

# A Logic of Epistemic Actions

Alexandru Baltag\*

## 1 Introduction

The subject of this paper is a logic that combines in a specific way a multi-agent epistemic logic with a dynamic logic of “epistemic actions”. The paper continues and improves on the work in our previous paper [BMS] (joint work with L.S. Moss and S. Solecki). The origins of the subject are in Fagin et al [FHMV], where the authors analyse knowledge in distributed systems, using a mixture of epistemic logic  $S5_M$  and temporal logic. The fundamental issues, examples and insights that gave rise to our logic come from the work in [FHMV]. But their approach runs into several problems. First, the resulting logic is *too strong*: in general, it is not decidable, and for many classes of systems is not even finitely axiomatizable. Secondly, from a different perspective, their logic seems to be *not expressive enough*: there is no notion of *updating knowledge (information)*; one can talk about the change of information induced by specific actions, but only about what happens “next” (or “always” or “sometimes” in the future), and this is determined by the model. (In their setting, the semantics is given by “runs”, i.e. temporal sequences of Kripke structures.) When they actually analyse concrete examples (e.g. The Muddy Children Puzzle), they do *not use their formal logic* only, but also “external” (semantic) reasoning about models; in effect, they simply “update” their structures from the outside (in the meta-language) according to informal intuitions, but without any attempt to give a systematic treatment of this operation.

The seminal idea of our work comes from a paper of Gerbrandy and Groeneveld [GG]. The idea is to combine Fagin-style epistemic logic with the work of Veltman [V] on update semantics. The authors introduce special kinds of epistemic actions, namely *public announcements* (“group updates”). Their logic is strong enough to capture all the reasoning involved in The Muddy Children Puzzle. In his Ph.D. dissertation [G], Gerbrandy improves and extends these ideas with a “program-update” logic. Our own work, developed in [BMS], started from observing some odd (or at least not always desirable) features of Gerbrandy’s and Groeneveld’s public announcements. Namely, they have “group-learning” actions of the form  $\mathcal{L}_A\varphi$ , with the intended meaning “the agents in the group  $A$  learn in common that  $\varphi$  is true”. The problem is that, with their definition, agents that are *outside* the group  $A$  (the “outsiders”) do not in any way *suspect* that the group-announcement is happening. Of course, they wouldn’t *know* it is happening (since they are not part of the “inside” group), but (by the GG definition) these outsiders are not even allowed to *consider the possibility* that such an announcement might be happening. As a result, they are totally “mislead”: after this action, in the resulting Kripke structure, the outsiders “live in an ideal world” (i.e. they do not “access” the actual world anymore).

---

\*CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands. Email: [abaltag@cwi.nl](mailto:abaltag@cwi.nl)

To put it differently, even if the initial model was a “knowledge structure” (i.e.  $S5_m$ -model), updating it with any announcement with at least two insiders and one outsider will result in a non-knowledge (non- $S5_m$ , more specifically non-reflexive) structure: the outsider acquires false beliefs about the world. Such an interesting “deceiving” situation is indeed possible in real life and we would like to still have such an action in our logic; but we wouldn’t want to impose that *every* group-announcement-with-outsiders be necessarily deceiving! Moreover, we would like to give the outsiders a better chance: not only that they could *suspect* something is going on, but on the basis of this suspicion they might *act*, attempting to confirm their suspicions. They could, for instance, wiretap or intercept the communications of the “insiders”.

More generally, the idea is to take at face value the notion of *epistemic update* and study it as an object in itself, in full generality. An epistemic update is a partial function  $F$  mapping epistemic models to epistemic models. We think of these functions as ways of “updating” a model, corresponding to changes in the information states of the models. For this reason, we shall only consider in this paper (as in [BMS]) update functions that are “truly epistemic”, i.e. in which nothing else changes but the information possessed by agents; in other words, the “facts about the world” do not change by such an update. More specifically, we want to understand epistemic updates as being induced by “epistemic actions”. Epistemic actions are (possibly complex) actions performed by agents in order to change their information states. Examples of such actions: *message-passing*, *public announcements*, “*card-showing*” in a card game, *misleading (or secret) actions* (in which non-participants are deceived into believing no action has been taken), *gratuitous suspicion* (agents suspect some action is taking place, but this is not the case), *wiretapping* (secret interception) of messages and more generally *secret-breaking, suspicious actions* (which are secretly taking place, but which nevertheless are being suspected to be taking place by some of the non-participants) etc. In general, the interesting type of actions that our system can capture are “half-transparent-half-hidden-actions”. For example, a move in a game can be such that some players “see” some part (or feature) of what is happening but not the whole move; nevertheless, if the “move” is legal they will necessarily “suspect” it, i.e. regard it as a possibility.

In both [BMS] and the present paper, we only try to represent the *epistemic content* of such actions, not their other aspects (intentional, factual etc.). We model the seeming complexity of such actions by endowing them with *internal structure*. Namely, there are (at least) two important epistemic features of such an action: (1) its *presuppositions* or preconditions of happening; these refer to the actual world or to the agent’s beliefs about the world *before* the action, and they define the applicability of this particular action to this particular world: not every action can happen in every world; (2) the agent’s views or beliefs about the very action that is taking place; i.e. action’s “appearance” to the agents. These beliefs are given, as in the case of “static” epistemic models, by accessibility relations corresponding to each agent, to represent what are the actions that agent considers as “alternatives” of the real action; i.e. what are the actions that each agent “cannot distinguish” from the real action.

In [BMS] we have introduced *finite* epistemic actions as *syntactic objects*. We used them as labels for actions in traditional dynamic-logic style, and we axiomatized the resulting logic. In the present paper, I am studying actions as *semantical objects*, on the same level with Kripke models. I define natural operations on actions, develop a “calculus of epistemic actions” and I prove two “normal form” representation theorems. For the associated logic, this allows me to replace the “syntactic actions” from our previous work with more “syntactically looking” (i.e.

recursively built) action-expressions, that *denote* the “real” (semantical) actions.

As an application, I use this setting and the logic to study modified versions of The Muddy Children Puzzle: some children cheat, by sending signals to tell their friends they are dirty; the others might not suspect it, which can lead to a totally wrong line of reasoning on their part, ending in a wrong answer; or they could be more cautious and suspicious, which allows them to use other agent’s wrong answers to find the truth more quickly than in the classical puzzle. I hope this could be the beginning of a more general study of the “logic of cheating at games”.

