Logics for Insecure Communication

Alexandru Baltag*

Abstract

We present a general setting for dynamic-epistemic logics for communication, which can express *announcements* of different types (public or private, truthful or not, reliable or not, secure or not etc.), *queries* of the corresponding types, complex *dialogues, dialogue games, strategies (protocols) for communication* etc. We analyze some examples, we discuss various properties of dialogues (normalcy, responsiveness, publicity, truthfulness, appropriateness of questions) which usually are tacitly assumed, and we formally define interesting types of dialogue which break these assumptions (rhetorical questions, cheating questions, Socratic dialogues, cheating by impersonation). We give an algorithm for computing the beliefs of the agents at the output-states, given their beliefs at the input-state and the the dynamic-epistemic features of the communication act. In the full paper, we give a complete and decidable axiom-system for each such logic.

1 Introduction.

In this paper we attempt to formalize a general notion of *communication actions* in multi-agent systems. Roughly speaking, these are actions that either affect directly the information states of the agents, or enable other such communication actions, but do not affect the "objective facts" of the world. We build on ideas that we have developed in our previous papers [BMS1], [BMS2], [B1], [B2] and [B3]. There we have introduced a notion of *epistemic actions*[1], and used a *combination of dynamic and epistemic logic*, to describe the epistemic features of communication acts such as *public announcements, secret messages, communication with a suspicious outsider, secret interception of messages by the outsider* etc. In this paper, we enrich our setting with a more expressive syntax, which allows us to cover, not just *anonymous announcements*, but also *signed, addressed, time-stamped announcements*, as well as *questions* of various types, and more generally *dialogues*.

Another difference between this paper and our previous approaches is that we derive communication acts from basic *communication types*, which formally are nothing but n-ary propositional connectives, endowed with some extra-structure: most importantly, for each agent we are given some "epistemic" accessibility relation on the set of communication types, which tells how each type "looks" to each agent. In addition to the epistemic structure, each communication type is endowed with a set of *precondition indexes* and a *content*. As we shall see, the first will indicate the possibility conditions for the resulting communication act. The content of an communication type will be a "message-type", composed of a "mode" (assertive or interrogative), a time-stamp, an "issue-index" (giving the index of a sentence in the list, which constitutes the "issue" of this communication), the names of the sender and of the intended recipient, and a list of other agents that are expected to read the message. The semantics for questions themselves is in terms of partitions, which is a rather standard approach[2]: putting a question "opens

*CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands. Email: abaltag@cwi.nl. Website: http://www.cwi.nl/~abaltag

[1] Maybe *doxastic* actions would be a more accurate name, but it doesn't sound so great. It's true that we didn't make any effort to restrict ourselves to $S5$, and so the only way we can talk about "knowledge" is as "true belief". There is no problem in restricting the present approach to $S5$, it is just that we *are interested* in "abnormal" situations of cheating, lying and being deceived.

[2] See [GS1], [GS2], [GSV].

up" a pair of binary partitions, one which divides the possible worlds according to whether or not the question was raised in them, while the second partition divides the worlds according to whether or not the correct answer to the question is "yes" or "no". In addition, as any message, a questioning message has a sender and a list of receivers, which form the intended "audience". [3]

Example 1 Our starting example concerns a "love triangle": suppose that Alice and Bob are a couple, but Alice has just started an affair with Charles. At some point, Alice sends to Charles an email, saying: "Don't worry, Bob doesn't know about us". But suppose now that Bob accidentally reads the message. Then, paradoxically enough, after seeing the message which says he doesn't know..., he will know! So, in this case, *learning the message is a way to falsify it.* [4]

Example 2 Let us modify the previous example to introduce "cheating regarding cheating". Suppose that in fact Alice was faithful, despite all the attempts made by Charles to seduce her. Out of despair, Charles comes up with a cool plan of how to break up the marriage: he sends an email which is identical to the one in the previous example, bearing Alice's signature and telling Charles not to worry about Bob; moreover, he makes sure somehow that Bob will have the opportunity to read the message. As a result, Bob will be mislead into "knowing" that he has been cheated.

Example 3: As observed a long time ago, the well-known Muddy Children Puzzle poses a problem to both naive belief-revision and naive semantics of dialogue. Suppose there are 4 children, exactly 3 of them have dirty faces, and each can see the faces of the others, but doesn't see his/her own face. The father publicly announces "One of you is dirty". But, apparently, this message is not informative to any of the children: the statement was already known to everybody![5] Then the father does another paradoxical thing: starts repeating over and over the same question "Do you know if you are dirty or not?"; after each question, the children have to answer publicly, sincerely and simultaneously, without taking any guesses. But (if the father knows that his children are good logicians then) at each step the father knows already the answer to his question, before even asking it! In a way, it seems the father's questions are *"abnormal"*, in that they don't actually aim at filling a gap in father's knowledge; but instead they are part of a Socratic strategy of teaching-through-questions.

Example 3: Let us modify the last example a bit. Suppose that the children who answer "Yes, I know" to the previous "knowledge" question are immediately asked to say if they actually are dirty or not. Suppose the children are rewarded for answering as quickly as possible, but they are punished for incorrect answers; suppose also that, after the second round of questions, two of the dirty children "cheat" on the others by secretly announcing each other that they're dirty, while none of the others suspects this can happen. Then one can easily see that the third dirty child will be totally deceived, coming to the "logical" conclusion that... she is clean! So she ends up by being punished for her credulity, despite her impeccable logic.

We are looking for a logic which can express all the above situations and the communication acts involved (including the questions), and which can faithfully express the reasoning in these examples (including "learning-by-questions", "being-deceived-by-logic" and "being-saved-by-suspicion"). The basic idea underlying all our papers on the subject is that *in order to understand such "belief-changing actions" we need to formalize the "belief components" of the actions themselves.* We do this by endowing them with an *internal epistemic structure*: the action's *appearance to each agent a* is described using

[3]The rest of the pragmatic features of the questioning act (whether or not the inquirer knows the answer beforehand, if he/she expects a truthful, or at least sincere, answer, if the question was secret or was overheard by outsiders, what is the view of every agent on what the question was or even if it was a question at all, and on what "channel" is the answer expected to be received, on a public channel or in a private communication etc etc) - all these details will be captured by the above-mentioned *internal epistemic structure* of the associated communication type.

