

# Establishing Mutual Beliefs by Joint Attention: Towards a Formal Model of Public Events

**Emiliano Lorini (emiliano.lorini@istc.cnr.it)**

Institute of Cognitive Sciences and Technologies, Viale Marx 15  
00137, Roma, ITALY

**Luca Tummolini (luca.tummolini@istc.cnr.it)**

Institute of Cognitive Sciences and Technologies, Viale Marx 15  
00137, Roma, ITALY

**Andreas Herzig (herzig@irit.fr)**

Institut de Recherche en Informatique de Toulouse, 118 Route de Narbonne  
F-3106, Toulouse, Cedex 4, FRANCE

## Abstract

While the role of mutual beliefs in coordination and collaboration has been extensively acknowledged, the cognitive processes supporting their establishment are left unexplained or simply assumed. Notions like “public event” or “public announcement” usually refer to events or speech acts that create such mutual information states. The goal of this paper is to provide a formal model of the conditions under which mutual beliefs can be established. Agents should be able to perceive and reason about each other epistemic activities in a shared world. To express such reasoning a simple version of propositional dynamic logic with converse operator (CPDL) is adopted.

**Keywords:** Mutual Belief achievement; Common Knowledge; Joint Attention.

## Introduction

The notion of *common* or *mutual belief* is a widespread interpretive concept shared by many diverse disciplines<sup>1</sup>. Since the seminal work of Lewis (1969), it has been widely adopted as a crucial notion to explain coordination in a variety of social settings from discourse understanding and definite reference (Clark & Marshall 1981), to strategic reasoning in game theory (Bacharach 1992; Geanakoplos 1992), to collaborative and group activity in AI (Grosz & Kraus 1996). To act effectively in these situations, it is not enough for a group of agents that they all believe something; they should also have attitudes towards the mental states of their peers. Consequently, the problem of the genesis of common belief is of fundamental importance for modeling social interaction between cognitive agents. Such genesis is often considered as related either to *public events* or to *public announcements*. In the former situation a common belief is a consequence of an event whose occurrence is so evident (*viz.* public) that agents cannot but recognize it as when, during a soccer match, players mutually believe that they are playing soccer. In the latter, common belief is the product of a special event that is a

communication process as when the referee publicly announces that one player is expelled. From there on each player believes that each other player believes and so on... that one of them has been expelled.

Intuitively, an event is considered “public” as long as its occurrence is epistemically accessible by everybody such that it becomes common knowledge between them. Such a definition is usually given for granted but can be explicitly stated as:  $Public(e) \leftrightarrow (Happens(e) \rightarrow CB(Happens(e)))$ .

However what are the “intuitive” conditions that make an event to be qualified as public? What are the reasons to believe that an occurring event is commonly believed?

To achieve a common belief, agents need to be aware of each other current epistemic activities (attending to, looking at) both at the event itself and at each other epistemic activities. Looking at each other (i.e. by eye contact) while accessing the event provides reasons to accept the mutual information state. Such condition is usually described as *joint attentional state* (see for instance Tomasello 1999).

Moving beyond traditional approaches to public announcements mostly focused on belief update at the group level (Baltag *et al.* 2003), in this paper first we introduce a propositional dynamic logic to reason about epistemic and pragmatic actions and beliefs. Then we advance a logic for perception and mutual perception that let the agents infer from the fact that they are jointly attending at something (i.e. by mutually checking whether something is true in the world) that a certain proposition is mutually believed by them. After discussing our model we conclude and point to future work.

## Language

Let denumerable sets  $AGT = \{1, \dots, n\}$  of agents,  $\Pi$  of propositional symbols  $\{p, q, r, \dots\}$  and  $ACT$  of atomic pragmatic actions  $\{a, b, c, \dots\}$  be given. The language  $L$  is the smallest superset of  $\Pi$  such that: if  $\varphi, \psi \in L$ ,  $i \in AGT$ ,  $\alpha \in ACT'$  then  $\neg\varphi$ ,  $\varphi \vee \psi$ ,  $Bel_i\varphi$ ,  $\langle\alpha\rangle_i\varphi$ ,  $\langle O(\varphi)\rangle_i\psi$ ,  $\langle P(\varphi)\rangle_i\psi$ ,  $\langle T(\varphi)\rangle_i\psi \in L$  where  $ACT'$  is the smallest superset of  $ACT$  such that if  $\varphi \in L$  and  $\alpha_1, \alpha_2 \in ACT'$  then

<sup>1</sup> While knowledge is generally considered as justified true belief in this paper we will adopt only the weaker belief mental attitude.

