ from ϕ itself) is either believed or intended by h; i.e., either $\psi \in I_s {B'_h}^*$ or $\psi \in I_s {I'_h}^*$.
I2. *s doesn't believe* that h is already in that mental state: $\psi \notin B_s B_h^*$ (resp. $\psi \notin B_s I_h^*$).
I3. *s believes* that if it performs the INFORM act, then h will be in a mental state in which it believes (resp. intends) ψ; i.e., $\psi \in B_s {B'_h}^*$ (resp. $\psi \in B_s {I'_h}^*$).
I4. s can perform the INFORM; that is, INFORM's precondition holds before and after performing the speech act; i.e., $\phi \in B_s^*$ and $\phi \in {B'_s}^*$.

5 Contextual Abduction and Revision

Before introducing abduction for contexts, let us briefly recall the main concepts of causal theory, abduction, and abductive explanation introduced by Konolige in [15]. Roughly speaking, abduction is an abstract hypothetical inferential schema that, given a causal theory of the domain, and an observation on a set of observable effects, looks for an explanation for them. An explanation for an observation is a minimal set of hypothesis, chosen among a set of possible causes, which if "added" to the causal theory, justify the observation. Syntactic propositional accounts of abduction formalize *causes* and *effects* as literals of a finite propositional language L, and the causal theory of the domain (*domain theory*) as a propositional theory of L.

A *simple causal theory* on a finite propositional language L, is a tuple $T = \langle C, E, \Sigma \rangle$, where C and E are sets of literals of L, and Σ is a theory on L. An *abductive explanation* (ABE) of an observation $O \subseteq E$, under a domain theory Σ, is a finite set $A \subseteq C$ such that:

- $\Sigma \cup A \nvdash \bot$ (A is consistent with Σ)
- $\Sigma \cup A \vdash O$
- A is subset-minimal over sets satisfying the first two conditions.

A *Simple Multi-Context Causal Theory* MT for a multi-context MC[3] is a family, $\{T_\alpha = \langle C_\alpha, E_\alpha, \Sigma_\alpha \rangle\}$ composed of a Simple Causal Theory T_α for each context α of MC. We introduce the extra hypothesis that for all α, $E_\alpha \subseteq C_\alpha$. This is because we accept the fact that each effect can be regarded as the explanation of itself.

We define

$$\mathbf{C} \stackrel{\triangle}{=} \{\alpha : \sigma \mid \sigma \in C_\alpha\}$$

$$\mathbf{\Sigma} \stackrel{\triangle}{=} \{\alpha : \sigma \mid \sigma \in \Sigma_\alpha\}$$

respectively, the *causes* and the *domain theory* of MT. An *Abductive Explanation* (ABE) for an observation $O \subseteq E_\alpha$ in a context α under the domain theory $\mathbf{\Sigma}$, is a set $\mathbf{A} \subseteq \mathbf{C}$, such that:

1. $\forall \beta,\ \mathbf{\Sigma} \cup \mathbf{A} \nvdash_{\mathsf{MC}} \beta : \bot$: $\mathbf{A}$ is consistent with $\mathbf{\Sigma}$ in any context.
2. $\mathbf{\Sigma} \cup \mathbf{A} \vdash_{\mathsf{MC}} \alpha : O$: the observation O can be derived in α from the set of axioms $\mathbf{A}$ and the domain theory $\mathbf{\Sigma}$.
3. $\mathbf{A}$ is the minimal set satisfying conditions 1 and 2. I.e., for any set $\mathbf{B} \subseteq \mathbf{C}$ satisfying condition 1 and 2, $\mathbf{B} \not\subset \mathbf{A}$.

From the decidability of MC (see [12]), and the fact that the set of possible explanations is finite, we can conclude that the problem of finding an ABE of an observation O under the domain theory $\mathbf{\Sigma}$ is decidable.