[4]A different formulation of this paradox was first introduced by Jelle Gerbrandy in [G]. This example shows that standard belief-revision postulates may fail to hold in such complex learning actions, in which the message to be learned refers to the knowledge of the hearer.

[5]As it is well-known though, this announcement adds information to the system: the children implicitly learn some new fact, namely the fact that what each of them used to know *in private* is now *public knowledge.*

Kripke accessibility relations $\stackrel{a}{\rightarrow}$, between an action and its possible alternative actions[6] suspected by agent a. The structure of our actions is similar to the one of the communication types, in that it contains two more components, besides the epistemic structure: the action's *set of preconditions* and the action's *content*. The first is a set of sentences, describing the conditions of applicability for the given action: not every action can happen in every world. The second component gives the actual "content" of the action, what is "really" going on. In former papers, we have considered actions which *change facts of the world*, in which case the action's content was a description of this change. Here (as in [BMS1] and [BMS2]), we do not consider fact-changing actions, but only *purely epistemic* actions. Aside from the desire for simplicity, our reason for this restriction is that we want to focus here on *communication actions*. So we consider the content of a basic action to be given by a *message*, i.e. a complex object indicating what is the issue (i.e. a sentence), but also who sent the message to whom, who else was supposed to read the message [7], and whether it was meant to be a question or an assertion. We model the *update of a state by an action* as an operation of *"conditional multiplication" of the two Kripke structures* (the static and the dynamic one): the space of output-states will be taken to be a subset of the Cartesian product of the two structures, in which we have deleted the "impossible pairs", i.e. the pairs (s, α) arising from input-states s which did *not* fulfill some precondition of the action α. We endow this set of output-states with a Kripke structure, by taking the "product arrows": $(s, \alpha) \rightarrow_a (t, \beta)$ iff $s \rightarrow_a t$ and $\alpha \rightarrow_a \beta$. This reflects the idea of "multiplicating independent uncertainties": the uncertainties of each agent regarding the current state and the uncertainties regarding the current action are assumed to be independent, in the case of simple actions, so they have to be 'multiplied' to obtain the uncertainty regarding the output state. In addition to this, we record both *the fact* that a specific message (assertion or question) was sent and whether or not its "issue" was in fact true[8], using some special atomic sentences $!_i j(a, b, B)$, $?_i j(a, b, B)$ and $y_i j(a, b, B)$. We mention here that we have a complete decidable system for the resulting dynamic-epistemic logic, the completeness proof being a rather straightforward modification of the ones for our previous logics in [BMS], [BMS2], [B1], [B2]. We use this logic to characterize various types of dialogue, various rules and conventional assumptions of such dialogues: truthful communication, sincerity of answers, "normality" , "responsiveness" ,"politeness" and "appropriateness" of questions etc. We analyze specific examples of dialogue games, such as "Socratic" ones, in which questions are used to increase the knowledge of the audience (while is the answer is already known by the agent asking the question). We analyze the Muddy Children Puzzle as an example of such a Socratic dialogue; we also analyze some other versions of the puzzle, in which some agents can "cheat", either by lying or by engaging in "secret" communication.

2 The Syntax.

We introduce here modal languages to talk about communication actions. We assume as given a set $AtSen$ of atomic sentences, denoted by $p, q, \ldots$, and a finite set Ag of agents, denoted by $a, b, \ldots$. We use capital letters $A, B, \ldots \subseteq Ag$ to denote *finite sets of agents*.

Communication Signatures. Our syntax will be given in terms of an *communication signature*, which will provide the *basic types of communication acts* allowed in the language. In fact, we do not restrict ourselves to one language, but we study *classes of languages*, generated by various signatures. An communication signature consists of the following: **(1)** for each $n \in N$, a *finite* set Σ_n of propositional connectives of arity $n+1$, called *communication types*; these sets Σ_n are assumed to be *mutually disjoint*;

[6]This is completely similar to the way such epistemic accessibility relations are usually used to represent agent's uncertainty concerning the current state of the system: it is just that here we represent agent's uncertainty concerning the *current action* taking place.

[7]Think of the list of names receiving carbon-copies of an email. This gives the "intended audience": if communication is reliable, then there will be common knowledge of this message among the sender, the recipient and the agents in this list.

[8]In the case of an assertion, this means the claim made was true; in the case of question, it means the correct answer was "yes".

(2) for each $\sigma \in \Sigma_n$ and each agent $a \in Ag$, some set $\sigma_a \subseteq \Sigma_n$, called *the appearance of σ to agent a* (while the types $\sigma' \in \sigma_a$ are called *agent a's epistemic alternatives for the type σ*);**(3)** for each $\sigma \in \Sigma_n$, some set $PRE_\sigma \subseteq \{1, \ldots, n\}$ of *precondition indices*; **(4)** finally, we are given a partial function[9] CON from Σ to $\{?,!\} \times N \times \{1, \ldots, n\} \times Ag \times Ag \times \mathcal{P}(Ag)$, associating a *content* to some of the communication types; the content is a tuple (ϵ, i, j, a, B), also written as $\epsilon_i j(a, b, B)$ and called *messages-type*; in such a message-type, $\epsilon \in \{?,!\}$ is called the *mode* of the message (indicating either a question or an assertion), $i \in N$ is the *time stamp* (indicating the time of the action[10]), $j \leq n$ is the *issue index* (that will point to a sentence which forms "the issue", via pointing to the index of this sentence in a given list); the agent $a \in Ag$ is called *the sender of the message* (or the *speaker*), the agent b is the *main recipient* (to whom the message is addressed) and the group $B \subseteq Ag$ gives *the list of the other (secondary) recipients* (the "intended audience" of this message). These components are required to satisfy two extra-conditions, to be listed below. We put $\Sigma = \bigcup_{n \in N} \Sigma_n$ to be the set of all our communication types, and we shall refer to the whole signature as to Σ, keeping the rest of the structure implicit. For each agent a, we can also define *epistemic accessibility relations*[11] $\stackrel{a}{\rightarrow} \subseteq \Sigma_n \times \Sigma_n$ between communication types, by putting: $\sigma \stackrel{a}{\rightarrow} \sigma'$ iff $\sigma' \in \sigma_a$. The above-mentioned conditions are: **(*) Full Introspection.** $\sigma'_a = \sigma_a$ for every $\sigma' \in \sigma_a$; **(**) Conscious Communication:** if $\epsilon k(a, B) \in CON_\sigma$ then also $\epsilon k(a, B) \in CON_{\sigma'}$, for every $\sigma' \in \sigma_a$. [12]