- $\alpha_1; \alpha_2 \in \text{ACT}$  “execute  $\alpha_1$  then execute  $\alpha_2$ ”
- $\alpha_1 \cup \alpha_2 \in \text{ACT}$  “choose either  $\alpha_1$  or  $\alpha_2$  nondeterministically and execute it”

### Beliefs

We use a modal logic KD45 as the logic for belief, i.e. an agent does not entertain inconsistent beliefs and is aware of his beliefs and disbeliefs. In the models for each agent  $i \in \text{AGT}$  and possible world  $w \in W$  there is an associated set of possible worlds  $B_i(w) \subseteq W$  where  $B$  is a mapping function  $B: W \times \text{AGT} \rightarrow 2^W$ . So for every agent in  $\text{AGT}$  there is a modal operator  $\text{Bel}_i$  and  $\text{Bel}_i \varphi$  expressing that agent  $i$  believes that  $\varphi$ . The truth condition for  $\text{Bel}_i$  says that  $M, w \models \text{Bel}_i \varphi$  if and only if  $\varphi$  holds in all worlds that are compatible with agent  $i$ 's beliefs, i.e.  $M, w \models \text{Bel}_i \varphi$  if and only if  $M, v \models \varphi$  for every  $v \in B_i(w)$ . So,  $B_i$  is an accessibility relation that is serial, transitive and euclidean.

### Pragmatic Actions

We use a simple version of CPDL (Propositional dynamic logic with converse operator) for modeling the action component. The empty action is noted  $\emptyset$ . To each action  $a$  is associated the modal operator  $[a]$  and its dual operator  $\langle a \rangle$ . The formula  $[a] \varphi$  reads “always  $\varphi$  is true after action  $a$  executed by agent  $i$ ” whereas  $\langle a \rangle \varphi$  reads “possibly  $\varphi$  is true after action  $a$  executed by agent  $i$ ” ( $\neg[a] \neg \varphi = \langle a \rangle \varphi$ ).  $[a] \text{false}$  expresses that  $a$  cannot be executed by agent  $i$ . On the other side  $\langle a \rangle \text{true}$  expresses that  $a$  can be executed by agent  $i$ . In the models for each agent  $i \in \text{AGT}$  and possible world  $w \in W$  and pragmatic action  $a \in \text{ACT}$  there is an associated set of possible worlds  $r_0(i, a)(w) \subseteq W$  where  $r_0$  is a mapping function  $r_0: W \times \text{AGT} \times \text{ACT} \rightarrow 2^W$ . The truth condition for  $[a] \varphi$  says that  $M, w \models [a] \varphi$  if and only if  $\varphi$  holds in all worlds  $w'$  that are results of action  $a$  executed by agent  $i$  when applied in world  $w$ <sup>2</sup>, i.e.  $M, w \models [a] \varphi$  if and only if  $M, v \models \varphi$  for every  $v \in r_0(w, i, a)(w)$ .

We adopt the standard axiomatic of PDL (see Harel *et al.* 2000) that for our fragment is nothing but the multimodal logic  $K$ . We use the converse operator  $\bar{\phantom{x}}$ : a program operator that allows a program to be “run backwards”. The mapping function  $r_0$  is extended to converse action by stipulating that  $r_0(w, i, a\bar{\phantom{x}}) = r_0(w, i, a)$ <sup>3</sup>. Notice that standard PDL already provides something that is similar to an Observe-action (defined below): the standard test operator  $?$ <sup>4</sup>. That operator is of no use here, because it is defined as  $[\varphi?] \psi \leftrightarrow (\varphi \rightarrow \psi)$ .

<sup>2</sup> Notice that the semantics in dynamic logic for complex programs is easily defined in an inductive way. Take for instance the semantic for sequential composition of programs  $M, w \models [\alpha_1; \alpha_2] \varphi$  if and only if  $M, v \models \varphi$  for every  $v \in r_0(i, \alpha_2)(r_0(i, \alpha_1)(w))$ .

<sup>3</sup> In CPDL two axioms for the converse operator are added to the standard PDL axiomatic:  $\varphi \rightarrow [a] \langle a \bar{\phantom{x}} \rangle \varphi$ ;  $\varphi \rightarrow [a \bar{\phantom{x}}] \langle a \rangle \varphi$ .

<sup>4</sup> In standard PDL  $\psi?$  means “if  $\psi$  is true proceeds with the program otherwise fail”.

### Perceptual Actions

Beyond usual pragmatic actions, for the aim of this paper we need to represent also a specific kind of *epistemic* action. More generally, we define epistemic actions as *actions that are specialized for epistemic results: i.e. for the acquisition of knowledge or the verification (confirmation) of beliefs* (Lorini & Castelfranchi 2004). *Pure* epistemic actions are actions that *never* change the state of the world; they just change the knowledge of the agent<sup>5</sup>. Here we are interested only to epistemic Actions of perceptual kind: *Observe that*  $\varphi$  ( $O(\varphi)$ ), *Perceive that*  $\varphi$  ( $P(\varphi)$ ), *Test-if*  $\varphi$  ( $T(\varphi)$ ). We assume that *Observe that*, *Perceive that* and *Test-if* are applicable to all formulas defined in our language. An agent can *Observe that* ( $\cdot$ ), *Perceive that* ( $\cdot$ ) and *Test-if* ( $\cdot$ ) either state of affairs (propositions) or the execution of actions. In principle it is not possible for an agent to perceive or to observe or to test his own beliefs. More generally, perception of mental states is not admitted in the present formalism. We define:

- *Perceive that*  $\varphi$  the action of perceiving some  $\varphi$  in the external world. *Perceive that*  $\varphi$  not necessarily implies that what the agent (for instance agent  $i$ ) actually perceives (in this case  $\varphi$ ) is true in the external world.