Example 1. Let us consider the situation where the domain theory is composed by the set of axiom $\mathbf{\Sigma} = \{(2)\text{–}(5)\}$, and the set of causes, effects, of the Simple Causal Theory of each context of MT is defined as follows:

- $E_{B_h} = E_{B_h B_s} = E_{B_h B_s B_h} = \{conditioning\_on, temp\_higher\_20°\}$
  beliefs and nested beliefs about temperature and about the conditioning are considered observable effects.
- $E_{I_h} = E_{B_h I_s} = \{conditioning\_on, stop\_working\}$
  intentions and beliefs about the speaker's intention regarding the conditioning and working are considered observable effects.

[3] In the rest of the paper, the reference to the multi-context MC is left implicit.

- For any other context α, $E_\alpha = \emptyset$
  we are not interested in effects in contexts different from the one mentioned above.
- $C_{B_h} = C_{B_h B_s} = C_{B_h B_s B_h} = \{\mathit{conditioning\_on}, \mathit{temp\_higher\_20°}\}$
  beliefs and nested beliefs about the temperature and about the conditioning are considered possible causes of the observable effects.
- $C_{I_h} = C_{B_h I_s} = \{\mathit{conditioning\_on}, \mathit{stop\_working}\}$
  beliefs and intentions regarding the speaker's intention regarding the conditioning and working are considered possible causes of the observable effects.
- For any other context α, $C_\alpha = \emptyset$
  we are not interested in causes in contexts different from the one mentioned above.

An ABE of the observation $I_h : \mathit{conditioning\_on}$ is $\{B_h B_s : \neg \mathit{conditioning\_on}\}$. Indeed we have that:

1. $\mathbf{\Sigma}, B_h B_s : \neg \mathit{conditioning\_on} \not\vdash_{\mathsf{MC}} \beta : \bot$, for any context β of MC;
2. $\mathbf{\Sigma}, B_h B_s : \neg \mathit{conditioning\_on} \vdash_{\mathsf{MC}} I_h : \mathit{conditioning\_on}$, see deduction in Figure 3;
3. $\{B_h B_s : \neg \mathit{conditioning\_on}\}$ is minimal.

An ABE of the observation $B_h I_s : \mathit{stop\_working}$ is

$$\{B_h B_s : \mathit{temp\_higher\_20°}, B_h B_s : \mathit{conditioning\_on}\} \tag{6}$$

Notice that $B_h B_s : \mathit{temp\_higher\_20°}$ can be derived from $B_h B_s : \neg \mathit{conditioning\_on}$ (by axiom (5); and therefore $B_h B_s : \mathit{temp\_higher\_20°}$ could be replaced by $B_h B_s : \neg \mathit{conditioning\_on}$ in (6) in order to obtain a second ABE for $B_h I_s : \mathit{stop\_working}$. On the other hand we cannot accept

$$\{B_h B_s : \neg \mathit{conditioning\_on}, B_h B_s : \mathit{conditioning\_on}\}$$

as an ABE since it violates the consistency condition.

Let us now describe how contextual abduction and revision can be exploited by the hearer to update its mental state. When the hearer receives a message, it can do a number of observations on the speaker's mental state; these observations concern the conditions that have induced s to send such a message. For instance, when h receives an INFORM from s, h can observe that the conditions I1–I4 on the speaker's mental state, described in Section 4, must hold. Such observations, however, are not unconditionally accepted by the hearer, rather it looks for a set of explanations for them, and only if it finds satisfactory explanations, it updates its mental state accordingly.