The Language. In addition to the given atomic sentences in $AtSen$ we assume, for each pair $(a, b, B) \in Ag \times Ag \times \mathcal{P}(Ag)$ of an agent and a set of agents, to be given some constant atomic sentences $!_i j(a, b, B)$, $?_i j(a, b, B)$ and $y_i j(a, b, B)$ (for $i \in N$), which will have a special semantical behavior. Put $\overline{AtSent} = AtSen \cup \{\chi_i j(a, b, B) : \chi \in \{?, !, y\}\}$. As usually, we assume as given the standard propositional connectives and the epistemic modalities. The additional feature is that, given an communication signature Σ, each communication type $\sigma \in \Sigma_n$ is a new *propositional connective* of arity $n + 1$, written in a special notation, namely as $[\sigma\ \varphi_1, \ldots, \varphi_n]\psi$. We define the set $L[\Sigma]$ of *sentences* over the signature Σ by recursion, as the least set which includes all atomic sentences $p \in AtSent$, $!_i j(a, b, B)$, $?_i j(a, b, B)$ and $y_i j(a, b, B)$, and is closed under negation $\neg$, conjunction $\wedge$, the standard epistemic modalities $\Box_a \varphi$ ("belief") and $\Box^*_A \varphi$ ("common knowledge"), and under all the propositional connectives $\sigma \in \Sigma$ in our signature: if $\sigma \in \Sigma_n$ and $\varphi_1, \ldots, \varphi_n, \psi$ are in $L[\Sigma]$ then $[\sigma\ \varphi_1, \ldots \varphi_n]\psi$ is in $L[\Sigma]$.

We make the notation $Ans_i j(a, b, B) =: \{y_i j(a, b, B), \neg y_i j(a, b, B)\}$ to be the *set of possible answers* (yes,no) to the questioning message $?_i j(a, b, B)$ sent by agent a to agent b.

Communication Acts and Communication Sequences. The syntactic objects of the form $\sigma\ \vec{\varphi}$, which appear inside brackets as dynamic modalities in the above definition, are called *communication acts*. We denote by Act_Σ the set of all such actions. We read the sentence $[\sigma\ \vec{\varphi}]\psi$ as saying that "after the communication act $\sigma\ \vec{\varphi}$ is performed at the current state, sentence ψ will be true at the output state". A *communication sequence* (or an 'abstract dialogue') is a "word" $\alpha \in Act^*_\Sigma$ over this alphabet, i.e. a finite sequence of communication acts. We can naturally extend the notation $[\alpha]\varphi$ from basic actions to communication sequences, by the following abbreviation: put $[\alpha_1 \alpha_2 \ldots \alpha_n]\varphi =: [\alpha_1][\alpha_2] \ldots [\alpha_n]\varphi$, if all $\alpha_i \in Act_\Sigma$. The meaning of the sentence $[\alpha]\psi$ is that "after the whole communication sequence α is performed on the current state, sentence ψ will be true at the output state". The *accessibility* ("uncertainty") relations $\stackrel{a}{\rightarrow} \subseteq \Sigma \times \Sigma$ between communication types can be extended in a natural manner to accessibility relations $\stackrel{a}{\rightarrow} \subseteq Act^*_\Sigma \times Act^*_\Sigma$ between communication acts and sequences, by putting: $\sigma\ \vec{\varphi} \stackrel{a}{\rightarrow} \sigma'\ \vec{\varphi}$ whenever we have $\sigma \stackrel{a}{\rightarrow} \sigma'$; and $\alpha_1 \alpha_2 \ldots \alpha_n \stackrel{a}{\rightarrow} \alpha'_1 \alpha'_2 \ldots \alpha'_n$ whenever all $\alpha_i, \alpha'_i \in Act_\Sigma$

[9] We make it partial in order to allow communication acts with no content, in which *no message is sent.*

[10] We assume all agents have synchronized clocks.

[11] There is no need to label these relations with an extra-subscript n, since the Σ_n's are pairwise disjoint. So we can use systematic ambiguity to denote all these (formally different) relations with the same notation $\stackrel{a}{\rightarrow}$, for each given agent a.

[12] The first condition says that the relations $\stackrel{a}{\rightarrow}$ are Euclidean and transitive, hence endowing our communication types with an $S4$-structure: the agents are assumed to be fully introspective with respect to their actions. The second condition says that agents know what messages they send (including their mode, the issue and the intended audience). We do *not* require similar knowledge conditions for the recipients of the message, since we wish to cover the cases when the *(potentially) unsuccessful* acts of communication.

are s.t. $\alpha_i \stackrel{a}{\to} \alpha_i'$. Similarly, we use the set PRE_σ of precondition indices to define *the set of presuppositions* of a communication act, and we can then extend this function to communication sequences by recursion: $PRE_{\sigma \cdot \varphi} =: \{\varphi_i : i \in PRE_\sigma\}$; $PRE_{\alpha\cdot\beta} =: \{\varphi \wedge [\alpha]\psi : \varphi \in PRE_\alpha, \psi \in PRE_\beta\}$, where $\alpha \in Act_\Sigma^*$, $\beta \in Act_\Sigma$ and $\cdot$ is the operation of concatenation. As announced earlier, PRE gives us the action's "conditions of possibility". The set PRE is always a finite set of sentences, so we can take its conjunction to form a single sentence $pre_\alpha = \bigwedge PRE_\alpha$, which defines the *domain of applicability* of action (communication sequence) α. In a similar manner we can use the content of an communication type to define *the content of a communication act or sequence*: $CON_{\sigma \cdot \varphi} =: \epsilon_i\varphi_j(a,b,B)$ iff $CON_\sigma = \epsilon_i j(a,b,B)$; $CON_{\alpha\cdot\beta} =: CON_\alpha \cdot CON_\beta$ (where $\cdot$ is word concatenation). The content of a communication act is called a *message*.