- *Observe that*  $\varphi$  is the action of observing some  $\varphi$  in the external world. *Observe that*  $\varphi$  is the action executed by agent  $i$  of perceiving something “correctly” (of perceiving something that is true in the external world).

An *Observe that* action as well as a *Perceive that* action has the following property: an agent  $i$ 's *Observe that*  $\varphi$  action (*Perceive that*  $\varphi$  action) cannot be perceived, observed or tested by another agent  $j$  nor from agent  $i$  himself.

- *Test-if*  $\varphi$  is the action of testing whether  $\varphi$  is true or not<sup>6</sup>. A *Test-if*  $\varphi$  action is always the precursor of either a *Perceive that*  $\varphi$  or a *Perceive that*  $\neg \varphi$  action and under some particular conditions (when the environment is not noisy) is the precursor of either an *Observe that*  $\varphi$  action or an *Observe that*  $\neg \varphi$  action. Moreover, a *Test-if*  $\varphi$  action has the property of being perceivable, observable and testable from the executor and from other agents (different from the executor). Assuming that *Test-if* actions can be perceived can be questionable. Indeed it could seem more reasonable to consider *Test-if* actions as mental processes, without the assumption that they can be perceived from the external environment. From a cognitive point of view an agent perceptually tests a proposition that  $p$  if and only if he is matching the *perceptual schema* of the object designated in  $p$  with the sensorial stimuli and no inferential process is involved in the process of verification. However our present formalism and analysis applies to the scenarios in which the

<sup>5</sup> Differently *parasitic* epistemic actions exploit pragmatic actions to achieve higher order epistemic goals and thus necessarily change the state of the world (see for example Kirsh & Maglio 1994).

<sup>6</sup> In the present analysis the distinction between Test-whether and Test-what action is not considered (see Harrah 2002 for the distinction between What-question and Whether-question in the logic of Questioning).

testing mental process has an *external counterpart* that is perceivable from the external environment (i.e. by detecting overt attentional shifts that can be seen as signals of this internal activity). We assume that the external counterpart of the *Test-if* action does not affect the dynamics of the environment<sup>7</sup>. As for pragmatic actions also epistemic actions of the perceptual kind are associated with the modal operator  $[.]$  and its dual operator  $\langle \cdot \rangle$ . For example, the formula  $[P(\varphi)]_i \psi$  reads “always  $\psi$  is true after agent  $i$  has perceived that  $\varphi$ ”.

**Axiomatic for Perceptual Actions and definition of Test if actions.** In the rest of the paper we will omit a logical analysis of Observation actions (correct perceptions) and we will strictly focus on Perception and Test-if actions. The following logical axioms characterize perceptual actions:

AXIOM 1.  $\psi \rightarrow [P(\varphi)]_i \psi$  if  $\psi$  is objective

AXIOM 2.  $\langle P(\varphi) \rangle_i \psi \rightarrow [P(\varphi)]_i \psi$

Axiom 1 says that perceptual actions are actions that do not change the physical environment (if some  $\psi$  is true before the perceptual action and  $\psi$  is an objective state of the world then  $\psi$  is still true after that agent  $i$  has perceived that  $\varphi$ ). Axiom 2 says that perceptual actions are deterministic (if possibly after that agent  $i$  perceives that  $\varphi$ ,  $\psi$  is true, then always after that agent  $i$  perceives that  $\varphi$ ,  $\psi$  is true).

**Introspection axiom for perception (IP)**

$[P(\varphi)]_i \text{Bel}_i \langle P(\varphi) \rangle_i \text{true}$

IAP says that always after that an agent  $i$  perceives some  $\varphi$ , he ends up believing to have perceived  $\varphi$ <sup>8</sup>.

**Axiom of Restricted Perception (RP)**

RP1.  $[P(\langle P(\varphi) \rangle_i \text{true})]_i \text{false}$

RP2.  $[P(\text{Bel}_j \varphi)]_i \text{false}$

RP1 and RP2 respectively say that an agent  $i$  can never perceive what agent  $j$  (or agent  $i$  himself) perceives or what agent  $j$  (or agent  $i$  himself) believes.