## 2 A Modal Logic for Epistemic Actions.

I introduce here a modal language to describe the update of epistemic structures by epistemic actions. (Our notion of epistemic structure is the standard one, which will be formally defined in the next section, but here we assume an informal understanding of this notion.) Our language  $L$  is obtained by putting together standard epistemic logic (with “common knowledge” operators) with a dynamic-logic of epistemic actions. For agents  $a$  and sets of agents  $A$ , we have the following standard epistemic modalities  $\Box_a$  (the *belief*, or *knowledge*, operator) and  $\Box_A^*$  (the *common belief*, or *common knowledge*, operator). The sentence  $\Box_a\varphi$  will denote the fact that *agent  $a$  believes that  $\varphi$* , while  $\Box_A^*\varphi$  will mean that  *$\varphi$  is a common belief among all the agents of the group  $A$* . In addition, we have *action-expressions* to denote *epistemic actions*, i.e. “programs” updating epistemic situations. As we shall see in the semantics given in the next section, such programs take epistemic structures as input and produce another epistemic structure as output. Our action expressions are generated from some basic ones by “process-algebra”-like operations. Finally, we also have, for each such action expression  $\alpha$ , a “dynamic-logic”-type modality  $[\alpha]$ ; the sentence  $[\alpha]\varphi$  denotes the fact that *after action  $\alpha$ , sentence  $\varphi$  becomes true*, or more precisely, that if  $\alpha$  can be executed then its output-structure satisfies  $\varphi$ .

**Syntax.** We assume as given a set  $AtProp$  of atomic propositions, denoted by  $P, Q, \dots$ , a set  $Ag$  of agents, denoted by  $a, b, \dots$ , and a set  $Var$  of *action variables*  $x, y, \dots$ . As before, we use capital letters  $A, B, \dots \subseteq Ag$  to denote *finite sets of agents*. We define now by simultaneous recursion: a set  $L$  of *propositions* (denoted by  $\varphi, \psi, \dots$ ) and, for each finite (possibly empty) list  $\vec{x}$  of variables, a set  $Act_L(\vec{x})$  of *action expressions in the free variables  $\vec{x}$* . The elements of  $Act_L(\vec{x})$  will be denoted by  $\alpha(\vec{x}), \beta(\vec{x}), \dots$ . The elements of the set  $Act_L(\emptyset)$  (i.e. the expressions with no free variables) will be denoted by  $\alpha, \beta, \dots$  and called *closed action expressions*, or more simply *actions*. (But they are not to be confused with the semantical structures with the same name, to be introduced later: formally they are different, but of course closed action expressions are simply *names* for finite actions.)

$$\begin{array}{llllllll} \varphi, & \psi & =: & P & \neg\varphi & \varphi \wedge \psi & \Box_a\varphi & \Box_A^*\varphi & [\alpha]\varphi \\ \alpha(\vec{x}), & \beta(\vec{x}) & =: & ?\varphi & x^a & \alpha(\vec{x})^a & \alpha(\vec{x}) + \beta(\vec{x}) & \alpha(\vec{x}) \circ \beta(\vec{x}) & \mu y. \alpha(y, \vec{x}) \end{array}$$

(In this definition, we assume that  $x$  denotes any variable in the vector  $\vec{x}$  and  $y$  denotes any variable which is *not* in  $\vec{x}$ .)

The semantics will be given in the next section. Informally, the meanings of our action-constructions are: “test  $\varphi$ ”  $?\varphi$  is the action that *tests the truth* of a proposition  $\varphi$ , i.e the program which accepts an epistemic structure as input iff  $\varphi$  is true (in which case it returns some trivial structure corresponding to “termination”). The action  $\alpha^a$  is the action in which

agent  $a$  “suspects” (regards as an epistemic possibility) that some action  $\alpha$  might be happening (while in reality no action happens, except for a getting suspicious). The sum  $\alpha + \beta$  is a sort of *parallel composition* of the two actions, while  $\alpha \circ \beta$  is their *sequential composition*. Finally,  $\mu$  is a fixed-point operator, allowing us to define self-referential, “circular” actions (in which for instance an action in which an agent  $a$  can suspect, or even know, the very action that is happening). The only reason we introduce variables  $x, y$ , and action expressions  $\alpha(\vec{x})$  having free variables, is to define such self-referential actions via fixed-point constructions  $\mu y. \alpha(y, \vec{x})$ . As one can easily see from the above formal definition, we only allow *guarded* action-expressions, i.e. in which every free variable  $x$  occurs only in the scope of some “suspicion-operator”  $\alpha(y, \vec{x})^a$ . This will ensure the existence of the corresponding fixed-points. One can define the *substitution of a variable  $y$  in an action expression  $\alpha(y, \vec{x})$  by an action  $\beta$*  in the usual manner, and denote its output by  $\alpha(\beta, \vec{x})$ . As seen from the informal description above, we intend to model our actions as carrying themselves an “epistemic action structure”, given by what agents “see”, “know” or “suspect” about the action. This will be formally expressed by a *suspicion relation*  $\rightarrow_a \subseteq Act_L(\emptyset) \times Act_L(\emptyset)$  between closed action expressions. Similarly, one can see from the description of the test action  $?\varphi$  that some actions are not always possible, since their application requires some “preconditions” or “presuppositions” to be satisfied. This will be formally expressed by a *precondition function*  $PRE : Act_L(\emptyset) \rightarrow \mathcal{P}(L)$ , assigning to each closed action expression its *set of preconditions*. The intuition is that, for instance, the “test” action  $?\varphi$  is “possible” only if  $\varphi$  is true; in case this action is possible, its execution consists in “termination”, i.e. no agent “does” anything, in particular no agent “suspects” anything. Similarly, an action of “suspicion” of the form  $\alpha^a$  can always happen (as possibly “gratuitous” suspicion), so it does not require precondition; in such an action, agent  $a$  “suspects” some action  $\alpha$ , while the other agents don’t suspect anything. The parallel composition action  $\alpha + \beta$  is possible only if both actions are possible, so its set of preconditions is the union of the preconditions of  $\alpha$  and of  $\beta$ ; when  $\alpha + \beta$  happens, each agent “suspects” (regards as possible alternative actions) simultaneously *both* the actions that he would suspect in  $\alpha$  and the ones he would suspect in  $\beta$ . The sequential composition action  $\alpha \circ \beta$  is possible only when  $\alpha$  is first possible and then, after  $\alpha$  is executed,  $\beta$  becomes possible; in other words, the set of preconditions of  $\alpha \circ \beta$  consists of the preconditions of  $\alpha$  and the sentences expressing the fact that, after  $\alpha$ , the preconditions of  $\beta$  become true. Also, the execution of  $\alpha \circ \beta$  consists first of the execution of  $\alpha$  followed by the execution of  $\beta$ ; when this happens, each agent suspects first the actions  $\alpha'$  which he suspects when  $\alpha$  is happening, and then suspects the actions  $\beta'$  suspected in  $\beta$ ; in other words, the agent suspects all the possible (consistent) sequential compositions  $\alpha' \circ \beta'$  of such alternative actions  $\alpha', \beta'$ . Finally, the fixed-point action  $\mu x. \alpha(x)$  is supposed to be equivalent to the result of substituting  $x$  in the expression  $\alpha(x)$  by the very same action  $\mu x. \alpha(x)$ : so its preconditions and epistemic structure are the same as the ones of  $\alpha(\mu x. \alpha(x))$ .