Examples:

"Target Logics".[13] The following examples are very natural kinds of announcements, the associated logics being given as examples of our "target logics' in [BMS2]. The communication type in which *nothing happens* is of the type *skip*, defined by putting: $\Sigma_0 = \{skip\}, skip_a = \{skip\}$ for all agents A, $PRE_{skip} = \emptyset$, while CON_{skip} is undefined. The only communication act of this type is $skip =: skip()$, having $pre_{skip} = true$, $skip_a = skip$ (and CON_{skip} undefined). The communication type *truthful public announcement by an agent a* (with the default assumption of successful communication) is the type $Pub_i(a,b)$ of actions in which a truthful message is send (with time-stamp i) by agent a to agent b over a public channel (immediately accessed by all agents). The signature is essentially given by a one-point loop for every agent: $\Sigma_1 = \{Pub_a\}, Pub_i(a,b)_c = \{Pub_i(a,b)\}$ for all agents $c \in Ag$), $PRE_{Pub_i(a,b)} = \{1\}$, $CON_{Pub_a} = !_i 1(a,b,Ag)$. Given any sentence φ, this type generates a communication act $Pub_i\varphi(a,b)$ of "agent a publicly (and truthfully) announcing to b that φ (with time-stamp i)", action with the following structure: $(Pub_i\varphi(a,b)_c = \{Pub_i\varphi(a,b)\}$ for every $c \in Ag$ (i.e. this is totally public, fully transparent action), $PRE_{Pub_i(a,b)} = \{\varphi\}$ (i.e. this action can happen iff φ is actually true) and $CON_{PUB_i(a,b)} = !_i\varphi(a,b,Ag)$ (i.e. this action consists in a truthfully broadcasting the message φ addressed to b, with time stamp i). Similarly, for each set $B \subseteq Ag$ of agents, we can consider the communication type $Pri_i(a,b,B)$ of *completely private (and completely secure, truthful, believable etc.) announcement addressed by agent a to agent b and broadcasted also to group B* (with the implicit assumptions that the communication is successful, reliable and secure: it is common knowledge inside the group $B \cup \{a,b\}$ that the message was sent, but none of the outsiders $c \notin B\{a\}$ suspects what happened: they think nothing happened). This type is given by: $\Sigma_1 = \{Pri_i(a,b,B), skip^1\}$, with $Pri_i(a,b,B)_c = \{Pri_i(a,b,B)\}$ for all $c \in B \cup \{a,b\}$, $Pri_i(a,b,B)_c = \{skip^1\}$ for $c \notin B \cup \{a,b\}$, $PRE_{Pri_i(a,b,B)} = \{1\}$, $CON_{Pri_i(a,b,B)} = !_i 1(a,b,B)$, and with $(skip^1)_b = \{skip^1\}$ for all b, $PRE_{skip^1} = \emptyset$, while CON_{skip^1} is undefined. This type generates communication acts of the form $Pri_i\varphi(a,b,B)$, in which the message φ addressed to b is truthfully broadcasted to group $B \cup \{b\}$ by agent a, while the outsiders don't suspect this happens. [14] The communication type $Prss_i^k(a,b,B)$ of a *private announcement sent by a to b, broadcasted to group B, with secure suspicion of k possible announcements by the outsiders*, has the following signature: $\Sigma_k = \{1,2,\ldots,k\} \cup \{1'2',\ldots,k'\} \cup \{skip^k\}$; $m_c = m_c' = \{1',\ldots,k'\}$ for $c \in B \cup \{a,b\}$; $m_c = \{1,\ldots,k\}$ for $c \notin B \cup \{a,b\}$; $m_c' = \{skip^k\}$ for $c \notin B \cup \{a,b\}$; and finally $(skip^k)_c = \{skip^k\}$ for *all* agents c; $PRE_m = PRE_{m'} = \{i\}$ and $PRE_{skip^k} = \emptyset$; $CON_m = CON_{m'} = !_i m(a,b,B)$, CON_{skip^k} undefined. Many other types of examples are possible (-see [BMS1], and especially [BMS2], for more examples of our target logics). We can also represent *disbelief of a message*, *lying* and *mistaken announcements*, *failed communication, interception of secret messages* ("wiretapping") and *misleading epistemic actions* (e.g. lying and "cheating").

Interrogatives. We can similarly express all various kinds of queries, with epistemic structures mirroring the above ones: for instance, the type $PubQ_i(a,b)$ of "public questioning of agent b by agent a" has exactly the same structure as $Pub_I(a,b)$ above, except that $PRE_{PubQ_i(a,b)} = \emptyset$ and $CON_{PubQ_i(a,b)} = ?_i 1(a,b,Ag)$. The resulting communiction act $PubQ_i\varphi(a,b)$ is the act of *agent a publicly asking b whether φ is true or not.* There are no default assumptions. But using richer signatures, we can

[13] See [BMS1,[BMS2],[D],[GG], [JB].