**Definition of Noise/Not Noise and Test if actions**

We introduce the formula  $\text{Noise}(i, \varphi)$  which is built on the *special* binary predicate  $\text{Noise}(\cdot)$  that relates agents to formulas and denotes a disjunction of all noise conditions for agent  $i$ 's perception of formula  $\varphi$  ( $\text{Noise}(i, \varphi) = \psi_1 \vee \dots \vee \psi_n$ )<sup>9</sup>. Moreover, we postulate the existence of a *Noise axiom* for each for agent  $i$ 's perception of formula  $\varphi$  of the following form.

**Noise axiom (NA)**<sup>10</sup>

$\langle P(\varphi) \rangle_i \text{true} \wedge \neg \varphi \rightarrow \text{Noise}(i, \varphi)$ <sup>11</sup>

This law establishes that whenever an agent has perceived that  $\varphi$  and  $\varphi$  is actually false then it means that there is Noise for agent  $i$ 's perception of formula  $\varphi$ .

**Definition of Test if action**

$\langle T(\varphi) \rangle_i \text{true} =$

$(\varphi \wedge \neg \text{Noise}(i, \varphi) \rightarrow \langle P(\varphi) \rangle_i \text{true}) \wedge$

$(\varphi \wedge \text{Noise}(i, \varphi) \rightarrow \langle P(\neg \varphi) \rangle_i \text{true}) \wedge$

$(\neg \varphi \wedge \neg \text{Noise}(i, \neg \varphi) \rightarrow \langle P(\neg \varphi) \rangle_i \text{true}) \wedge$

$(\neg \varphi \wedge \text{Noise}(i, \neg \varphi) \rightarrow \langle P(\varphi) \rangle_i \text{true})$

Informally, an agent  $i$ 's *Test if*  $\varphi$  action is defined as the nondeterministic choice between having perceived  $\varphi$  or perceived  $\neg \varphi$  under some conditions of Noise with respect to  $\varphi$  (and  $\neg \varphi$ ) and the actual truth value of  $\varphi$ <sup>12</sup>.

**Proposition 1**<sup>13</sup>

$\langle P(\varphi) \rangle_i \text{true} \wedge \text{Bel}_i (\langle P(\varphi) \rangle_i \text{true} \rightarrow \varphi) \rightarrow \text{Bel}_i \varphi$

Proposition 1 says that whenever an agent  $i$  has perceived that  $\varphi$  and believes that if he has perceived  $\varphi$ ,  $\varphi$  is effectively true then the agent believes that  $\varphi$ <sup>14</sup>.

## Mutual Social Perception

### The problem of Background Expectations

As previously remarked (see Axioms RP1 and RP2), an agent  $i$  can never perceive (or observe) what another agent  $j$  perceives (or observes) while can perceive (and observe) that another agent  $j$  is testing *something*, since *Test-if*

---

Amplificatory Noise (not perceiving what we are actually testing versus perceiving something that is not true in the external world). Moreover we do not take into account the notion of Deforming Noise (given a test for verifying whether  $\varphi$  or  $\psi$  is true, if the  $\varphi$  is true the agent perceives  $\psi$ ).

<sup>11</sup> Notice that this axiom is equivalent to the *Not-Noise axiom*  $\neg \text{Noise}(i, \varphi) \rightarrow (\langle P(\varphi) \rangle_i \text{true} \rightarrow \varphi)$ .

<sup>12</sup> The expression ‘An agent  $i$  has tested if  $\varphi$ ’ is equivalent to: if  $\varphi$  is true and there is not noise for agent  $i$  with respect to  $\varphi$  then agent  $i$  has perceived  $\varphi$  and, if  $\varphi$  is true and there is noise for agent  $i$  with respect to  $\varphi$  then agent  $i$  has perceived  $\neg \varphi$  and, if  $\neg \varphi$  is true and there is not noise for agent  $i$  with respect to  $\neg \varphi$  then agent  $i$  has perceived  $\neg \varphi$  and, if  $\neg \varphi$  is true and there is noise for agent  $i$  with respect to  $\neg \varphi$  then agent  $i$  has perceived  $\varphi$ .

<sup>13</sup> The proposition can be derived by means of standard modal principles of Beliefs and the Introspection Axiom for perception given above.

<sup>14</sup> The present model underlines a Belief Update Semantic that is close to the semantic given in Herzig & Longin (2000). The main idea of this belief update semantic is the following: whenever a given agent  $i$  has perceived that a certain action  $a$  will possibly happen next and the agent believes that his own perception was not affected by Noise, the agent updates his belief base; he “mentally” simulates the execution of action  $a$  in his Belief Model, discarding all the mental simulation-resulting states that are not connected by action  $a$  and accepting all the mental simulation-resulting states that are connected by action  $a$ .

<sup>7</sup> A more fine-grained formalism should distinguish between the internal counterpart of a Test Action and the external one.

<sup>8</sup> See Del Val *et al.* 1997 for a similar axiom in a logic of perception.