Using the notation  $[\alpha]\Phi =: \{[\alpha]\varphi : \varphi \in \Phi\}$ , we formally define by recursion the function  $PRE$  and the relations  $\rightarrow_a$  between closed action expressions in the following way:

$$\begin{aligned}
PRE(?\varphi) &=: \{\varphi\} \\
PRE(\alpha^a) &=: \emptyset \\
PRE(\alpha + \beta) &=: PRE(\alpha) \cup PRE(\beta) \\
PRE(\alpha \circ \beta) &=: PRE(\alpha) \cup [\alpha]PRE(\beta) \\
PRE(\mu x. \alpha(x)) &=: PRE(\alpha(\mu x. \alpha(x)))
\end{aligned}$$

and

$$\alpha^a \rightarrow_a \alpha,$$

if  $\alpha \rightarrow_a \gamma$  then  $\alpha + \beta \rightarrow_a \gamma$ ; similarly, if  $\beta \rightarrow_a \gamma$ ,

if  $\alpha \rightarrow_a \beta$  and  $\alpha' \rightarrow_a \beta'$  then  $\alpha \circ \alpha' \rightarrow_a \beta \circ \beta'$ ,

if  $\alpha(\mu x.\alpha(x)) \rightarrow_a \beta$  then  $\mu x.\alpha(x) \rightarrow_a \beta$ .

One can easily prove that a closed action expression has only *finitely many preconditions*. This allows us to define for each such expression  $\alpha$  a unique sentence  $pre(\alpha)$ , called *the presupposition of  $\alpha$* , and given by the conjunction of all the preconditions:  $pre(\alpha) =: \bigwedge PRE(\alpha)$ . Similarly, one can define, for each set  $A$  of agents, the *iterated suspicion relation*  $\rightarrow_A^*$  between closed action expressions as the reflexive-transitive closure of the union  $\bigcup_{a \in A} \rightarrow_a$  of all ; or, equivalently, by the recursive conditions: (1)  $\alpha \rightarrow_A^* \alpha$  and (2) if  $\alpha \rightarrow_A^* \beta$  and  $\beta \rightarrow_a \gamma$  for some  $a \in A$ , then  $\alpha \rightarrow_A^* \gamma$ . Again, one can easily see that, for any closed action expression, the set of all its  $\rightarrow_A^*$ -successors is *finite*. By taking  $A$  to be the set  $Ag$  of all agents, we define the *set of epistemic alternatives of an action*  $ALT(\alpha) = \{\beta : \alpha \rightarrow_{Ag}^* \beta\}$ . (So, in particular,  $ALT(\alpha)$  is also finite.) Finally, for a closed action-expression, we make the following notation:

$$\|\alpha\| =: ((ALT(\alpha), (\rightarrow_a)_{a \in Ag}, PRE^\alpha), \alpha),$$

where  $PRE^\alpha$  is the restriction of  $PRE$  to  $ALT(\alpha)$  (given by  $PRE^\alpha(\beta) =: PRE(\beta)$  for every  $\beta \in ALT(\alpha)$ ). This object  $\|\alpha\|$  will be a “real action” (in the sense of the next section), giving a semantical interpretation to our action expression  $\alpha$ .

### 3 Models and Actions.

As in the previous section, we fix a set  $Ag$  of agents and a set  $AtProp$  of atomic propositions, and we consider the language  $L$  of epistemic action logic, as defined in the previous section. We define first a notion of *epistemic action*, which is formally more general than the standard notion of epistemic model: as we shall see, *from a formal point of view, models are just actions of a special type*. Nevertheless, the interpretation of a structure as a model is intuitively very different from the interpretation of the same structure as an action: the first notion is *static*, while the second is *dynamic*. The “tokens” (elements) of an epistemic model are to be understood as states or “possible worlds, while the elements of an action structure are “action-tokens” (possible actions). For this reason, we shall use slightly different notations for the two notions.

A *Kripke frame over  $Ag$*  is a triplet  $(K, (\rightarrow_a^K)_{a \in Ag})$ , where  $K$  is a set of *tokens*, each  $\rightarrow_a^K$  is a binary accessibility relation on  $K$ . As usually, we skip the superscript  $K$  and write  $\rightarrow_a$  for the accessibility relations when the underlying structure is unambiguously fixed. Given a set  $L_0 \subseteq L$  of propositions of our language, an *epistemic action structure over  $L_0$*  is a triplet  $\mathbf{K} = (K, (\rightarrow_a^K)_{a \in Ag}, PRE^K)$ , consisting of a Kripke frame  $(K, (\rightarrow_a^K)_{a \in Ag})$  and a *precondition function*  $PRE : K \rightarrow \mathcal{P}(L_0)$ , associating to each  $k \in K$  some set of propositions in  $L_0$ . An *epistemic action* is just a pointed epistemic structure, i.e. a pair  $\alpha = (\mathbf{K}, k_0)$  of an action structure and a designated token  $k_0 \in K$ . The class of all actions over  $L_0$  will be denoted by  $Act_{L_0}$ . The class of all actions over our (fixed) language  $L$  is denoted by  $Act = Act_L$ .

When we interpret such a structure as an action, the designated token  $k_0$  is called *the “actual action” of  $\alpha$  (or the “top” of  $\alpha$ )* and is supposed to represent the “real” action taking place; the members of  $K$  are called *action tokens* and they represent possible alternatives for this action, some of which may be (wrongly) believed, or just suspected, to be happening by some of the agents; accordingly, the accessibility relations  $\rightarrow_a$  are called *suspicion relations*, and we read  $k \rightarrow_a^K k'$  as “when action  $k$  happens, agent  $a$  suspects that action  $k'$  may be happening”; the members of  $PRE^K(k)$  are called the *preconditions of action(-token)  $k$* . These are to be understood as being all the conditions that any “world” would have to fulfill in order for the action-token  $k$  to be “possible” in that world. An action structure is said to be *finite* if the underlying set  $K$  of tokens is finite. The action  $\alpha = (\mathbf{K}, k)$  is said to be *finite* if the underlying structure  $\mathbf{K}$  is finite. For finite action structures  $\mathbf{K}$ , we can define a *presupposition function*  $pre^K : K \rightarrow L$  associating to each token the conjunction of all its preconditions:  $pre^K(k) =: \bigwedge PRE^K(k)$ . (Note that, if  $PRE^K(k) = \emptyset$ , then  $pre^K(k) = T$ , the universally true proposition.)