[14] Observe that, as actions, *skip* and $skip^1$ are equivalent; indeed, as epistemic structures, they will be bisimilar; we only distinguish between them since, formally, *skip* is a unary connective, while $skip^1$ is a binary propositional connective...

express more specific kinds of communication acts, such as the action of *"normal" public questioning*, i.e. the act $PubNQ_i\varphi(a,b)$ of *"agent a publicly asking a question φ, for which he doesn't know the answer, but believes it is possible that somebody in the audience knows it"*: this act has exactly the same structure as $PubQ_a(\varphi)$, except that $PRE_{PubNQ_i\varphi(a,b)} = \{\neg\Box_a\varphi \wedge \neg\Box_a\neg\varphi \wedge \neg\Box_a\neg\Box_a(\bigwedge_{b\in Ag}\neg\Box_b\varphi \wedge \bigwedge_{b\in Ag}\neg\Box_b\neg\varphi)\}$. In a similar manner we can express the corresponding act of *"normal" private questioning without outsider suspicion* (broadcasted by agent a to group B), *"normal" private questioning with secure suspicion*, *"normal" private questioning which is secretly overheard (intercepted) by outsider* etc; but we also can describe various "ab-normal" questioning acts: deceiving questions (e.g. "deceiving rhetorical questions, asked by an agent who knows the answer, but by questioning is successfully inducing in others the belief that he doesn't know the answer); "pedagogical" (Socratic) questions ($\Box_b\varphi$)? ("Do you know if φ?"), with the default assumption that the speaker a knows the answer, knows that nobody in the audience Ag (of all agents) knows how to answer now, but believes that, after putting the question and so triggering a public I-don't-know answer from b, someone in the audience (say d) will learn the answer (from b's ignorance). [15]

3 The Semantics.

In this paper, an *epistemic model* is an a multi-modal $S4$- Kripke model $(W, \stackrel{a}{\to})_{a\in Ag}, V)$, with the accessibility relations $\stackrel{a}{\to}$ being transitive and Euclidean for each agent a, and with the valuations $V : \overline{AtSen} \to \mathcal{P}(W)$ having the property that *only finitely many atomic sentences of the form $\chi_i j(a,b,B)$ are true at any given state* (for $\chi \in \{?,!,y\}$). This last condition ensures that we always have fresh time-stamps. We write $v \models^W p$ iff $v \in V(p)$. An *epistemic state* is just a model-world pair (W,v) of an epistemic model W and a designated world w (called the *current state*). We denote by St the class of all epistemic states. Instead of specifying epistemic relations $\stackrel{a}{\to}$, we can alternatively (but equivalently) define (as in the case of communication acts) an epistemic state as being endowed with some maps $\cdot_a : W \longrightarrow \mathcal{P}(W)$, for each agent a, called *appearance maps* and required to satisfy the condition: $w'_a = w_a$ for every $w' \in w_a$. Indeed, we can take $w_a = \{w' : w \stackrel{a}{\to} w'\}$ to be the set of all its $\stackrel{a}{\to}$-successors (also called a-epistemic alternatives of world w). For a world $w \in W$ in a model, we put $i(w) =: 1 + max\{i : w \models \chi_i j(a,b,B)$, for some $\chi \in \{?,!,y\}, j \in N, a,b \in Ag, B \subseteq Ag\}$ to be the least "fresh" time-stamp. All these notions are in fact relative to the underlying model (so, to be precise, we should write $w \to_a^W w'$ etc.). To make them model-independent, we can "lift" them to the level of epistemic states, by "identifying" an epistemic state with its "root", and transferring the properties of the root to the state. In this way, we can define $s \models p$ iff $s = (W,v)$ for some model (W,V) and world v s.t. $v \models^W p$; $i(s) = i(w)$ iff $s = (W,w)$; and $s \stackrel{a}{\to} s'$ iff $s = (W,v), s' = (W,v')$ for the some model W and worlds s.t. $v \stackrel{a}{\to}{}^W v'$. This allows us to abstractly specify epistemic states s, without explicitly constructing the underlying models, by just giving their truth-conditions for atomic sentences and their sets of *epistemic alternatives* $s_a = \{s' : s \stackrel{a}{\to} s'\}$, for each agent a. [16] Put also, for a state s and a set A of agents, s_A to be the set of all its "iterated A-successors", i.e. of all the states s' which can be reached from s by any finite chain of arrows $\stackrel{a}{\to}$'s, corresponding only to agents $a \in A$.

For a given signature Σ, we simultaneously define a notion of *truth* $\models \subseteq St \times L[\Sigma]$, as a binary relation between epistemic states and sentences, and a partial *update* operation $\cdot : St \times Act_\Sigma \to St$, as a partially defined binary operation taking pairs of epistemic states and communication acts[17] into epistemic states. The *truth-clauses* are simply obtained by adding to standard epistemic logic with common knowledge an extra-clause referring to the dynamic modalities, and saying that meaning of $[\alpha]\varphi$ is that *after updating with α, φ becomes true at the output-state*: the valuation takes care of the truth-clauses for all atomic

[15] This Socratic action has as precondition the sentence: $(\Box_a(\varphi) \vee \Box_a(\neg\varphi)) \wedge \bigwedge_{c\in Ag}\Box_a(\neg\Box_c\varphi \wedge \neg\Box_c\neg\varphi) \wedge \Box_a[Pub_i(\neg\Box_b\varphi \wedge \neg\Box_b\neg\varphi)(b,a)]\Box_d$.

[16] These sets can circularly refer back to the state to be defined.

[17] We can naturally extend the update operation to arbitrary communication sequences, although it is not necessary for our semantics: put by recursion $s.(\alpha \cdot \beta) = (s.\alpha).\beta$. One can easily check that $s \models [\alpha \cdot \beta]\varphi$ iff $s.(\alpha \cdot \beta) \models \varphi$.

sentences (including the constants $q(a,B)$ and $y(a,B)$); $s \models \neg\varphi$ iff $s \not\models \varphi$; $s \models \varphi \wedge \psi$ iff $s \models \varphi$ and $s \models \psi$; $s \models \Box_a\varphi$ iff $s' \models \varphi$ for every $s' \in s_a$; $s \models \Box^*_A\varphi$ iff $s' \models \varphi$ for every $s' \in s_A$; and finally, $s \models [\alpha]\varphi$ iff $s.\alpha \models \varphi$ whenever $s.\alpha$ is defined. The *update of a state s with an action $\alpha \in Act_\Sigma$* is defined *checking first whether the state satisfies the precondition*; then *multiplicating the uncertainties (epistemic arrows) regarding the state with the uncertainties regarding the action* for each agent a, while keeping the valuation the same; and then changing the valuation of the atomic sentences of type $?, !, y$ which record the sending of the message and its truth-value: (1) $s.\alpha$ is defined iff $s \models pre_\alpha$; (2) $(s.\alpha)_a =: \{s'.\alpha' : s' \in s_a, \alpha' \in \alpha_a\}$, and $s.\alpha \models P$ iff $s \models P$ (for $P \in AtSent$ "fact of the world"); (3) if $\alpha = \sigma\,\vec{\varphi}$, put: $s.\alpha \models !_i j(a,b,B)$ iff either $s \models !_i j(a,b,B)$ or $i = i(s), CON_\sigma = !_i j(a,b,B)$; $s.\alpha \models ?_i j(a,b,B)$ iff either $s \models ?_i j(a,b,B)$ but $CON_\alpha \notin \{!_k\varphi(b,a,B) : \varphi \in Ans_i j(a,b,B), k \in N\}$, or if $i = i(s), CON_\sigma = ?_i j(a,b,B)$; $s.\alpha \models y_i j(a,b,B)$ iff either $s \models y_i j(a,b,B)$ or $i = i(s), CON_\sigma = \epsilon_i j(a,b,B), s \models \varphi$ for some $\epsilon \in \{?,!\}$.