<sup>9</sup> An alternative solution to deal with the *Noise* notion is based on the definition of Noise as a modal operator (based on the normal K-system of modal logic). This second solution is more appropriate for a semantic analysis of the *Noise axiom*. We leave this issue to further developments of the model.

<sup>10</sup> For the sake of simplicity, we condensate in the same definition the notion of Deleting Noise (Obstacle) and the notion of

actions are assumed as the interface between external and internal world.

However, given that an agent  $i$  can perceive that another agent  $j$  is testing if *something* is true or not, how can he perceive *what* agent  $j$  is actually testing? In fact, while *agent  $j$ 's Test-if action can be perceived, its "object" cannot.*

We argue that for an agent  $i$  to perceive *what* the proposition-object of agent  $j$ 's test-if action is, two kinds of conditions are required:

1. *Indexical cues.* Agent  $i$  perceives that agent  $j$ 's sensors are directed towards a specific region of the space  $S$  where several objects  $O_1, \dots, O_n$  are located.

2. *Background expectations in agent  $i$ 's mind.* Once agent  $i$  has individuated the spatial region  $S$ , some background expectations are necessary for orienting the perception of agent  $i$  towards the predicted object of agent  $j$ 's test, for making agent  $i$  able to establish which object among  $O_1, \dots, O_n$  is pointed by agent  $j$ . Assuming that the reference object is a propositional atom  $p$ , we can state that: if an agent  $i$  perceives that another agent  $j$  is testing if  $p$  is true or not then agent  $i$  had some previous background expectation that drove his own perception.

The following statement makes explicit the kind of background expectation that could be involved in the perception of other agent's tests:

#### Background Expectation Axiom

$[P(\langle T(p) \rangle; \text{true})^-]_i \text{Bel}_i \langle T(p) \rangle; \text{true}$

The statement says that if an agent  $i$  has perceived agent  $j$ 's Test-if action (with respect to  $p$ ) then agent  $i$  was expecting (before perceiving it) an agent  $j$ 's Test-if action (with respect to  $p$ )<sup>15</sup>.

#### The problem of the "undistinguished Test"

Imagine two agents  $i$  and  $j$  interacting in a physical environment. Agent  $i$  has already perceived agent  $j$  looking at a certain region of the space  $S$  that "contains" a number of objects (propositions  $p, q, r$  etc...). On the basis of indexical cues and background expectations agent  $i$  has disambiguated the situation. Agent  $i$  infers that agent  $j$  has tested whether  $p$  is true or not (according to agent  $i$ ,  $p$  is relevant for agent  $j$ ). Now agent  $i$  turns his sensors towards agent  $j$ 's sensors and agent  $j$  does the same. Each agent has his sensors directed towards the sensors of the others.

We argue that at this level a 3 phases process must be understood. The process is completely independent from the

<sup>15</sup> It seems reasonable that different kinds of background expectations and beliefs could be identifiable at this level. For instance, we could require that the necessary and sufficient condition for specifying the object  $O_i$  (the proposition  $p$ ) of a *Test-if* action of another agent is *not expecting* (before perceiving the *Test-if* action) that the object of the *Test-if* action *will not* be  $O_i$  (the tested proposition will *not* be  $p$ ). This is a weaker condition that the one given in the Background Expectation Axiom.

Moreover, Agent  $i$ 's perception of agent  $j$ 's Test-if action can also be driven by some background belief concerning agent  $j$ 's intentions and goals ("agent  $j$  intends to test whether  $p$  is true or not", "agent  $j$  wants to test whether  $p$  is true or not").

agent  $i$ 's perception of agent  $j$ 's epistemic activity on the external world that has been described above.

1. According to agent  $i$  proposition  $p$  is *relevant* for agent  $j$  and according to agent  $i$  agent  $j$  believes that proposition  $p$  is *relevant* for agent  $i$ .

2. The beliefs specified at point 1 (agent  $i$  believes that  $p$  is relevant for agent  $j$  and the belief that agent  $j$  believes that  $p$  is relevant for agent  $i$ ) give to agent  $i$  the reasons for believing that agent  $j$  has directed his sensors towards agent  $i$ 's sensors *in order to* check the epistemic activity of agent  $i$  with respect to  $p$ .

3. Agent  $i$  cannot identify which kind of *Test-if* action concerning  $p$  (at which level of nesting) agent  $j$  has executed.

