"ONE IS A LONELY NUMBER": LOGIC AND COMMUNICATION

JOHAN VAN BENTHEM

Abstract. Logic is not just about single-agent notions like reasoning, or zero-agent notions like truth, but also about communication between two or more people. What we tell and ask each other can be just as logical as what we infer in Olympic solitude. We show how such interactive phenomena can be studied systematically by merging epistemic and dynamic logic, leading to new types of question.

§1. Logic in a social setting.

1.1. Questions and answers. Consider the simplest type of communication: a question-answer episode between two agents. Here is an example. I approach you in a busy Roman street, A.D. 180, intent on rescuing my former general Maximus, now held as a gladiator, and ask:

Q   Is this the road to the Colosseum?

As a well-informed and helpful Roman citizen, you answer truly:

A   Yes.

This is the sort of thing that we all do competently millions of times in our lives. There is nothing to it. But what is going on? I learn the fact that this is the road to the Colosseum. But much more happens. By asking the question, I convey to you that I do not know the answer, and also, that I think it possible that you do. This information flows before you have said anything at all. Then, by answering, you do not just convey the topographical fact to me. You also bring it about that you know that I know, I know that you know I know, etc. This knowledge up to every finite depth of mutual reflection is called *common knowledge*. It involves a mixture of factual information and iterated information about what others know.

These *epistemic overtones* concerning our mutual information are not mere side-effects. They may steer further concrete actions. Some bystanders' knowing that I know the road to Maximus' location may lead them to rush off and

warn the Emperor Commodus about this inquisitive Dutch visitor — while my knowing that they know I know may lead me to prevent them from doing just that. So epistemic overtones are ubiquitous and important, and we are good at computing them! In particular, we are well-attuned to fine differences in group knowledge. Everyone's knowing individually that your partner is unfaithful is unpleasant, but shame explodes when you meet people and know they all know that they know.

This is just the tip of an iceberg. I have described one type of question, but there are others. If you are my student, you would not assume that my classroom question shows that I do not know the answer. It need not even convey that I think you know, since my purpose may be to expose your ignorance. Such phenomena have been studied from lots of angles. Philosophers of language have developed speech act theory, linguists study the semantics of questions, computer scientists study communication mechanisms, and game theorists have their signaling games. All these perspectives are important — but there is also a foothold for *logic*. This paper aims to show that communication is a typical arena for logical analysis. Logical models help in raising and solving basic issues not recognized before.

1.2. The puzzle of the Muddy Children. Subtleties of information flow are often high-lighted in puzzles, some with a long history of appeal to broad audiences. A perennial example are the Muddy Children:

> *After playing outside, two of three children have mud on their fore-heads. They all see the others, but not themselves, so they do not know their own status. Now their Father comes and says: "At least one of you is dirty". He then asks: "Does anyone know if he is dirty?" The children answer truthfully. As this question-answer episode repeats, what will happen?*
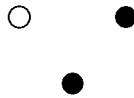
Nobody knows in the first round. But upon seeing this, the muddy children will both know in the second round, as each of them can argue as follows.

> *"If I were clean, the one dirty child I see would have seen only clean children around her, and so she would have known that she was dirty at once. But she did not. So I must be dirty, too!"*

This is symmetric for both muddy children — so both know in the second round. The third child knows it is clean one round later, after they have announced that. The puzzle is easily generalized to other numbers of clean and dirty children. Many variants are still emerging, as one can check by a simple Internet search.

Puzzles have a serious thrust, as they highlight subtle features of communication beyond simple questions and answers. E.g., consider the plausible putative *Learning Principle* stating that what we hear in public becomes common knowledge. This holds for announcing simple facts — such as the one in Tacitus that, long before international UN peace-keepers, German imperial guards already walked the streets of Rome. But the Principle is not valid in general! In the first round of Muddy Children, the muddy ones both announced the true fact that they did not know their status. But the result of that announcement was not that this ignorance became common knowledge. The announcement rather produced its own falsity, since the muddy children knew their status in the second round. Communicative acts involve *timing* and *information change*, and these may change truth values in complex ways. As we shall see, one of the virtues of logic is that it can help us keep all this straight.