As announced, we formally define an *epistemic model* to be just just an epistemic action structure over *AtProp*, i.e. an action structure  $\mathbf{W}$  such that the precondition  $PRE : K \rightarrow \mathcal{P}(AtProp)$  takes *only atomic sentences as values*. When we interpret such a structure as a model, we use the notations  $W, w, \mathbf{W}, V$  instead of  $K, k, \mathbf{K}, PRE$ , and so we write our models as  $\mathbf{W} = (W, (\rightarrow_a^W)_{a \in Ag}, V^W)$ , with  $V : W \rightarrow \mathcal{P}(AtProp)$ . An *epistemic state*, or a “pointed model” is just an action over *AtProp*, i.e. a “pointed model”: a pair  $s = (\mathbf{W}, w_0)$  of an epistemic model and a designated state-token  $w_0 \in W$ . In this case, the tokens  $w \in W$  are called *state tokens* or *possible worlds* of  $s$ , the accessibility relation  $\rightarrow_a^W$  is called *the indistinguishability relation for agent  $a$* ,  $w_0$  is called the *actual state (world)* (or the “top”) of the model,  $v$  is called the *valuation function* and the fact that  $p \in V^W(w)$  as read as “ $p$  is true at state-token  $w$ ”. We denote by *Mod* the class of all pointed models (states). Clearly, we have  $Mod \subseteq Act$ .

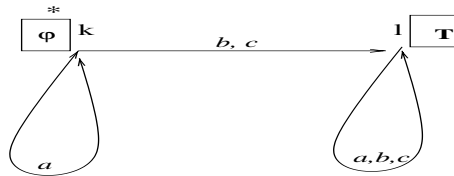
To be precise, we have labelled the accessibility relations and the precondition function with superscripts for both actions and models, to indicate the underlying structure, as in  $\rightarrow_a^W$  and  $\rightarrow_a^K$ . To avoid writing this extra-superscript all the time, we shall “lift” our “local” accessibility relations  $\rightarrow_a$  between tokens *inside* structures (actions or models) to “universal” accessibility relations *between* pointed structures (actions or states); namely, an action (state) is  $a$ -accessible from another action (state) whenever we have that: the two underlying action structures (models) coincide, and the “top” of the first action (state) is  $a$ -accessible from the “top” of the second. More precisely, e.g. for actions  $\alpha = (K, (\rightarrow_a)_{a \in Ag}, k_0, PRE^K)$ ,  $\beta = (K', (\rightarrow'_a)_{a \in Ag}, k'_0, PRE^{K'})$ , we put  $\alpha \rightarrow_a \beta$  iff:  $K = K'$ ,  $\rightarrow_b = \rightarrow'_b$  for all  $b \in Ag$ ,  $PRE^K = PRE^{K'}$  and  $k_0 \rightarrow_a k'_0$ . Similarly, we put:  $PRE(\alpha) =: PRE^K(k_0)$ ,  $pre(\alpha) =: pre^K(k_0)$ . So we can talk about “the precondition of an action  $\alpha$ ” etc. Since  $Mod \subseteq Act$ , this induces corresponding relations and valuation function on epistemic states (=pointed models): e.g., for states  $s = (\mathbf{W}, w), s' \in Mod$ , we can write  $s \rightarrow_a s'$ ,  $V(s) = V^W(w)$  etc. Naturally, the class *Act*, endowed with these accessibility relations and precondition maps, forms a “large” Kripke structure, and *Mod* forms a substructure of *Act*. This is a technical trick giving us a way to make our accessibility relations “model-independent”. Similarly, the use of states, as “pointed” models, gives us a way of making the *truth (satisfaction) relation* to be “model-independent”. Usually, truth in modal logic is a ternary relation  $(\mathbf{M}, w) \models \varphi$  between a state token  $w$  (of a model), the (underlying) model  $\mathbf{M}$  and a proposition  $\varphi$ . By using states  $s = (\mathbf{W}, w_0)$ , we can (and will) simply write  $s \models \varphi$ .

To define common knowledge, we need to introduce iterated accessibility relations: for each group  $A \subseteq Ag$  of agents, define relation  $\rightarrow_A^*$  as the reflexive-transitive closure of the union  $\bigcup_{a \in A} \rightarrow_a$ ; in other words: we have  $\alpha \rightarrow_A^* \beta$  iff there exists an  $A$ -chain  $\alpha = \alpha_0 \rightarrow_{a_0} \alpha_1 \rightarrow_{a_1} \dots \rightarrow_{a_n} \beta$ , with  $a_i \in A$  for every  $i$ . This is the semantic counterpart of the similar syntactic relation defined in section 2. The standard Kripke-style definition of *belief* (or *knowledge*) in an epistemic model can be easily seen to be equivalent to the following definition: we say that, in *epistemic state*  $s$ , *agent*  $a$  *believes that*  $\varphi$  (write  $s \models \Box_a \varphi$ ) if  $\varphi$  is true in all models  $s'$  such that  $s \rightarrow_a s'$ . Similarly, one can define *common knowledge* inside a *set of agents*: for a set  $A \subseteq Ag$  of agents, we say that  $\varphi$  is *common belief in state*  $s$  *among the agents in the group*  $A$  if  $\varphi$  is true in all states that are accessible from  $s$  by a finite sequence of  $A$ -arrows; i.e. in all states  $s'$  such that  $s \rightarrow_A^* s'$ .

**Examples of actions:** Let us fix our set of agents  $Ag = \{a, b, c\}$ .

1. **(Private, Truthfull, Conscious, Introspective) Learning:** Agent  $a$  learns (discovers) that some proposition  $\varphi$  is true. The act of learning is done *in private*: while it is happening, nobody else knows, *or even suspects*, that it is happening. (Accordingly, after this action, agents  $b$  and  $c$  remain in the same information-state as before.) The act of learning is indeed learning and not just a belief-revision, in the sense that it is *truthfull*:  $\varphi$  is actually true. The act of learning is *conscious* and *introspective*, in the sense that agent  $a$  knows what she is doing and knows that nothing else happens in the meantime.

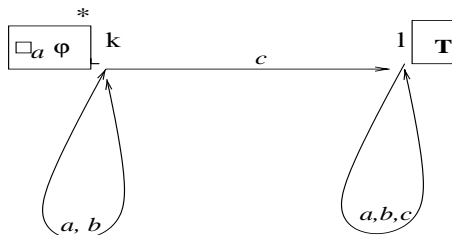
This action  $\alpha$  can be described by an action structure with two action-tokens,  $K = \{k, l\}$ . Here  $k$  represents the “real” action that is taking place (learning of  $\varphi$  by agent  $a$ ), action which has as presupposition the truth of  $\varphi$ ,  $PRE(k) = \{\varphi\}$ : one cannot truthfully learn something false. (If we wanted to model a notion of “truthfull and informative (non-redundant) learning, we would have to add as extra-presupposition the fact that agent  $a$  doesn’t know  $\varphi$  before the action, i.e. we would put  $PRE(k) = \{\varphi, \neg \Box_a \varphi\}$ .) On the other hand,  $l$  represents *the action that agents  $b$  and  $c$  think that is taking place*, namely nothing:  $l$  will just be the “trivial” action in which nothing changes. This trivial action can “happen” anywhere, so it has “no precondition”, i.e. its presupposition is always true:  $PRE(l) = \emptyset$ . Also, the trivial action is completely “transparent”, in the sense that, if it happens, then everybody knows it is happening; so it is its own only successor:  $l \rightarrow_a l, l \rightarrow_b l, l \rightarrow_c l$  (and no others). On the contrary, action-token  $k$  “looks like” the trivial one  $l$  from the point of view of  $b$  and  $c$ , i.e.  $k \rightarrow_b l, k \rightarrow_c l$ , while the same action-token  $k$  is “transparent” to  $a$ , who knows that  $k$  is happening, so she considers  $k$  as its own only alterenative:  $k \rightarrow_a k$ . The picture is



, where the action-tokens are represented by boxes that surround their own presuppositions and the star is used to mark the designated “top” action-token (the “actual action”).