When agent  $i$  perceives that agent  $j$  is doing some testing activity on agent  $i$ 's testing activity with respect to  $p$ , he will not be able to specify at which level of nesting agent  $j$  is actually testing the testing activity of agent  $i$  with respect to  $p$ . "Is he actually testing whether I test whether  $p$  is true or not? Or is he actually testing whether I test whether he tests whether  $p$  is true or not? Or is he actually testing whether I test whether he tests whether I test whether  $p$  is true or not? ... and so on. The following axiom describes the previous reasoning:

#### Axiom of undistinguished Test (UTA)

$(\langle P(\langle T(\langle T(\varphi) \rangle; \text{true})^- \rangle; \text{true})^- \rangle; \text{true} \leftrightarrow$   
 $(\langle P(\langle T(\langle T(\langle T(\varphi) \rangle; \text{true})^- \rangle; \text{true})^- \rangle; \text{true})^- \rangle; \text{true}$

Axiom UTA says that if an agent  $i$  has perceived a 2-order Test of agent  $j$  on the  $\varphi$ -testing activity of agent  $i$  then agent  $i$  has perceived a 3-order Test of agent  $j$  on agent  $i$  testing activity on the  $\varphi$ -testing activity of agent  $i$  (Indistinguishability Relation). Given any formula  $\varphi$ , the axiom can be extended to cover all Indistinguishability Relations between 2-order Tests of the other agent and  $n$ -order tests of the other agent. The axiom states that whenever agent  $i$  has evaluated  $\varphi$  to be relevant for agent  $j$  and he has attributed to agent  $j$  the belief that  $\varphi$  is relevant for agent  $i$ , and he has perceived agent  $j$ 's sensors directed towards agent  $i$ 's sensors, he cannot establish at which level agent  $j$  is actually testing agent  $i$ 's testing activity with respect to formula  $\varphi$ . All nested tests are in fact realized by the same physical action. From a phenomenological perspective, a  $n$ -order agent  $j$ 's test-if action on agent  $i$ 's epistemic activity and a  $(n-m)$ -order agent  $j$ 's test-if action (for  $m < n$ ) on agent  $i$ 's epistemic activity cannot be distinguished by agent  $i$ .

#### Implementation of Mutual Belief via Mutual Perception

Before presenting the main result of this section a further axiom is necessary. Notice that the following axiom is given with respect to the special predicate Noise(.) introduced above.

#### Axiom of of Noise/Not-Noise Equivalence for nested Tests (NET)

a.  $\text{Not-Noise}(i, \langle T(\langle T(\varphi) \rangle; \text{true})^- \rangle; \text{true})^- \rightarrow$

$$\text{Not-Noise}(i, \langle T(\langle T(\langle T(\varphi) \rangle_j \text{true}) \rangle_i \text{true}) \rangle_j \text{true})$$

$$b. \text{Noise}(i, \langle T(\langle T(\varphi) \rangle_j \text{true}) \rangle_i \text{true}) \rightarrow$$

$$\text{Noise}(i, \langle T(\langle T(\langle T(\varphi) \rangle_j \text{true}) \rangle_i \text{true}) \rangle_j \text{true})$$

The two axioms establish that if there is (not) Noise for agent  $i$  with respect to agent  $j$ 's test on agent  $i$ 's test if  $\varphi$  then there is (not) Noise for agent  $i$  with respect to agent  $j$ 's test on agent  $i$ 's test on agent  $j$ 's test if  $\varphi$ . We argue that the two axioms are reasonable because at a given moment 2-level tests, 3-level tests, n-level tests (from  $j$ 's sensors to  $i$ 's sensors) are physically realized in the same way and their conditions of Noise (with respect to the same agent) are equivalent. Let us now establish the following theorem:

### Theorem of Mutual Belief Implementation

$$\langle P(\langle T(\langle T(p) \rangle_j \text{true}) \rangle_i \text{true}) \rangle_j \text{true} \wedge$$

$$\bigwedge_{k>1} (\text{Bel}_i \text{Bel}_j)^k (\text{Not-Noise}(i, \langle T(\langle T(p) \rangle_j \text{true}) \rangle_i \text{true}) \wedge$$

$$\text{Not-Noise}(j, \langle T(\langle T(p) \rangle_j \text{true}) \rangle_i \text{true})) \wedge$$

$$\bigwedge_{k>0} (\text{Bel}_i \text{Bel}_j)^k \text{Bel}_i (\text{Not-Noise}(i, \langle T(\langle T(p) \rangle_j \text{true}) \rangle_i \text{true}) \wedge$$

$$\wedge \text{Not-Noise}(j, \langle T(\langle T(p) \rangle_j \text{true}) \rangle_i \text{true})) \rightarrow$$

$$(\bigwedge_{k>1} (\text{Bel}_i \text{Bel}_j)^k p \wedge \bigwedge_{k>0} (\text{Bel}_i \text{Bel}_j)^k \text{Bel}_i p)$$

The theorem establishes the sufficient conditions for guaranteeing the implementation of nested beliefs (from level 3 to level n) in the mind of agent  $i$ <sup>16</sup>. Extending the result of the theorem to the 2-agents case we obtain the sufficient conditions for implementing the infinite conjunctive chain of nested beliefs (from level 3 to level n) that constitutes the structure of a Mutual Belief that  $p$ .