As described in Section 2, hearer's mental state can be represented by a set of axioms AX in a multi-context system MC . In order to explain observations deriving from communication, the hearer must be provided with a simple multi-context causal theory on MC . The domain theory of MC is always part of h's mental state, in symbols $\mathbf{\Sigma} \subseteq \mathsf{AX}$. The domain theory is never revised by the hearer. On the other hand AX contains other revisable belief and intentions, which can change along the dialogue. As argued above, these changes are the result of accepting the explanations of some observation. As a consequence the portion of AX which is not in $\mathbf{\Sigma}$ must be a subset of the causes; in symbols $\mathsf{AX} = \mathbf{\Sigma} \cup \mathbf{X}$, where $\mathbf{X} \subseteq \mathbf{C}$ is the only part that can be modified by the reception of the speech act (i.e., $\mathsf{AX}' = \mathbf{\Sigma} \cup \mathbf{X}'$). We call $\mathbf{X}$ the set of *current explanations*. $\mathbf{X}$ can be changed by applying three basic operations: *abductive expansion*, *abductive contraction*, and *abductive revision*.

Abductive Expansion Abductive Expansion, denoted by $+$, is applied when $\mathbf{X}$ must be extended to make possible to derive the observation ϕ in a context α:

$$\mathbf{X} + \alpha : \phi \stackrel{\triangle}{=} \mathbf{X} \cup \mathbf{A}$$

where $\mathbf{A}$ is an ABE of $\alpha : \phi$ under the domain theory $\mathbf{\Sigma} \cup \mathbf{X}$. Expansion does not specify how $\mathbf{A}$ is chosen among the set of minimal explanations. This choice might be based on a partial order on ABEs. In this paper we do not consider the effect and the specific definition of such an order. This order strictly depends on the application domain and on the meaning and the degree of plausibility of the explanations. Similarly we do not consider methods to represent and compute such a partial order.

Abductive Contraction The operation of contraction, denoted by $-$, is applied if $\mathbf{X}$ is in contrast with a new observation. So, some formulas must be removed from $\mathbf{X}$. For any formula ϕ in a context α:

$$\mathbf{X} - \alpha : \phi \stackrel{\triangle}{=} \mathbf{X} \setminus \mathbf{Y}$$

where $\mathbf{Y}$ is a minimal hitting set[4] of $\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$, where $\{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$ is the set of all the ABEs of ϕ in the context α under the domain theory $\mathbf{\Sigma}$, which are contained in $\mathbf{X}$. Again we have not specified how the hitting set is chosen in all the $\mathbf{A}_i$. As before, this will be related to the ordering on the ABEs.

Abductive Revision The operation of revision, denoted by $*$, must be performed when $\mathbf{X}$ explains something which is inconsistent with the observation. For any formula ϕ in a context α, we define the operator:

$$\mathbf{X} * \alpha : \phi = (\mathbf{X} - \alpha : \neg\phi) + \alpha : \phi$$

This operative definition comes from the Levi identity (see [6, 5]).

6 Update from INFORM

Checking preconditions (Condition I4) Let us suppose that s performs $\mathrm{INFORM}(s, h, \phi)$. Being aware of the fact that condition **I4** holds, h may update its image of the speaker's mental state by imposing that the precondition of $\mathrm{INFORM}(s, h, \phi)$ holds on its images of speaker's beliefs[5]:

$$\phi \in B_h {B'_s}^*$$

The new mental state is obtained by updating the set of current explanations $\mathbf{X}$ to $\mathbf{X}'$ as follows:

$$\mathbf{X}' = \mathbf{X} * B_h B_s : \phi$$

We can distinguish two cases: either $\mathbf{\Sigma} \cup \mathbf{X}$ is consistent with $B_h B_s : \phi$, (i.e., $\neg\phi \notin B_h B_s^*$) or it is inconsistent with $B_h B_s : \phi$ (i.e.,$\neg\phi \in B_h B_s^*$). In the first case h just expands $\mathbf{X}$ with an explanation of $B_h B_s : \phi$; in the second case, h computes $\mathbf{X}'$ first by contracting $\mathbf{X}$, in order to have a new set $\mathbf{Y} = \mathbf{X} - B_h B_s : \neg\phi$, such that $\mathbf{\Sigma} \cup \mathbf{Y} \not\vdash_{\mathsf{MC}} B_h B_s : \neg\phi$, and then expands $\mathbf{Y}$ into a set $\mathbf{X}' = \mathbf{Y} + B_h B_s : \phi$, by adding an explanation of $B_h B_s : \phi$. In the resulting mental state we have, therefore that $\phi \in B_h B_s'^*$.