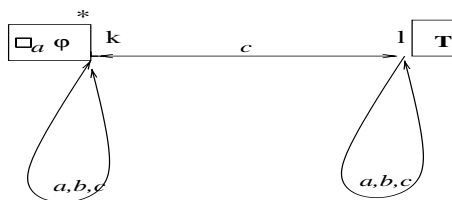
2. **Secure Group Announcements with no Suspicion:** Suppose  $a$  and  $b$  get together, without  $c$  suspecting this (or, alternatively, suppose  $a$  and  $b$  have common access to a secret, reliable and secure communication channel). Agent  $a$  makes a sincere announcement  $\varphi$  at this gathering (or sends a sincere message over this channel). Here, “sincere” means that  $a$  actually

believes  $\varphi$  to be true, and we actually assume more, namely that  $a$  and  $b$  trust each other. As mentioned,  $c$  does not suspect that this is happening: he trusts  $a$  and  $b$  and does not even consider the possibility of such a secret communication. (Or, alternatively, one can say that the act of communication is done in such a misleading way, that it appears to  $c$  as if nothing happened, and that nothing could happen.) This action  $\alpha$  can be described by

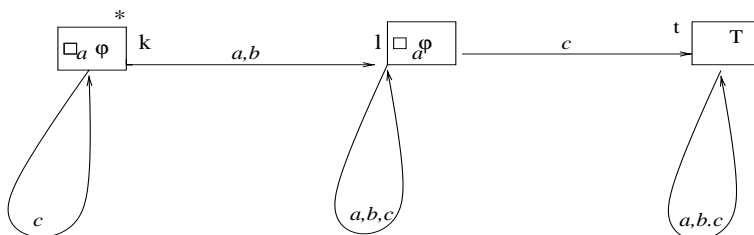


Here,  $K = \{k, l\}$  as before,  $k \rightarrow_a k, k \rightarrow_b k, k \rightarrow_c l, l \rightarrow_a l, l \rightarrow_b l, l \rightarrow_c l$ ,  $PRE(k) = \{\Box_a \varphi\}$  (since the announcement is “sincere”, so the presupposition is that  $a$  believes  $\varphi$ ),  $PRE(l) = \emptyset$  (the universally true condition).

**3. Reliable, Secure Group announcements with a suspicious outsider:** As in example 2, but now  $c$  is suspicious: he doesn’t trust  $a$  and  $b$  so much, so he suspects this group announcement might be happening. He does not necessarily believe it is happening, but he doesn’t exclude such a possibility. On the other hand,  $a$  and  $b$  know this, and moreover they have *common knowledge of this suspicious character of  $c$* .



**4. Group Announcements with a (Secure) Wiretap:** As in the last example, but now  $c$  is not only suspicious, but extremely curious: he actually wiretaps the conversation between  $a$  and  $b$  (or violates their mail etc.). So  $c$  knows about the announcement, while  $a$  and  $b$  don’t suspect this: they just do not consider wiretapping as a real possibility; but they still know that  $c$  is suspicious, so they do suspect that  $c$  suspects something. But (due to his wiretapping)  $c$  knows all this (including their suspicion about his suspicion).



**5. Other Examples:** Announcements with suspicion of being wiretaped, Discovery of a wiretap, Lying, Communication over unreliable channels, Generalized Suspicion (“paranoia”) etc.

## 4 Semantics and Update

The semantics of our (closed) action-expressions is simple: in section 2, we have already associated to each closed action expression  $\alpha$  an object  $\|\alpha\| =: ((ALT(\alpha), (\rightarrow_a)_{a \in Ag}, PRE^\alpha), \alpha)$ . This is easily seen now to be an epistemic action, in which the tokens are all the “epistemic alternatives” of action-expression  $\alpha$ ,  $\rightarrow_a$  and  $PRE^\alpha$  are the syntactic relations and function defined in section 2 by induction on the complexity of our expressions, and the designated token is the expression  $\alpha$  itself. This defines an *interpretation function*  $\|\cdot\| : Act_L(\emptyset) \rightarrow Act$ , which assigns an epistemic action to each closed action-expression. This function gives the semantics for action expressions and it obviously has the following properties:  $\|\alpha\| \rightarrow_a \gamma$  iff  $\gamma = \|\beta\|$  for some  $\beta \in ALT(\alpha)$  s.t.  $\alpha \rightarrow_a \beta$ ; and  $PRE(\|\alpha\|) = PRE(\alpha)$ . This allows us to use systematic ambiguity (as we have already done), by using the same variables  $\alpha, \beta, \dots$  for both actions and closed action-expressions, and the same notations  $\rightarrow_a, PRE$  for both. Observe that we cannot identify  $Act_L(\emptyset)$  with  $Act$ , but only *embed* the first one in the second via the interpretation function  $\|\cdot\|$ . This is because the interpretation of a closed action-expression is always a *finite* action. We shall see that the converse is also true: *every finite action is the interpretation of some action-expression*.

We turn now to the interpretation of our logic  $L$ . Our models are the above mentioned *epistemic states*, i.e. epistemic models with a designated world. We need to define two notions: first, an operation of *update of a state by an epistemic action*, which expresses the effect of applying the action to the given state; and secondly, a semantic relation of *truth (satisfaction) of a proposition in given state*. There is a certain circularity involved: the first notion requires some understanding of the second, since the precondition of an action is a proposition; on the other hand, for defining the second notion (truth) we need some understanding of the first (since in order to check the truth of an “update proposition”  $[\alpha]\varphi$ , talking about something being true *after* applying the action denoted by  $\alpha$  we need to first know how to apply this action to a state). We solve this problem by defining the two notion *simultaneously* by *double recursion on the complexity of the propositions involved* (including the preconditions of our actions). (This is only possible because our language  $L \cup Act_L$  is well-founded, being defined itself by double recursion.)