The sufficient conditions are the following:

1. an agent  $i$ 's perception of a 2-order Test of agent  $j$  on the  $p$ -testing activity of agent  $i$  and an agent  $j$ 's perception of a 2-order Test of agent  $i$  on the  $p$ -testing activity of agent  $j$ ;

2. a mutual belief that there is not noise for agent  $i$  with respect to his perception (specified in condition 1) and a mutual belief that there is not noise for agent  $j$  with respect to his perception (specified again in condition 1).

Provided that both agent  $i$  and agent  $j$  enter in the mutual perception already holding the belief that  $p$  and the belief that the other believes that  $p$ , the conditions specified in the theorem are sufficient for guaranteeing the full implementation of a mutual belief that  $p$  (nested beliefs from level 1 to level n).

## Discussion

Although, several authors have discussed the problem of mutual belief achievement, the seminal contribution of Lewis (1969) is still accepted. Recently Cubitt & Sugden

(2003) have proposed a complete formalization. Let us here introduce Lewis' conditions in a really simplified way within our formalism.

Condition 1. For all agents  $i$ :  $\text{Bel}_i A$

Condition 2. For all agents  $i$ :  $\text{Bel}_i (A \rightarrow \text{Bel}_i A)$

Condition 3. For all agents  $i$ :  $\text{Bel}_i (A \rightarrow p)$

Condition 4. For all agents  $i, j$  and for all propositions  $y$ :  
 $\text{Bel}_i (A \rightarrow y) \rightarrow \text{Bel}_i \text{Bel}_j (A \rightarrow y)$ .

$A$  is defined as the *reflexive common indicator* that  $p$ :  $A$  is responsible for the generation of higher-order beliefs in the mutual belief structure. Condition 1 and 2 are the conditions of *public announcement*, i.e. if  $A$  is true then all agents believes  $A$ , for all agents  $i$  if agent  $i$  believes that  $A$  holds then all other agents believe that  $A$  holds. Condition 4 is very close to the property that Lewis states as 'suitable ancillary premises regarding our rationality, inductive standards, and background information': all agents believe to share the same inductive standards, i.e. for all propositions  $y$  and for all agents if agent  $i$  believes that  $A$  implies  $y$  then agent  $i$  believes that agent  $j$  believes that  $A$  implies  $y$ . We think that Conditions 1,2 and 4 do not hold in all interaction contexts. Even assuming that all agents believe that  $A$  is true, there are cases in which not all agents in the community believe that the other agents believe that  $A$ . On the other side there are cases in which agents do not believe to have the same inductive standards. Take for example the case where the state of affairs  $A$  is "the alarm goes off" and where the proposition  $p$  is "there is a fire in the building". Moreover, assume that agent  $i$  perceives that the "the alarm goes off". We think that it is quite problematic to assume that in this context agent  $i$  holds the belief that all agents in the building will perceive the alarm and to assume that agent  $i$  holds the belief that all agents in the building believe that "alarm" means "fire in the building". Indeed it is not so obvious to assume that conditions 2 and 4 are holding in this situation. An analysis of what makes an event public is missing at this point. Were the other agents attending at the state of affairs  $A$  ("the alarm goes off") that I have perceived? "Are the other agents sharing my same knowledge? ("Does everybody know that alarm means fire in the building"?). To interpret the example with our model we substitute condition 2 and 4 given by Lewis with the notion of Mutual Perception. Take for instance the 2 agents case and focus on agent's  $i$  reasoning and epistemic activity: 1) agent  $i$  believes that "the alarm goes off"; 2) agent  $i$  believes that agent  $j$  believes that "the alarm goes off"; 3) agent  $i$  believes that agent  $j$  believes that "alarm means fire", 4) agent  $i$  perceives that agent  $j$  is testing whether "the alarm goes off" or not *in order* to verify whether "there is a fire in the building" or not<sup>17</sup>, 5) agent  $i$  perceives agent  $j$  directing his sensors towards agent  $i$ 's sensors in order to check the epistemic activity of agent  $i$  with respect to "the alarm goes off" (and

<sup>16</sup> The formal proof of the theorem is not given in this paper. The proof is by induction and is based on standard modal principles of Beliefs, the Introspection Axiom for perception, the Noise Axiom, the Axiom of undistinguished Test and the Axiom of Noise/Not-Noise Equivalence for nested Tests.