The meaning of the last clauses is that: an assertive message is recorded as "announcement made" if either was already recorded as such before the last action or if the last action involved sending this message; an interrogative message is recorded as an "unanswered question" either if it was already recorded as such and the last action did not involve answering it, or if the last action involves raising this question; a "yes" corresponding to a message is true either if it was already true before the last action, or if the last action involved sending this message and the issue raised by the message (i.e. the claim made or the positive answer to the question raised) was true when it was sent. As one can see from this, we record *all* the categorical announcements, but only the questions that were not answered yet.[18] The above semantics considers communication acts as *instantaneous*, although they can *fail*, by not reaching their recipients. But we can still *model time delays*, by concatenating (sequentially composing) unsuccessful communication acts with later actions which "look" to the recipients as (instantaneous) successful communication acts (although in fact nothing is sent at this time). The reason this modeling works is that, although the meaning of a sentence (claim or question) φ changes in "time" (i.e. it may be different in an updated model from the original one), the above semantics allows us to keep track of *the original meaning* of the claims made or of the questions asked, via the time-stamped atomic sentences. This also allows us to correctly model actions which answer questions that were raised several rounds before: once they were raised, the corresponding atomic sentence y_i captures their original extensional meaning (being true at the states where the correct answer was affirmative then); if the recipient cannot answer the question at the time (or if communication fails and he is not aware of being asked the question), but at a later time becomes aware of both the question and its answer, the answer y_i that he sends back answers indeed the original question (and not its current version, which may have a different meaning). In a similar manner, although our language does not formally allow *parallel communication* (i.e. a set of messages and questions being exchanged in the same time), the content of each of our basic communication acts being one single message, we can still model such parallel communication as sequential composition (concatenation) of various acts; we can do this by skillfully manipulating the *time-stamps* and the *epistemic structures* of these acts.[19]

Proposition 3.1: Completeness and Decidability. *For every communication signature Σ, there exists a sound and complete axiomatic proof system for the corresponding logic. In fact, these systems are obtainable in a uniform way from a generic system (with axiom-schemes), by taking all the instances of the axioms which are in the given signature language. The proof method*[20] *implies also that the logic is decidable.*

[18]This is just a trick, allowing us to know when a question is answered: the information concerning the fact that this particular question was raised in this context is not lost, since it is recorded in the form of its "answering" message.

[19]It is not very hard to change the present setting such as to allow parallel communication from the beginning: just take the "content" of a basic action to be a *set of messages*, possibly with different senders and recipients. We chose not to do it only for simplicity reasons, after observing that we can in effect capture all of them in the present setting.

[20]The proof is based on a rather easy adaptation of our argument in the joint paper [BMS]. The presence of questions and of the non-standard atoms only affects the axioms concerning the atomic sentences. The argument uses a Fisher-Ladner style filtration and a rewriting system for the language to establish that the logic has the finite model property.

4 Types of Dialogues.

A *dialogue game*, in our sense, will be a pair $G = (S_G, Act_G)$ of a finite set $S_G \subseteq St$ of *initial epistemic states* and a finite set of $Act_G \subseteq Act_\Sigma$ of communication acts, called *legal (communication) moves.*[21] In a dialogue game G, the set Act_a of the *moves available to agent a* is the set of those moves in which a sends some message: $Act_a = \{\alpha \in Act_G : CON_\alpha = \epsilon_i\varphi(a,b,B)$ for some $\epsilon, i, \varphi, b, B\}$. Given a dialogue game, a *legal dialogue* is a string $s_0, \alpha_0, s_1, \alpha_1, \ldots, \alpha_{n-1}, s_n$, s.t. $s_0 \in S_G$, $\alpha_i \in Act_G$ and $s_{i+1} = s_i.\alpha_i$ for all i. An agent b is *responsive* (and sincere) in a dialogue game to an agent a if all b's available moves α are of one of the following three kinds: (1) moves s.t. $\{\Box_b ?_i j(a,b,B), \Box_B \varphi\} \subseteq PRE_\alpha$ and $CON_\alpha = !_k\varphi(b,a,B)$, for some $i,j,k \in N$, $B \subseteq Ag$, $\varphi \in Ans_i j(a,b,B)$; (2) moves s.t. $\Box_b ?_i j(a,b,B) \in PRE_\alpha$, but $\{\neg\Box_b\varphi : \varphi \in Ans_i j(a,b,B)\} \subseteq PRE_\alpha$ and $CON_\alpha = !_k(\bigwedge\{\neg\Box_b\varphi : \varphi \in Ans_i j(a,b,B)\})(b,a,B)$; (3) moves s.t. $\{\neg\Box_b ?_i j(a,b,B) : i,j,B$ s.t. $Con_\sigma = ?_i j(a,b,B)$, for some $\sigma\, \vec{\varphi} \in Act_a\} \subseteq PRE_\alpha$. In words: an agent b is responsive to a if his available moves, in case he is aware of any questions being asked to him by a in front of an audience B, are sincere answering acts or sincere announcements that he doesn't know the correct answer; and moreover, these answering messages are broadcasted to *the same audience* as the questions.[22]

A questioning act α is *normal* if it has among its preconditions the sentences saying that the sender of the underlying questioning message does not know the answer, and that he considers possible that somebody in the audience knows the answer: i.e. if $CON_\alpha = ?_i\varphi_i(a,b,B)$ then: $\{\neg\Box_a\varphi, \neg\Box_a\varphi\} \subseteq PRE_\alpha$ and also $(\neg\Box_a\neg(\bigvee_{b\in B}(\Box_b\varphi \vee \Box_b\neg\varphi))) \in PRE_\alpha$. Assuming that it is common knowledge that all the questioning moves of a dialogue game are normal can make each of it more "informative" than it would otherwise be: you can learn from listening to a question that the speaker doesn't know the answer and that he considers possible that you may know the answer. Nevertheless, normality can fail, as in the Muddy Children Puzzle.