1.3. Logical models of public communication. A logical description of our question-answer episode is easy to give. First, we need to picture the *relevant information* states, after that, we say how they are *updated*.

Answering a question. One initial information model for the group $\{Q, A\}$ of you and I has two states with "ϕ" ,"$\neg\phi$", with ϕ "this is the road to the Colosseum". We draw these states as points in a diagram. Also, we indicate agents' uncertainties between states. The labeled line shows that Q cannot distinguish between the two:

$$\phi \; \bullet \!\!\!\overset{Q}{\rule{3em}{0.4pt}}\!\!\! \circ \; \neg\phi$$
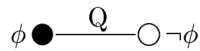
The black dot is an outside marker for the actual world where the agents live. There are no uncertainty lines for A. This reflects the fact that the Roman local A knows if this is the road to the Colosseum. But Q, though uninformed about the facts, sees that A knows in each eventuality, and hence he knows that A knows. This information about other's information is an excellent reason for asking a question.
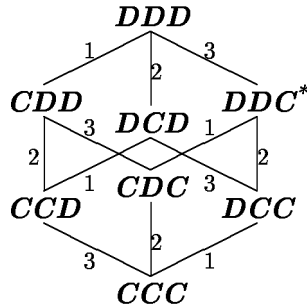
Next, A's answer triggers an *update* of this information model. In this simple case, A's answer eliminates the option *not-ϕ*, thereby changing the initial situation into the following one-point diagram:
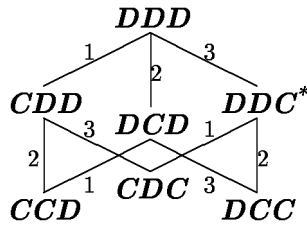
$$\phi \; \bullet$$

This picture has only one possible state of the world, where the proposition ϕ holds, and no uncertainty line for anyone. This indicates that ϕ is now common knowledge between you and me. Cognoscenti will recognize where we are heading. Information states are models for the modal logic S5 in its multi-agent version, and communication consists in actions which change such models. In what follows, we mean by "knowledge" only this much: "according to the agent's information".

Muddy Children: the movie. Here is a video of information updates for Muddy
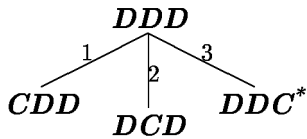
Children. States of the world assign D (dirty) or C (clean) to each child: 8 in total. In any of these, a child has one uncertainty. It knows about the others' faces, but cannot distinguish the state from one where its own D/C value is different:

$$
\begin{array}{c}
\textbf{DDD} \\
{}^{1}\diagup \;\; |_{2} \;\; {}^{3}\diagdown \\
\textbf{CDD} \quad \textbf{DCD} \quad \textbf{DDC}^{*} \\
{}^{3}\quad {}^{1} \\
2 \qquad\qquad 2 \\
{}^{1}\textbf{CDC}\,{}^{3} \\
\textbf{CCD} \qquad \textbf{DCC} \\
{}_{3}\;\; |_{2} \;\; {}_{1} \\
\textbf{CCC}
\end{array}
$$

Updates start with the Father's elimination of the world CCC:

$$
\begin{array}{c}
\textbf{DDD} \\
{}^{1}\diagup \;\; |_{2} \;\; {}^{3}\diagdown \\
\textbf{CDD} \quad \textbf{DCD} \quad \textbf{DDC}^{*} \\
{}^{3}\quad {}^{1} \\
2 \qquad\qquad 2 \\
{}^{1}\textbf{CDC}\,{}^{3} \\
\textbf{CCD} \qquad \textbf{DCC}
\end{array}
$$

When no one knows his status, the bottom worlds disappear:

$$
\begin{array}{c}
\textbf{DDD} \\
{}^{1}\diagup \;\; |_{2} \;\; {}^{3}\diagdown \\
\textbf{CDD} \qquad \qquad \textbf{DDC}^{*} \\
\textbf{DCD}
\end{array}
$$