<sup>17</sup> A background belief allows agent  $i$  to disambiguate the epistemic activity of agent  $j$ . Agent  $i$  believes that the fact "there is fire in the building" is relevant for agent  $j$ .

indirectly for checking the epistemic activity of agent  $i$  with respect to “there is a fire in the building”<sup>18</sup>. Assume that 1, 2, 3, 4 and 5 hold also for agent  $j$ . If we add to 1,2,3,4 and 5 for  $i$  and  $j$  the Mutual Belief that “there is not noise for agent  $i$  and  $j$  with respect to the state of affairs ‘the alarm goes off’”, we achieve (on the basis of the theorem of Mutual Belief Implementation) the Mutual belief that “the alarm goes off” and given the shared rule “alarm means fire” the full achievement of the Mutual Belief that “there is a fire in the building”. We want to suggest in this work that an analysis of mutual perception scenarios is needed in order to understand how mutual belief can be achieved in a community of agents. An analysis of agents’ reasoning about shared conditions of Noise should be included at this level. This kind of approach would provide a specification of the social context features that potentially guarantee the implementation of the mutual belief. This approach is very close to the one of Clark & Marshall (1981) who advanced the simultaneity assumption<sup>19</sup> and the attention assumption<sup>20</sup> as bases for mutual belief implementation.

### Conclusion

In this paper, we have discussed the conditions that guarantee the achievement of mutual belief in a community of agents. Mutual perception has been addressed as a fundamental process for understanding how knowledge gets shared. We have decomposed the notion of public announcement by trying to individuate the conditions that make either the realization of a natural event or the realization of an action or the truth of a state of affairs mutually believed. The theorem of mutual belief implementation represents the formal result that makes explicit those conditions. In the theorem the condition of “mutual belief that the environment is not noisy for all the agents in the community” is required for the implementation of the mutual belief that  $p$ . This result is in accordance with the well-known theorem by Fagin *et al.* (1995) stating that “if the communication channel is noisy (and there is mutual belief about that) no communication protocol can guarantee mutual knowledge achievement”. However our condition is even stronger. Our future work will be devoted to prove the *Impossibility Theorem* saying that “if there is not mutual belief about the reliability of the channel (Not Noise) then there is not communication protocol that guarantees mutual belief achievement”.

<sup>18</sup> Together with the background belief given at point 2 an additional background belief is operating at this level: agent  $i$  believes that agent  $j$  believes that the fact “there is fire in the building” is relevant for agent  $i$ .

<sup>19</sup> Agent  $i$  sees that agent  $j$  has his eyes open and is looking simultaneously at her and object in the world. That is, she has evidence that she and the other are looking at each other and the object simultaneously.

<sup>20</sup> Agent  $i$  assumes that Agent  $j$  is not only looking at her and the object, but also attending to them.

### Acknowledgments

The first two authors of this research are supported by the European Project MindRACES (IST-511931). We would also like to thank Cristiano Castelfranchi for helpful comments on previous drafts of the paper.

### References

- Bacharach, M. (1992). The Acquisition of Common Knowledge. In Bicchieri C., Dalla Chiara M.L. (Eds.), *Knowledge, Belief and Strategic Interaction*, Cambridge, MA: Cambridge University Press.
- Baltag, A., Moss, L., Solecki, S. (1998), The Logic of Public Announcements, Common Knowledge and Private Suspicion. In *Proceedings TARK 1998*. (pp. 43-56), Los Altos: Morgan Kaufmann Publishers.
- Clark, H. H. & Marshall C. (1981). Definite reference and mutual knowledge. In A. Joshi, B. Webber, and I. Sag, editors, *Elements of Discourse Understanding*, pages 10–63. Cambridge, MA: Cambridge University Press.
- Cubitt, R.P. & Sugden, R. (2003) Common Knowledge, Salience and Convention: a reconstruction of David Lewis’ Game Theory. *Economics and Philosophy*, 19,175–210.
- Del Val, A., Shoham, Y., Maynard-Reid II, P. (1997) Qualitative Reasoning about Perception and Belief. In *Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence*. San Mateo, CA: Morgan Kaufmann.
- Fagin, R., Halpern, J. Y., Moses, Y., Vardi, M. Y. (1995). *Reasoning about Knowledge*. MIT Press, Cambridge.
- Geanakoplos, J. (1992). Common Knowledge. *Journal of Economic Perspectives*, 6, 53-82.
- Grosz, B.J. & Kraus, S. (1996). Collaborative Plans for Complex Group Action. *Artificial Intelligence*, 86, 269-357.
- Harel, D., Kozen D., Tiuryn, J.(2000) *Dynamic Logic*, Cambridge, MA: MIT Press.
- Harrah, D. (2002). The Logic of Questions. In D. Gabbay and F. Guenther (Eds.), *Handbook of Philosophical Logic*, vol. 8, pp. 1-60.
- Herzig, A. & Longin, D. (2002). Sensing and revision in a modal logic of belief and action. *Proceedings of the Fifteenth European Conference on Artificial Intelligence*, (pp. 307-311). The Netherlands: IOL Press.
- Kirsh, D & Maglio, P.P. (1994). On Distinguishing Epistemic from Pragmatic Actions. *Cognitive Science*, 18, 513-549.
- Lewis, D. (1969). *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- Lorini, E. & Castelfranchi, C. (2004). The role of epistemic actions in expectations. *Proceedings of ABIALS 2004*, Los Angeles, 17 July.
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.