So we define by simultaneous recursion a partial operation  $update . : Mod \times Act \rightarrow Mod$  mapping pairs of states and actions into states, and a binary relation *satisfaction*  $\models \subseteq Mod \times L$  between states and propositions:

(1). **Update of a State by an Action:** Let us be given an epistemic state  $s = (\mathbf{W}, w_0)$ , based on an epistemic model  $\mathbf{W} = (W, (\rightarrow_a^W)_{a \in Ag}, V^W)$ , and an epistemic action  $\alpha = (\mathbf{K}, k_0)$ , based on an action-structure  $\mathbf{K} = (K, (\rightarrow_a^K)_{a \in Ag}, PRE^K)$ . We say that a *state-token*  $w \in W$  *survives an action-token*  $k \in K$ , if the preconditions of  $k$  are true in  $\mathbf{W}$  at state-token  $w$ , i.e. if  $(\mathbf{W}, w) \models PRE^K(k)$ . (This recursively involves, of course, the notion of “truth”, to be defined below.) Similarly, a state  $s$  (having  $w_0$  as its “top”) *survives action*  $\alpha$  (having  $k_0$  as its top) if  $w_0$  survives  $k_0$ , i.e. if  $s \models PRE(\alpha)$ .

We need to define the effect of applying action  $\alpha$  to state  $s$ . An epistemic action changes only the epistemic aspects of the model (i.e. updates the information possessed by the agents), but *not the facts* of the “real world”. The output will be a new epistemic state  $s.\alpha$ . But first we need to check if the given action is *possible at all* in the given state. This happens if the state survives the action, i.e. if  $s \models PRE(\alpha)$ . So our function  $.(update)$  will be a partial

function from  $Mod \times Act$  to  $Mod$ , having as domain the class of all pairs  $(M, \alpha)$  such that  $M$  survives action  $\alpha$ . Our actions will be deterministic. Hence, each “possible action” (i.e. each action-token  $k \in K$  applied to each “possible state”  $w \in W$  gives a possible output-state  $w.k$ , provided that the action is indeed possible at that state, i.e. provided that  $w$  survives  $k$ . If, given the actual state  $w$  and the actual action  $\alpha$ , some agent  $a$ , with her limited information, thinks the input-state might be  $w'$  and that the action happening might be  $\alpha'$ , then she will think that the resulting (output-)state might be  $w'.\alpha'$ . This leads to the following definition: The *update map*  $\cdot$  is a partial function with the above-mentioned domain:  $s.\alpha$  is defined iff  $s \models PRE(\alpha)$ . For  $s$  and  $\alpha$  given as above, satisfying  $s \models PRE(\alpha)$  the update map is defined by  $s.\alpha =: (\mathbf{W.K}, w_0.k_0)$ , with  $\mathbf{W.K} = (W.K, (\rightarrow_a)_{a \in Ag}, V)$ , where:  $W.K =: \{(w, k) : w \in W, k \in K, (\mathbf{W}, w) \models PRE^K(k)\}$ ,  $(w, k) \rightarrow_a (w', k')$  iff both  $w \rightarrow_a^W w'$  and  $k \rightarrow_a^K k'$ , the “top” (actual state) of  $s.\alpha$  is  $w_0.k_0 =: (w_0, k_0)$ , and, finally,  $V(w, k) =: V^W(k)$  (i.e. epistemic actions do not change the truth of atomic propositions, taken to represent the “facts” of the actual world).

In this way, every action  $\alpha$  is seen to induce a (partial) *epistemic update function*  $F_\alpha$  from (the class of)  $Mod$  of epistemic states (which survive  $\alpha$ ) to epistemic state, function given by:  $F_\alpha(s) =: s.\alpha$ .

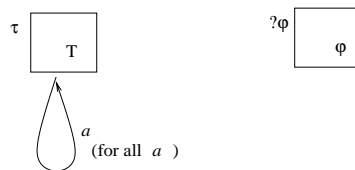
(2). **Truth:** For propositions  $\varphi \in L$  and for epistemic states  $s \in Mod$ , we define by recursion a relation  $s \models \varphi$ . For propositional and modal operators we have the usual recursive conditions: For atomic sentences,  $s \models P$  iff  $P \in V^s$ ; for negation,  $s \models \neg\varphi$  iff  $s \not\models \varphi$ ; for conjunction,  $s \models \varphi \wedge \psi$  iff  $s \models \varphi$  and  $s \models \psi$ ; for belief,  $s \models \Box_a \varphi$  iff  $s' \models \varphi$  whenever  $s \rightarrow_a s'$ ; for common belief,  $s \models \Box_A^* \varphi$  iff  $s' \models \varphi$  whenever  $s \rightarrow_A^* s'$ . Finally, for the “dynamic” modality:  $s \models [\alpha]\varphi$  iff  $s.\alpha \models \varphi$  whenever  $s \models PRE(\alpha)$ .

**Proposition 4.1** (*Completeness and Decidability:*) *There exists a proof system, which was showed in our paper [BMS] to be complete and decidable for the logic presented there. It is easy to see that the proof can be adapted to the present setting, so that we have the following complete and decidable logic of epistemic actions.*

## 5 A Calculus of Epistemic Actions

1. **Special Actions:** Among special actions, we mention: *the trivial action*  $\tau$ , that does not change any model, so its associated update function  $F_\tau$  is just the identity; ; and the “*test*  $\varphi$ ” action  $?\varphi$  that tests the truth of a proposition  $\varphi$ , by being possible only if  $\varphi$  is true, and in this case it returns the same state; if input-states which make  $\varphi$  do not survive this action.

Structurally, these actions can be defined by: let us fix an arbitrary token  $*$ . Define  $\tau = (\mathbf{K}^\tau, *)$ , where  $\mathbf{K}^\tau = (\{*\}, (\rightarrow_a^\tau)_{a \in Ag}, PRE^\tau)$  is given by: each  $\rightarrow_a^* = \{(*, *)\}$  and  $PRE^\tau(*) = \emptyset$ . Similarly, define  $?\varphi = (\mathbf{K}^{?\varphi}, *)$ , where  $\mathbf{K}^{?\varphi} = (\{*\}, (\rightarrow_a^{?\varphi})_{a \in Ag}, PRE^{?\varphi})$  is given by:  $\rightarrow_a^* = \emptyset$  and  $PRE^{?\varphi}(*) = \{\varphi\}$ . Pictures:



**2. Sequential composition of actions:** We define a partial operation  $\circ$  (“composition”) on actions, which will turn out to act on models exactly as a sequential composition of actions. Let  $\alpha = ((K, (\rightarrow_a^K)_{a \in Ag}, PRE^K), k_0)$ ,  $\beta = ((K', (\rightarrow_a^{K'})_{a \in Ag}, PRE^{K'}), k'_0)$  be actions. We define the *action composition*  $\alpha \circ \beta = ((K \times K', (\rightarrow_a)_{a \in Ag}, PRE), (k_0, k'_0))$ , where:  $(k, k') \rightarrow_a (l, l')$  iff  $k \rightarrow_a l, k' \rightarrow_a l'$ ; and  $PRE(k, k') =: PRE^K(k) \cup [\alpha]PRE^{K'}(k')$ .