What makes the Muddy Children Puzzle go is the combination of *truthfulness* with *responsiveness* of all the communication acts, but also the *public quality* of everything that goes on: our Example 3 shows that a basic default assumption in the classical Puzzle is the absence of any secret communication moves. This makes possible the father's "Socratic strategy". A question $?\varphi(a,B) \in CON_\alpha$ raised in an communication act α is said to be *abnormal* in this act if the first condition of normality fails: the speaker knows the answer, and moreover this is part of the default presupposition of this action: $(\Box_a\varphi \vee \Box_a\neg\varphi) \in PRE_\alpha$. The question is *rhetorical in* α if both normality conditions fail, and moreover this failure is part of the Precondition. A questioning act is *Socratic* if the underlying question is rhetorical in this act, but (in the context of the given dialogue game) putting the question ensures that, after the next action, somebody *will* know the answer. A questioning sequence(i.e. a string of questioning acts) is Socratic if (given the dialogue game), performing them in a sequence, whenever is possible, ensures that at the output some agent will know the answer. [23] Another way that normality can fail is through "cheating questions". In the Muddy Children, the failure of normality is public knowledge, nobody assumes that the father doesn't know who's dirty, although he keeps repeating the question. Being public knowledge, such a failure of normality is "legal", in a way: non-deceiving. But

[21] Compare these notions with the related ones in [H],[GSV], [GG].

[22] There is a very often used convention in normal communication, which enforces this rule: the implicit assumption is that a question addressed publicly to a group of people becomes a *common issue*: choosing to answer the question in private, to the sender only, is not a polite act with respect to the audience. Note that the muddy children are "responsive" to the father's questions in precisely this way.

[23] One can formally define a notion of *strategy for agent a* in a dialogue game as a special kind of *set of communication sequences* constructed from agent a's available moves and from *knowledge tests for a* (which are simple announcements of the form $Pri_i\varphi(a,a)$): this is a "syntactic" version of the corresponding notion of strategy in Game Theory. One can then characterize "Socratic strategies" as strategies employing only rhetorical questioning moves, but by whose application the participants end up by increasing their knowledge (given certain assumptions about the game, e.g. responsiveness, truthfulness etc). We mention here that we have a *full formal treatment* of the reasoning and of the Socratic strategy in Muddy Children (and in the modified versions from our Examples), in the frame given by our modal logic of communication acts. One can pursue this line of finding formal characterizations of "good, normal types" of dialogue, trying for instance to get a hold on Grice's famous maxims (Grice 1989, see [Gr]). But we are more interested in abnormalities here.

what if the listeners do assume (as they usually do) the normality of the questioning act? This opens the possibility for the inquirer a to ask cheating questions, confusing the listeners from group B into thinking he doesn't know some things he knows: this will be a communication act α s.t. $\Box_a\varphi \in PRE_\alpha$ for some φ, but s.t. $(\neg\Box_a\varphi) \in PRE_\beta$ for all $\beta \in \alpha_b$, with $b \in B$. Another way to cheat is to go the other way around, answering questions with super-informativity, but only when and if the extra-details are prone to lead to wrong conclusions. [24] Yet another form of cheating that we can model in our setting is the act of *agent a successfully impersonating* agent c in the eyes of the group $B \cup \{b\}$: agent a sends to b a message $CON_\alpha = \epsilon_i\varphi(c, b, B)$, which bears c's signature; moreover, none of the agents in $B \cup \{b\}$ suspects the fraud: they think the imaginary message send by d is common knowledge between them and d (i.e. there exists an "apparent action" β s.t. $\alpha_c = \{\beta\}$ for all $c \in B \cup \{b\}$, $CON_\alpha = CON_\beta$, $\beta_c = \beta_d = \{\beta\}$ for all $c \in B \cup \{c\}$).[25] Preventing such actions to happen is an important issue when checking the correctness of *security protocols* for communication.

5 Comparison with Other Work.

One of the seminal ideas of our work comes from a paper of Gerbrandy and Groeneveld [GG]. The idea was 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 or semi-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 on epistemic actions started from observing some odd (or at least not always desirable) features of Gerbrandy's and Groeneveld's semi-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 φ is true". The default assumption is that these announcements are *completely private and completely secure* ("secret"), and hence they correspond in our notation[26] to communication acts of the form $Pri_i\varphi(a, b, A)$ (for any $a, b \in A$): the outsiders don't suspect anything. But there obviously are many other kinds of learning actions by semi-public announcements that one would like to model (e.g. our actions $Prss_i^k\varphi_1\varphi_2 \ldots \varphi_k a, b, B)$ above, in which the the secret message actually sent is φ_i, while the outsiders suspect that any one of the messages $varphi_1, \ldots, \varphi_k$ might have been sent).

The work of H. P. van Ditmarsch, although related in content, did not influence much our own work, as we have discovered it later, and enjoyed some comments and communications with him on these issues. His work deals with a special kind of communication moves, namely the ones occurring in the game of Cluedo. On the other hand, he also needs "fact-changing" actions (e.g. take a card), but this is not a problem for our approach (cf [B3]).

There is a large literature on the semantics and pragmatics of questions, starting with Grice [GR] and continuing with many others, among whom I want to mention the work of J. Groenendijk [GS1],[GS2],[GSV] whose approach to questions-as-partitions has been partially incorporated here. For the others, I list below a short list references.