[4] Given a collection of sets $S = \{\mathbf{A}_1, \ldots, \mathbf{A}_n\}$, a set $\mathbf{H}$ is a hitting set of S if for each $\mathbf{A}_i$, $\mathbf{H} \cap \mathbf{A}_i \neq \emptyset$. A hitting set is minimal if for any other hitting set $\mathbf{H}'$, $\mathbf{H}' \not\subseteq \mathbf{H}$.

[5] Notice that this coincides also with the main effects of $INFORM(s, h, \phi)$ on the mental state of h.

Example 2. Suppose that s performs $\mathrm{INFORM}(s, h, \mathit{temp\_higher\_20°})$ when the current explanations $\mathbf{X} = \emptyset$. It is easy to see that $\mathbf{X}' = \mathbf{X} * B_h B_s : \mathit{temp\_higher\_20°} = \{B_h B_s : \neg \mathit{conditioning\_on}\}$. Indeed, since $\neg \mathit{temp\_higher\_20°} \notin B_h B_s^*$ the hearer computes the following minimal ABE:

$$\mathbf{A} = \{B_h B_s : \neg \mathit{conditioning\_on}\}$$

and expands $\mathbf{X}$ accordingly, resulting in $\mathbf{X}' = \{B_h B_s : \neg \mathit{conditioning\_on}\}$. Notice that, in this new mental state, h has the intention to switch the conditioning on, as we have $\mathbf{\Sigma} \cup \mathbf{X}' \vdash_{\mathsf{MC}} I_h : \mathit{conditioning\_on}$.

Intention recognition (condition I1) By intention recognition we mean the hearer's ability to recognize the intention that caused the speaker to perform the speech act. Condition (**I1**) states that a motivation for s to perform an $\mathrm{INFORM}(s, h, \phi)$ is its intention to change h's mental state so that h believes or intends some new formula. To discover this intention, h checks the differences between its mental state before and after s executes $\mathrm{INFORM}(s, h, \phi)$ ($\mathbf{X}$ and $\mathbf{X}'$ respectively) and then it extends $\mathbf{X}'$ to include the fact that s has the intention of causing those differences. For instance suppose that for the context $B_h B_s$[6] there is formula ψ such that:

$$\psi \notin B_h^* \quad \text{and} \quad \psi \in B_h'^* \tag{7}$$

This intuitively means that one of the effect that the speaker has obtained by its utterance is that, the hearer believes ψ. Therefore the hearer might suppose that this was an intention of the speaker. Namely the hearer makes the observation $B_h I_s B_h : \phi$. Then h revises $\mathbf{X}'$ by the observation $B_h I_s B_h : \psi$, i.e. h tries to find an explanation of the fact that the speaker has the intention to make itself to believe ψ. To do this h revises $\mathbf{X}'$ to obtain a new set of explanations $\mathbf{X}''$ defined as follows:

$$\mathbf{X}'' = \mathbf{X}' * B_h I_s B_h : \psi$$

Similar revision can be done on any other context of the hearer's belief state.

Example 3. Let us consider our example restricting the recognition problem to the possible speaker's intentions regarding hearer's beliefs and intentions. Suppose that s performs $\mathrm{INFORM}(s, h, \mathit{temp\_higher\_20°})$ and $\mathbf{X}'$ is computed as in Example 2. We have that h finds that $\mathit{temp\_higher\_20°}$ is not in B_h^* but it is in $B_h^{'*}$. So h revises $\mathbf{X}^{'}$ as follows:

$$\mathbf{X}'' = \mathbf{X}^{'} * B_h I_s B_h : \mathit{temp\_higher\_20°}$$

Moreover h finds that $\mathit{conditioning\_on}$ is a formula that does not belong to $\mathbf{X}$, but belongs to $\mathbf{X}'$. As before h revises $\mathbf{X}^{'}$ as follows:

$$\mathbf{X}'' = \mathbf{X}' * B_h I_s I_h : \mathit{conditioning\_on}$$

This means that h believes that s has the intention of making h believe that $\mathit{temp\_higher\_20°}$ and of making h intend to perform $\mathit{conditioning\_on}$.