One can check now that we have  $s.(\alpha \circ \beta) = (s.\alpha).\beta$ , and so indeed  $F_{\alpha \circ \beta} = F_\beta \circ F_\alpha$ , which shows that the action  $\alpha \circ \beta$  has indeed the same effect as the sequence of  $\alpha$  followed by  $\beta$ .

**3. Suspicion (or Exponentiation) of Actions and States:** We introduce an “atomic suspicion action”, corresponding to the action in which *agent a suspects (regards as a possibility) that some action  $\alpha$  might be happening, when in reality nothing happens* (except for a getting suspicious...). Given action  $\alpha = ((K, (\rightarrow_b^K)_{b \in Ag}, PRE^K), k_0)$  and agent  $a \in Ag$ , we define the action  $\alpha^a$  (read “ $a$  suspects  $\alpha$ ”) by  $\alpha^a = ((K^{(a)}, (\rightarrow_b)_{b \in Ag}, PRE), k_0^a)$ , where:  $k_0^a$  is some *new* action token  $k_0^a \notin K$  (chosen arbitrarily);  $k_0^a \rightarrow_a l$  in  $\alpha^a$  iff  $l = k_0$  and, in rest, for  $k, l \in K$ ,  $k \rightarrow_b l$  in  $\alpha^a$  iff  $k \rightarrow_b^K l$ ;  $PRE(k_0^a) = \emptyset$ ; in rest  $PRE$  stays the same for all other states:  $PRE(k) = PRE^K(k)$  for  $k \neq k_0$ . The same definition can be formally applied to epistemic states (considered as a particular case of actions).

**4. Pointed Sum (or Parallel Composition) of Actions and States** We want to model a notion of *epistemic parallelism*. Intuitively, the more complex Examples of the previous section can be seen as the unfolding in parallel of simpler actions:  $a$  announces  $b$  some proposition  $\varphi$ , while in the same time (if the channel is reliable)  $b$  “hears” the announcement, and in the same time  $b$  trusts  $a$  to be sincere, while in the same time  $c$  wiretaps (or intercepts) the announcement, and in the same time  $a$  and  $b$  suspect they might be wiretaped...

Let  $\alpha = ((K, (\rightarrow_a^K)_{a \in Ag}, PRE^K), k_0)$ ,  $\beta = ((K', (\rightarrow_a^{K'})_{a \in Ag}, PRE^{K'}), k'_0)$  be actions. We define their (*pointed*) *sum*  $\alpha_1 + \alpha_2 = ((K'', (\rightarrow_a)_{a \in Ag}, PRE), k''_0)$ , where:  $k''_0$  is some token (which can be assumed to be *new*, i.e. not in  $K \cup K'$ ; the set  $K''$  of tokens is given by  $K'' = K + K' + \{k''_0\}$ , where  $+$  is the direct sum;  $k \rightarrow_a^{K''} l$  in the sum iff: either  $k \rightarrow_a^K l$  or  $k \rightarrow_a^{K'} l$  or  $(k = k''_0$  and either  $k_0 \rightarrow_a^K l$  or  $k'_0 \rightarrow_a^{K'} l)$ ; the precondition function is given by:  $PRE(k''_0) =: PRE^K(k_0) \cup PRE^{K'}(k'_0)$  for the new “top”  $k''_0$ ,  $PRE(k) =: PRE^K(k)$  for  $k \in K$ ,  $PRE(k) =: PRE^{K'}(k)$  for  $k \in K'$ . Again (by formally considering states as actions), this operation induces a similar operation  $s + s'$  on states.

Intuitively, the pointed sum is obtained by taking disjoint copies of the two structures and identifying their “tops”. As intended, the pointed sum of two actions is a sort of parallel composition, or “sum”, of the suspicions involved in the two actions: one first checks that the preconditions of  $\alpha_1$  and  $\alpha_2$  are both satisfied, and then each agent regards as possible (alternative) actions everything that he would regard so in either  $\alpha_1$  or  $\alpha_2$ . But this simultaneous composition of the suspicion actions can be also regarded as an “*alternative*” (*disjunctive*) *sum of the knowledge actions*: in the sum  $\alpha + \beta$ , the agent’s knowledge decreases (or stays the same), the number of epistemic possibilities increases.

**5. Bisimulation Equivalence of Actions and States.** Our notion of equivalence for actions and states is given by bisimulation. The *bisimilarity relation* is the largest relation  $\sim \subseteq Act \times Act$  between epistemic actions such that, if  $\alpha \sim \beta$  then we have the following conditions: (i).  $PRE(\alpha) \cup \{T\} = PRE(\beta) \cup \{T\}$ , where  $T$  is the universally true proposition; (ii). for every  $\alpha' \leftarrow_a \alpha$  there exists  $\beta' \leftarrow_a \beta$  s.t.  $\alpha' \sim \beta'$ ; (iii). for every  $\beta' \leftarrow_a \beta$  there exists  $\alpha' \leftarrow_a \alpha$  s.t.  $\alpha' \sim \beta'$ . (The first condition (i). can be explained as identity of the set of preconditions, disregarding the trivial precondition  $T$ .) The same notion of bisimilarity can be

defined, in particular, *for states*. One can easily check that *bisimilarity is a congruence with respect to the above operations*. One can also check that, modulo bisimulation, the interpretation function  $\|\cdot\|$  sends the syntactic operations defined in section 2 to their semantic counterpart, and similarly that the interpretation of a fixed-point expression  $\mu x.\alpha(x)$  is indeed a fixed-point-modulo-bisimulation of the associated operator.

**Some relevant Identities:**

$$\begin{array}{llll}
s.(\alpha + \beta) & \sim & s.\alpha + s.\beta & \\
(s + t).\alpha & \sim & s.\alpha + s.t & \\
s.(\alpha \circ \beta) & \sim & (s.\alpha).\beta & \\
\alpha \circ (\beta + \gamma) & \sim & \alpha \circ \beta + \alpha \circ \gamma & \\
(\beta + \gamma) \circ \alpha & \sim & \beta \circ \alpha + \gamma \circ \alpha & \\
\alpha + \alpha & \sim & \alpha & \\
\alpha + \beta & \sim & \beta + \alpha & \\
(\alpha + \beta) + \gamma & \sim & \alpha + (\beta + \gamma) & \\
(\alpha \circ \beta) \circ \gamma & \sim & \alpha \circ (\beta \circ \gamma) & \\
\alpha^a \circ \beta^a & \sim & (\alpha \circ \beta)^a & \\
\alpha^a \circ \beta^b & \sim & ?T & \text{for } a \neq b \\
s^a.\alpha^a & \sim & (s.\alpha)^a & \\
s^a.\alpha^b & \sim & ?T & \text{for } a \neq b \\
s.\tau & \sim & s & \\
?T + \alpha & \sim & \alpha & \\
\tau \circ \alpha & \sim & \alpha \circ \tau & \cong \alpha
\end{array}$$

**Proposition 5.1 (“The First Representation Theorem”):** *Every finite action  $\alpha$  is bisimilar to a finite sum of “test” actions (of the form  $? \varphi$ ) and “suspicion” actions (of the form  $\beta^a$ ). More precisely, such a representation is given by:  $\alpha \sim \sum_{\varphi \in PRE(\alpha)} ?\varphi + \sum_{a \in Ag} \sum_{\beta \text{ s.t. } \alpha \rightarrow_a \beta} \beta^a$ .*

However, this representation does *not* provide us with a way to recursively construct any action using only the operations sum, test and exponential. The reason is that, in the above representation, the terms  $\beta$  involved can be equal to  $\alpha$  itself (or to sums involving  $\alpha$ ). An example is the normal representation of the trivial action:  $\tau \sim (? \tau) + \sum_{a \in Ag} \tau^a$ . To obtain a recursive representation of actions we need the fixed-point operator.