The final update is to

$$\textbf{DDC}^{*}$$

1.4. General communication. Update by elimination of worlds incompatible with a statement made publicly is a simple mechanism. Human communication in general is very complex, including many other propositional attitudes than knowledge, such as belief or doubt — and logically more challenging phenomena than announcing the truth, such as hiding and cheating. There are two main lines of research here. One is further in-depth analysis of public communication, which turns out to be a quite subtle affair. This will be the main topic of the present paper. The other direction is modeling more complex communicative actions, such as giving answers to questions which others do not hear, or which others overhear, etc. Natural language has a rich

vocabulary for all kinds of attitudes toward information, speech acts, secrets, and so on — reflecting our natural proficiency with these. We will discuss more complex models briefly later in Section 9. Actually, this might seem a hopeless enterprise, as our behaviour is so diverse and open-ended. But fortunately, there exist simple realistic settings highlighting key aspects, viz. *games* that people engage in, which will also be discussed toward the end.

Some crucial references for this research program are [13, 39, 6, 14], and the extensive new version 2003 of the latter basic reference: [2]. Also well-worth reading is [40], which contains a mathematical analysis of all the subtle information passing moves in the well-known parlour game "Cluedo". The present paper builds on these references and others, while also including a number of unpublished results by the author over the past few years. Most results in what follows have been presented at public events since 1999, including the keynote lecture "Update Delights" at the 11th *ESSLLI* Summer School, Birmingham, 2000. An *ILLC* report version of this paper has been available on-line since 1992.

§2. **The basics of update logic.** The logic of public information update can largely be assembled from existing systems. We first survey basic epistemic logic and dynamic logic, and then discuss their combination.

2.1. Epistemic logic. *Language.* Epistemic logic has the following explicit notation for talking about knowledge of agents:

$$K_j\phi \qquad \text{agent } j \text{ knows that } \phi.$$

With such a symbolism, we can also analyse further patterns:

$$\neg K_j\neg\phi \text{ (or } \diamond_j\phi) \qquad \text{agent } j \text{ considers it } possible that \phi$$
$$K_j\phi \vee K_j\neg\phi \qquad \text{agent } j \text{ knows } if \phi$$
$$K_j\neg K_i\phi \qquad \text{agent } j \text{ knows that agent } i \text{ does not know that } \phi$$

E.g., in asking a "normal" question, Q conveys that he does not know if ϕ:

$$\neg K_Q\phi \wedge \neg K_Q\neg\phi$$

and also that he thinks that A might know:

$$\diamond_Q(K_A\phi \vee K_A\neg\phi)$$

By answering affirmatively, A conveys that she knows that ϕ, but she also makes Q know that ϕ, that Q knows that A knows, and so on up to any finite iteration depth, leading to *common knowledge*, written as follows:

$$C_{\{Q,A\}}\phi$$

Models. Models for this epistemic language are of the form

$$M = (S, \{\sim_j \mid j \in G\}, V)$$

with (a) S a set of worlds, (b) V a valuation function for proposition letters, and (c) for each agent $j \in G$, an *equivalence relation* $\sim_j$ relating worlds s to all worlds that j cannot distinguish from it. These models M may be viewed as collective information states.

Semantics. Next, in these models, an agent j *knows* exactly those propositions that are true in all worlds she cannot distinguish from the current one. That is:

$$M, s \models K_j \phi \quad \text{iff} \quad M, t \models \phi \quad \text{for all } t \text{ s.t. } s \sim_j t$$

The related notation $\neg K_j \neg \phi$ or $\Diamond_j \phi$ works out to:

$$M, s \models \Diamond_j \phi \quad \text{iff} \quad M, t \models \phi \quad \text{for some } t \text{ s.t. } s \sim_j t$$

In addition, there are several useful operators of "group knowledge":

> *Universal knowledge* $E_G \phi$
> This is just the conjunction of all formulas $K_j \phi$ for $j \in G$.

> *Common knowledge* $C_G \phi$
> This says at s that ϕ is true in every state reachable from s by finitely many $\sim$-steps for members of group G (e.g.,$\sim_1, \sim_2, \sim_1, \sim_3$).