A final update (Condition I3) A further observation that the hearer can do as a consequence of itself getting to believe ϕ (see condition (7)) is that now s is in a mental stete in which s believes that its intention has been satisfied; namely that s believes that h believes ψ. (conditioning **I3**). As a consequence h can can expand $\mathbf{X}''$ in order to verify that $\psi \in B_h B_s B_h^*$ as follows:

$$\mathbf{X}''' = \mathbf{X}'' * B_h B_s B_h : \psi$$

[6] h could be interested only in the effects yielded in some particular contexts.

7 Related Work

The work presented in this paper relates with four main research areas: agent's internal structure, agent communication language, abduction, and belief revision. Concerning agent internal models, we based our approach on previous work [12, 1, 4].

Concerning the semantics of Agent Communication Languages, there are a number of proposal of a speech act based semantics for KQML and FIPA ACL (the two main agent communication Languages), however, all these languages do not specify how this semantics must be used by the agents involved in a dialog. Differently, in our approach we have defined a concrete and computationally feasible way for an agent to treat a specific set of communicative acts. In other words, given a set of communicative acts with e semantics expressed in terms of preconditions and main effects, we have defined a set of revision policies that an agent can follow to update its mental state whenever it receives a message.

In spite of the two well grounded research areas on "belief revision" and on "abduction", few works attempted to combine them in a unified treatment. In [16], Lobo and Uzcátegui define an abductive version of a large class of theory change operators. For any operator $*$ they define its abductive version $*_a$. Lobo's and Uzcátegui's objective is to define general abductive change operators (on the basis of existing theory change operators), while our work has the main goal to define a *specific* set of theory change operators suitable for observations generated by a communicative act.

Aravidian and Dung, in [2], state a number of rationality postulates for the contraction of knowledge basew.r.t. a sentence, and they define an abduction based algorithm for its computation. Their algorithm (based on hitting sets) is very closed to the definition of abductive contraction given in this paper. As a matter of facts, our definition of contraction fulfills their basic rationality postulates. A second important analogy is the fact they suppose that the knowledge base is composed of two subset: an "immutable theory" and an "updatable theory" which are the analogous to $\mathbf{\Sigma}$ and $\mathbf{X}$ defined in this paper. The main difference is that we extend this idea to the case of abductive extension and revision, and we have specialized the operators to a logic for belief and intentions.

Analogously, Pagnucco et al. [19, 20] introduce some rationality postulates for abductive expansion and, [18] argue that the notion of abduction corresponds to an attempt to determine an initial belief state from a contracted belief state and an epistemic input under certain conditions. In all these works the revision process is limited to the agent's beliefs. Introducing mental states, we extend the revision process to the agent's intentions. In multi-agent systems this is very important because it allows an agent to revise its intentions whenever new beliefs are acquired. Moreover, the use of image of mental states allows to maintain its beliefs about the mental state of the other agents always updated.