Finally, I should mention the (by now, standard) alternative approach to modeling the flow of information in multi-agent systems, based on a *mixture of epistemic logic and temporal logic* (rather than dynamic logic). The origins of this approach are in the work of Fagin et al [FHMV], where the authors analyze knowledge in distributed systems. I was aware of this alternative from the very beginning of my work in this area: the fundamental issues, examples and insights that gave rise to our logics come from the work in [FHMV]. But I think that are important differences between the two kinds of settings, which make a direct comparison (of expressivity etc) very difficult; on one hand, their logics cannot express

[24] The are nice examples like this in Game Theory, in which answering truthful questions, but only on "special" occasions, can be more devastating for the "enemy" than random lying.

[25] This is the type of the impersonating action happening in Example 2 in the Introduction.

[26] See the **Examples** of "target logics" above.

specific types of actions: the kind of action that is going can be recovered only by looking at the specific model); on the other hand their models are "systems" , i.e. very rich structures (sets of "runs"), in which the whole future evolution is given, while ours are "open", in the sense that many possible actions can be executed on one of our epistemic states. In a sense, their logic is *too strong*: in general, it is not decidable. [27] Both technically and philosophically, our approach is essentially different: our models are simpler and easier to handle, as we are trying, not just to keep them finite, but to keep them as small as possible.

In conclusion, I would like to stress the *main original points* of the paper. *By comparison with other approaches*, the main conceptual and technical novelty consisted in our product-semantics for update, in which we have endowed communication acts (as we did in previous papers with "epistemic actions") with their own, internal epistemic structure. In addition to its philosophical importance, this idea has clear technical advantages: if offers a simple, compact way to represent epistemic changes and to compute their effect; it has greatly simplified our prior work on completeness and decidability for various logics, some proposed by J. Gerbrandy and H. van Ditmarsch, some arising from our own work; in its "syntactical" version, the idea of endowing actions with an epistemic "appearance" was useful in formulating simple, intuitive axioms to describe the interplay between knowledge (belief) and change. *If we compare to our own earlier work*, the main new points are: *the syntax in terms of communication signatures* (-this seems to me much more readable and usable than our previous attempts, in which the language was looking less like a language, but like a semantical object, a model); secondly, *an account of questions, interrogative actions, in addition to the one for assertive communication*; more generally, a notion of *communication acts*, which includes more information (sender/receiver, mode, time-stamp) than the notion of epistemic actions from our previous papers; an attempt to analyze some features of *dialogues*, *dialogue games*, *misleading questions* and *"Socratic questions"*; finally, an extension of our previous completeness result to the present setting. As a project for future work, I would like to look more in detail at the connections with the work already done on *using modal logics for checking security for communication protocols* (e.g. the so-called BAN logic) and to study such issues in the present setting (enriched maybe with cryptographic primitives).

References (short list):

[B1] A. Baltag *A Logic of Epistemic Actions.* In the *Proceedings of the Workshop on "Foundations and Applications of Collective Agent Based Systems"* 11th European Summer School on Logic, Language and Information (ESSLLI'99), Utrecht University, Utrecht 1999.

[B2] A. Baltag *The Dynamic Logic of Epistemic Actions.* Unpublished manuscript, 2000.

[B3]A. Baltag *A Logic for Suspicious Players.* Accepted for publication by *Bulletin of Economic Research*, to appear in 2001. A preliminary version was presented at the *Proceedings of LOFT 4*, ICER, Torino, Italy, June 2000.

[BMS] A. Baltag, L.S. Moss, S. Solecki, *The Logic of Public Announcements, Common Knowledge and Private Suspicions* (Extended Abstract). In the *Proceedings of TARK'98*, pp. 43-56. Morgan Kaufmann Publishers. 1998. An extended version appeared also as CWI Technical Report SEN-R9922, November 1999. Electronic version available at http://www.cs.indiana.edu/cogsci/techreps/238.html

[BMS2] A. Baltag, L.S. Moss, S. Solecki, *The Logic of Public Announcements, Common Knowledge and Private Suspicions.* Improved version. Submitted for publication to APAL in November 1999. Accepted, December 2000. To appear 2001.

[27] Only thirty-two of the ninety-six logics of knowledge and time analyzed by Halpern and Vardi [HV] contain common-knowledge operators; out of these, all but twelve are undecidable.

[**BE**] A. Bleeker, J. van Eijk, *Epistemic Action and Change.* Presented at LOFT4, 2000. In the Proceedings.

[**D**] H. P. van Ditmarsch, *Knowledge games*, Ph. D. Dissertation, Groningen University, 2000.

[**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.

[**Gr**] P. Grice, *Logic and Conversation.* In: Studies in the Ways of Words, pp. 22-40. Harvard University Press, Cambridge, MA, 1989.

[**GS1**] J. Groenendijk and M. Stokhof, *Studies on the Semantrics of Questions and the Pragmatics of Answers.*, Jurrians BV, Amsterdam, 1984.

[**GS2**] J. Groenendijk and M. Stokhof, *Questions.* In: J. van Bentham and A. ter Meulen, Handbook of Logic and Language,pp/ 1055-1124, Elsevier, 1997.

[**GSV**] J. Groenendijk, M. Stokhof, F. Veltmann, *Coreference and Modality.* In: Hanbook of Contemporary Semantic Theory, pp. 179-213, Blacwell, Oxford, 1996.(Editor: Shalom Lapin)

[**H**] C.L. Hamblin, *Mathematical Models of Dialogue.* Theoria, 2:130-155, 1971.

[**JB**] J. van Benthem, *Information Update as Relativisation.* Unpublished Manuscript, February 2000.

[**JB2**] J. van Benthem, *Games in Dynamic-Epistemic Logic.* Presented at LOFT4. To appear in the Proceedings.

[**JB3**] J. van Benthem, *'Hintikka Self-applied': an essay on the epistemic logic of imperfect information games.* To appear in Lewis Hahn, ed., Hintikka Volume, *Library of Living Philosophers.*

[**P**] J. Plaza, *Logics of public communications.* In the *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, 1989.

[**V**] F. Veltman, *Defaults in Update Semantics*, JPL, 25:221-261, 1996