> *Implicit knowledge* $I_G \phi$
> This says that ϕ is true in all states which are related to s via the intersection of all uncertainty relations $\sim_j$ for $j \in G$.

Logic. Information models validate an epistemic logic that can describe and automate reasoning with knowledge and ignorance of interacting agents. Here are its major validities:

$$
\begin{array}{ll}
K_j(\phi \rightarrow \psi) \rightarrow (K_j \phi \rightarrow K_j \psi) & \textit{Knowledge Distribution} \\
K_j \phi \rightarrow \phi & \textit{Veridicality} \\
K_j \phi \rightarrow K_j K_j \phi & \textit{Positive Introspection} \\
\neg K_j \phi \rightarrow K_j \neg K_j \phi & \textit{Negative Introspection}
\end{array}
$$

The complete system is multi-S5, which serves in two different guises: describing the agents' own explicit reasoning, and describing our reasoning as theorists about them. And here are the required additional axioms for a complete logic with common knowledge:

$$
\begin{array}{ll}
C_G \phi \leftrightarrow \phi \wedge E_G C_G \phi & \textit{Equilibrium Axiom} \\
(\phi \wedge C_G(\phi \rightarrow E_G \phi)) \rightarrow C_G \phi & \textit{Induction Axiom}
\end{array}
$$

This is the standard decidable version of epistemic logic.

2.2. Dynamic logic. The usual logic of knowledge by itself can only describe static snapshots of a communication sequence. To describe the updates themselves, we must add actions explicitly.

Language. The language now has formulas F and program expressions P on a par. We write this two-level syntax in abbreviated form with so-called Backus-Naur notation:

$$F := \text{propositional atoms } p, q, r, \ldots \mid \neg F \mid (F \wedge F) \mid \Diamond_P F$$
$$P := \text{basic actions } a, b, c, \ldots \mid (P; P) \mid (P \cup P) \mid P^* \mid (F)?$$

Semantics. This formalism is interpreted over polymodal models

$$M = \langle S, \{ R_a \mid a \in A \}, V \rangle$$

which are intuitively seen as process graphs with states and possible basic transitions. The truth definition explains two notions in one recursion:

$$M, s \quad \models \phi \qquad \phi \text{ is true at state } s$$
$$M, s_1, s_2 \models \pi \qquad \text{the transition from } s_1 \text{ to } s_2 \text{ corresponds to}$$
$$\text{a successful execution for the program } \pi$$

Here are the inductive clauses:

▶ $M, s \quad \models p$ iff $s \in V(p)$
$M, s \quad \models \neg\psi$ iff $\text{not } M, s \models \psi$
$M, s \quad \models \phi_1 \wedge \phi_2$ iff $M, s \models \phi_1 \text{ and } M, s \models \phi_2$
$M, s \quad \models \Diamond_\pi\phi$ iff $\text{for some } s' \text{ with } M, s, s' \models \pi: M, s' \models \phi$

▶ $M, s_1, s_2 \models a$ iff $(s_1, s_2) \in R_a$
$M, s_1, s_2 \models \pi_1; \pi_2$ iff $\text{there exists a } s_3 \text{ with}$
$M, s_1, s_3 \models \pi_1 \text{ and } M, s_3, s_2 \models \pi_2$
$M, s_1, s_2 \models \pi_1 \cup \pi_2$ iff $M, s_1, s_2 \models \pi_1 \text{ or } M, s_1, s_2 \models \pi_2$
$M, s_1, s_2 \models \pi^*$ iff $\text{some finite sequence of } \pi\text{-transitions}$
$\text{in } M \text{ connects } s_1 \text{ with } s_2$
$M, s_1, s_2 \models (\phi)?$ iff $s_1 = s_2 \text{ and } M, s_1 \models \phi$

Thus, formulas have the usual Boolean operators, while the existential modality $\Diamond_\pi\phi$ is a weakest precondition true at only those states where program π can be performed to achieve the truth of ϕ. The program constructions are the usual regular operations of relational composition, Boolean choice, Kleene iteration, and tests for formulas. This system defines standard control operators on programs such as:

IF ε THEN π_1 ELSE π_2 $((\varepsilon)? ; \pi_1) \cup ((\neg\varepsilon)? ; \pi_2)$
WHILE ε DO π $((\varepsilon)? ; \pi)^* ; (\neg\varepsilon)?$

Logic. Dynamic logic expresses all of modal logic plus regular relational set algebra. Its complete set of validities is known (cf. [9]):

- ▶ All principles of the minimal modal logic for all modalities $\Box_\pi$
- ▶ Computation rules for weakest preconditions:

$$\Diamond_{\pi_1;\pi_2}\phi \;\leftrightarrow\; \Diamond_{\pi_1}\Diamond_{\pi_2}\phi$$
$$\Diamond_{\pi_1\cup\pi_2}\phi \leftrightarrow \Diamond_{\pi_1}\phi \vee \Diamond_{\pi_2}\phi$$
$$\Diamond_{\phi?}\psi \quad\leftrightarrow \phi \wedge \psi$$
$$\Diamond_{\pi^*}\phi \quad\leftrightarrow \phi \vee \Diamond_\pi\Diamond_{\pi^*}\phi$$

- ▶ Induction Axiom: $(\phi \wedge \Box_{\pi^*}(\phi \rightarrow \Box_\pi\phi)) \rightarrow \Box_{\pi^*}\phi$

Like basic epistemic logic, the logic of public announcement is decidable. This remains so with certain *extensions* of the basic language, such as the program construction ∩ of *intersection* — to which we will return below. Extended modal languages occur quite frequently in applications.

2.3. Dynamic epistemic logic. Analyzing communication requires a logic of knowledge in action, combining epistemic logic and dynamic logic. This may be done in at least two ways.

Abstract DEL. One can join the languages of epistemic and dynamic logic, and merge the signatures of their models. This yields abstract logics of knowledge and action: cf. [12] on planning, or [28] on imperfect information games. The general logic is the union of epistemic multi-S5 and dynamic logic. This is a good base for experimenting with further constraints. An example is agents having perfect memory for what went on in the course of communication (cf. [7]). This amounts to an extra commutation axiom

$$K_j\Box_a\phi \longrightarrow \Box_a K_j\phi$$

Abstract *DEL* may be the best setting for general studies of communication.

Concrete update logic. More concretely, in Section 1, public announcement of a proposition ϕ changed the current epistemic model M, s, with actual world s to a new one as follows:

eliminate all worlds which currently do not satisfy ϕ.
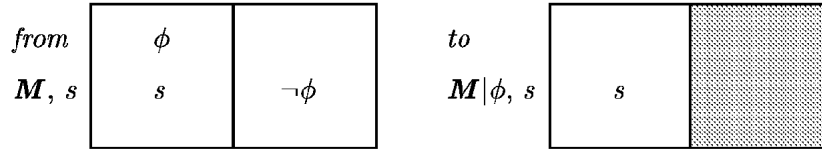
Thus, we work in a universe whose states are epistemic models — either all of them or just some family — and basic actions are public announcements $A!$ of assertions A from the epistemic language. These actions are *partial functions*. If A is true, then it can be truthfully announced with a unique update. From the standpoint of dynamic logic, this is just one concrete instance of abstract process models, plus epistemic extras. The appropriate logic has combined

dynamic-epistemic assertions

$$\Box_{A!} \, \phi \;\text{``after truthful announcement of A, ϕ holds.''}$$

The set-up of this system merges epistemic with dynamic logic, with some additions reflecting particulars of our update universe. There is a complete and decidable axiomatization ([17, 6]), with the following key axioms:

$$
\begin{aligned}
\diamondsuit_{A!} \, p &\quad\leftrightarrow\quad A \wedge p \text{ for atomic facts } p\\
\diamondsuit_{A!} \, \neg\phi &\quad\leftrightarrow\quad A \wedge \neg\diamondsuit_{A!} \, \phi\\
\diamondsuit_{A!} \, \phi \vee \psi &\quad\leftrightarrow\quad \diamondsuit_{A!} \, \phi \vee \diamondsuit_{A!} \, \psi\\
\diamondsuit_{A!} \, K_i\phi &\quad\leftrightarrow\quad A \wedge K_i(A \to \diamondsuit_{A!} \, \phi)
\end{aligned}
$$