Hindriks et al. in [14] provide an operational semantics for two pairs of communicative operators, ask and tell, and request and offer, based on transition systems. As in our approach, in their approach each agent contains a mental state, which is composed of two sets: beliefs and goals. In addition, each mental state contains a set of rules describing its evolution. The mental state of an agent refers to a particular moment during the agent evolution. The semantics of communicative operators is given in terms of a transition of an agent from a belief state to another. A first difference between their and our approachs is that they do not allow agents to have images of the others' mental states. A more radical difference regards the fact that, in their semantics, communicative acts do not have either preconditions or effects on the beliefs of an agent. In other words, the semantics does not contain an explicit relationship between the agent's beliefs and its communications. The main consequence of this choice is that the reception of a message does not necessarily yield the revision of the agent's mental state. Belief revision is considered as any other action. An agent can decide to revise its beliefs independently from the communication with other agents. Moreover, [14] does not provide any criteria for belief revision, which is assumed to be a pre-compiled function. Our proposal is therefore complementary to this approach as we provide a well founded and executable methods for revising beliefs after communication. Finally, Hindriks et al. deal with synchronous communication while we consider asynchronous communication. Synchronous communication means that a tell from agent A to agent B is necessarily preceded by an ask from agent B to agent A, and vice-versa any ask from B to A is followed by a tell in the opposite direction. This is true also for the other pair of communicative acts, namely $\mathsf{request}$ and offer. Deduction and abduction are not used to enlarge and/or modify agents beliefs. Rather, in our case deduction is a reasoning pattern which is used by an agent to generate an answer (tell) of an information request (ask); similarly abduction is used to generate an offers (offer) that fulfill a request ($\mathsf{request}$). Differently in our approach, which considers asynchronous communication, deduction and abduction are performed not only to generate the answer to a specific request, but to process the reception and the sending of any message.

8 Conclusion

In this paper we have made the fundamental assumption that there is a causal relationship between a speaker's mental state and its uttering a sentence. Following this idea, we have developed some abductive methods to recognize the speaker's mental state and then to update the hearer's image of the speaker's mental state. Inspiration for the causal relationship has been taken from the classical "Speech Acts Theory", namely we have adopted the plan-based vision of speech acts in representing them as STRIPS-like operators. We have investigated how it is possible to use both the preconditions and the effects of the speech act in order to update the addressee's image of the speaker's belief and intentions. Our work is based upon the use of multi-context systems for which we have extended the notion of casual theories, abduction and revision. The definition of an efficient and implementable algorithm that computes the set of all abductive explanations for an observation will be the object of our future work.

References

1. M. Benerecetti, F. Giunchiglia, and L. Serafini. Model Checking Multiagent Systems. *Journal of Logic and Computation, Special Issue on Computational & Logical Aspects of Multi-Agent Systems*, 8(3):401–423, 1998. Also IRST-Technical Report 9708-07, IRST, Trento, Italy.
2. S. Carberry and W. A. Pope. Knowledge revision, abduction and database updates. *Journal of Applied Nonclassical Logics*, 1995.
3. A. Cimatti and L. Serafini. Multi-Agent Reasoning with Belief Contexts II: Elaboration Tolerance. In *Proc. 1st Int. Conference on Multi-Agent Systems (ICMAS-95)*, pages 57–64, 1996. Also IRST-Technical Report 9412-09, IRST, Trento, Italy. *Commonsense-96*, Third Symposium on Logical Formalizations of Commonsense Reasoning, Stanford University, 1996.
4. P.R. Cohen and C.R. Perrault. Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212, 1979.
5. A.F. Dragoni and P. Giorgini. Revising beliefs received from multiple source. In M A Williams and H Rott, editors, *Frontiers of Belief Revision*, Applied Logic. Kluwer, 1999.
6. P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press., Cambridge Mass., 1988.
7. M.P. Georgeff. Communication and interaction in multiagent planning. In *AAAI 83*, pages 125–129, 1983.