**Proposition 5.2 (“Recursive Representation Theorem”):** *Every finite action is bisimilar to one which is recursively built using only tests  $? \varphi$ , sums  $\sum$ , exponentiation  $\alpha^a$  and fixed points  $\mu$ . More precisely: every finite action is bisimilar to the interpretation  $\|\alpha\|$  of some closed action-expression, which does not involve any occurrence of sequential composition  $\circ$ .*

**An Analysis of a Modified “Muddy Children” Puzzle:** As an application of our logic, I give a analysis of a “modified muddy children” puzzle, similar to the one given by Gerbrandy to the classical version of this puzzle. There are four children  $a, b, c, d$ , the first three are the muddy ones. Each can see the others but not himself. The father comes and says publicly: “At least one of you is muddy”. Then they play a game, in rounds. In each round they all simultaneously announce publicly one of the following: “I know I am muddy”, “I know I am not muddy”, “I don’t know (whether I am muddy or not)”. After many rounds (say four for

convenience), the game stops. The ones who gave a correct “definite” answer (“muddy” or “not”) win (say 10 points), the ones who gave a wrong answer lose (-10 points), the ones who still don’t know finish with 0 points.

In the classical puzzle, it can be proved that in certain assumptions (namely, that it’s common knowledge that all children are sincere in their answers, that they are “good logicians” and that they do not “take guesses”, but they answer only they know it) then all the dirty children win in three rounds and the others win in the fourth run. But one of the not so easily observable assumption is the absence of secret communications. Even if the children are sincere and do not “cheat” by lying or guessing, there are some more subtle forms of cheating. Let’s suppose for instance that, after the first round (but before the third), children  $a$  and  $b$  (very good friends, trusting and helping in each other even at the price of cheating, because... a friend in need is a friend indeed...) decide to “cheat” by sending each other secret signals to communicate the message: “You are dirty”. Naturally, in the second round, they both answer “Yes, I know I am dirty” and win. Child  $c$  is also a very trustful person, so trustful that she cannot imagine that such a dirty and secret communication between her dirty colleagues could have taken place. So, in the third round, she is confused: thinking that  $a$  and  $b$  used only their reasoning abilities to answer, she concludes that (the only way for this to have happened is if)  $a$  and  $b$  were the only dirty ones. So she hurries to answer “I know I am not muddy” and she loses! The fourth child  $d$  is the only “clean” one, and he has two possibilities: he either “gets suspicious”, i.e. starts entertaining the possibility (which soon becomes a certainty, after  $c$ ’s wrong answer) that  $a$  and  $b$  cheated; or he could still go on and think this is impossible. In the first case, his action of suspicion will help him to win in the end: after the third round, he gets convinced that  $a$  and  $b$  cheat, that  $c$  is deluded and that himself ( $d$ ) is clean, which he actually will say in the fourth run, winning! But in the second, he will “go crazy”: he will never understand what happened: after the third run, his set of beliefs is not only false, but is actually inconsistent!

Let, for each agent  $i \in Ag = \{a, b, c, d\}$ ,  $P_i$  be some atomic sentence, meaning “ $i$  is dirty”. As before,  $\tau$  is the trivial action. Following Gerbrandy, I make the following abbreviations:

$$\begin{aligned}
dirty_I &= \bigwedge_{i \in I} P_i \wedge \bigwedge_{i \notin I} \neg P_i \text{ (“} I \text{ is the set of all dirty kids”)} \\
Yes_i &= \Box_i P_i \text{ (“} i \text{ knows he is dirty”)} \\
No_i &= \Box_i \neg P_i \text{ (“} i \text{ knows he is not dirty”)} \\
Yes_I &= \bigwedge_{i \in I} Yes_i \\
No_I &= \bigwedge_{i \in I} No_i \\
?_i &= \neg Yes_i \wedge \neg No_i \text{ (“} i \text{ has no clue”)} \\
father &= P_a \vee P_b \vee P_c \vee P_d \text{ (“one of you is dirty”)} \\
\mathcal{L}_I \varphi &= \mu x. (? \varphi + \sum_{i \in I} x^i) \text{ for } I \subseteq Ag \text{ (“the agents in group } I \text{ learn } \varphi \text{ in common”)} \\
vision &= \bigwedge_{i \neq j \in Ag} ((P_i \rightarrow \Box_j P_i) \wedge (\neg P_i \rightarrow \Box_j \neg P_i)) \\
cheat_{a,b} &= \mu x. (? (P_a \wedge P_b) + x^a + x^b + \tau^c + x^d + (\mu y. (y^a + y^b + \tau^c + y^d + x^d))^d)
\end{aligned}$$

“Vision” says everybody sees the others. “ $Cheat_{a,b}$ ” is the full description of the cheating action. If  $Ag = \{a, b, c, d\}$  then the following is a theorem of our logic,:

$$\vdash dirty_{a,b,c} \wedge \Box_{Ag}^* vision \rightarrow [\mathcal{L}_{Ag} father][\mathcal{L}_{Ag} No][Cheat_{a,b}][\mathcal{L}_{Ag}(Yes_{a,b} \wedge No_{c,d})]\Box_c \neg P_c,$$

which proves that after the first two rounds (with the intermediate cheating)  $c$  comes to (falsely) believe that she is clean. This explains her answer in the third round. One can also prove a theorem that says that after this third round,  $d$  knows he is clean.

### References (short list):

- [BMS ] A. Baltag, L.S. Moss, S. Solecki, *The Logic of Public Announcements, Common Knowledge and Private Suspicions*, to be published. Presented at TARK'98.
- [FHMV ] R. Fagin, J. Halpern, Y. Moses, M. Vardi, *Reasoning about Knowledge*, MIT Press, 1995
- [GG ] J. Gerbrandy, W. Groeneveld, *Reasoning about information change*, JLLI 6 (1197) 147-169
- [G ] J. Gerbrandy, *Bisimulations on Planet Kripke*, Ph. D. Dissertation, University of Amsterdam, 1999.
- [V ] F. Veltman, *Defaults in Update Semantics*, JPL, 25:221-261, 1996