8. C. Ghidini. Modelling (Un)Bounded Beliefs. In P. Bouquet, L. Serafini, P. Brezillon, M. Benerecetti, and F. Castellani, editors, *Modelling and Using Context – Proceedings of the 2nd International and Interdisciplinary Conference, Context'99*, volume 1688 of *Lecture Notes in Artificial Intelligence*, pages 145–158. Springer Verlag - Heidelberg, 1999.
9. E. Giunchiglia and F. Giunchiglia. Ideal and Real Belief about Belief. *Journal of Logic and Computation*, 2000. To appear, Also IRST-Technical Report 9605-04, IRST, Trento, Italy.
10. F. Giunchiglia. Contextual reasoning. *Epistemologia, special issue on I Linguaggi e le Macchine*, XVI:345–364, 1993. Short version in Proceedings IJCAI'93 Workshop on Using Knowledge in its Context, Chambery, France, 1993, pp. 39–49. Also IRST-Technical Report 9211-20, IRST, Trento, Italy.
11. F. Giunchiglia and C. Ghidini. Local Models Semantics, or Contextual Reasoning = Locality + Compatibility. In *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 282–289. Morgan Kaufmann, 1998. Also IRST-Technical Report 9701-07, IRST, Trento, Italy.
12. F. Giunchiglia and L. Serafini. Multilanguage hierarchical logics (or: how we can do without modal logics). *Artificial Intelligence*, 65:29–70, 1994. Also IRST-Technical Report 9110-07, IRST, Trento, Italy.
13. Piotr J. Gmytrasiewicz and Edmund H. Durfe. A rigorous, operational formalization of recursive modeling. In *Proc. of the First International Conference on Multi-Agent Systems (ICMAS)*, pages 125–132, 1995.
14. Koen V. Hindriks, Frank S. de Boer, W. van der Hoek, and J.-J.Ch. Meyer. Semantics of communicating agents based on deduction and abduction. In *Proceedings of IJCAI'99 Workshop on ACL*, 1999.
15. K. Konolige. Abduction versus closure in causal theories. *Artificial Intelligence*, 53:255–272, 1992.
16. Jorge Lobo and Carlos Uzcátegui. Abductive change oprators. *Fundamenta Informaticae*, 27(4):319–418, 1996.
17. J.S. Rosenschein M.R. Gensereth, M.L. Ginsberg. Cooperation without communication. In *AAAI 86*, pages 51–57, 1986.
18. M. Pagnucco and N.Y. Foo. The relationship between abduction and changes in belief states. In *Proceedings of the ICLP93 Postconference Workshop on Abductive Reasoning*, pages 75–83, Budapest, Hungary, 1993.
19. M. Pagnucco, A.C. Nayak, and N.Y. Foo. Abductive expansion: The application of abductive inference to the process of belief change. In *Proceedings of the Seventh Australian Joint Conference on Artificial Intelligence*, pages 70–77, Armidale, Australia, 1994.
20. M. Pagnucco, A.C. Nayak, and N.Y. Foo. Abductive reasoning, belief expansion and nonmonotonic consequence. In *Proceedings of the ICLP'95 Joint Workshop on Deductive Databases and Logic Programming and Abduction in Deductive Databases and Knowledge-based Systems*, pages 143–158, Shonan Village Center, Japan, 1995.
21. A.S. Rao and M. Georgeff. Bdi agents: from theory to practice. In *Proc. of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 312–319, S. Francisco, CA, 1995.
22. A.S. Rao, M. Georgeff, and E.A. Sonenberg. Social plans: A preliminary report. In E. Werner and Y. Demazeau, editors, *Decentralized AI - Proc. of the Third European Workshop on Modeling Autonomous Agents in a Multi-Agent World (MAAMAW-91)*, pages 57–76, Amsterdam, The Netherlands, 1992. Elsevier Science Publishers B.V.
23. J.S. Rosenschein and M.R. Gensereth. Communication and cooperation. *Stanford Heuristic Programming Rep*, 1984.
24. J. R. Serale and D. Vanderveken. *Foundations of illuctionary Logic*. Cambridge University Press., 1985.
25. E. Werner. Toward a theory of communication and cooperation for multiagent planning. In *Proc. of TARK 88*, Los Altos, CA, 1988. Morgan Kaufmann Publisher.