Essentially, these compute preconditions $\diamondsuit_{A!} \, \phi$ by *relativizing* the postcondition ϕ to A. The axioms can also be stated with the modal box, leading to versions like

$$\Box_{A!}K_i\phi \longleftrightarrow A \longrightarrow K_i\Box_{A!}\phi$$

This is like the above law for Perfect Recall. As for common knowledge, the earlier epistemic language needs a little extension, with a *binary* version

$$C_G(A, \phi) \qquad \text{common knowledge of } \phi \text{ in the submodel defined by } A.$$

There is no definition for this in terms of just absolute common knowledge. Having added this feature, we can state the remaining reduction principle

$$\diamondsuit_{A!}C_G\phi \longleftrightarrow A \wedge C_G\big(A, \diamondsuit_{A!}\phi\big)$$

[37] develops this idea in detail. There are even further ways of combining knowledge and action. In particular, [24] defines a more thoroughly epistemized dynamic logic with transitions as objects in their own right.

DEL with program constructions. Public announcement is just one basic action. Conversation may involve more complex programming of what is said. Saying one thing after another amounts to program composition, choosing one's assertions involves choice, and Muddy Children even involved a guarded iteration:

$$\text{WHILE ``you don't know your status'' DO ``say so''.}$$

The basic logic of public update with the first two constructions is like its version with just basic announcements $A!$, because of the reduction axioms for composition and choice in dynamic logic. But with possible iteration of announcements, the system changes — and even loses its decidability ([2]).

§3. **Basic theory of information models.** *Special model classes.* Multi-S5 models can be quite complicated. But there are some subclasses of special interest. For instance, Muddy Children started with a full cube of 3-vectors,

with accessibility given as the special equivalence relation

$X \sim_j Y$ which holds if the j-th coordinates are equal: $(X)_j = (Y)_j$

Cube models are studied in algebraic logic ([10]) for their connections with assignment spaces over first-order models. But the subsequent Muddy Children updates led to submodels of such cubes. These already attain full epistemic generality ([20]):

THEOREM 3.1. *Every multi-*S5 *model is representable up to bisimulation as a submodel of a full cube.*

Other special model classes arise in the study of card games ([40]) and games in general ([29]).

Bisimulation. Epistemic and dynamic logic are both standard modal logics (cf. [4]) with the following structural model comparison:

DEFINITION 3.2. A *bisimulation* between two models M, N is a binary relation $\equiv$ between their states m, n such that, whenever $m \equiv n$, then (a) m, n satisfy the same proposition letters, (b1) if mRm', then there exists a world n' with nRn' and $m' \equiv n'$, and (b2) the analogous "zigzag clause" holds in the opposite direction.

Our question-answer example has a bisimulation with this variant:
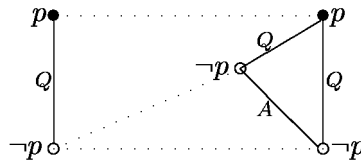
In a natural sense, these are two representations of the same information state. Bisimulation equivalence occurs naturally in update. Suppose the current model is like this, with the actual world indicated by the black dot:
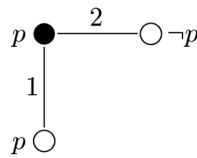
Note that all three worlds satisfy different epistemic formulas. Now, despite her uncertainty, *1* knows that p, and can say this — updating to the model
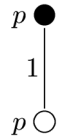
But this can be contracted via a bisimulation to the one-point model

$$p \; \bullet$$

In general, it is convenient to think of update steps with bisimulation contractions interleaved automatically.

Some basic results link bisimulation to truth of modal formulas. For convenience, we restrict attention to *finite models* — but this can be lifted.
Invariance and definability. Consider general models, or those of multi-S5.

LEMMA 3.3 (Invariance Lemma). *The following are equivalent*:

(a) M, s *and* N, t *are connected by a bisimulation*,
(b) M, s *and* N, t *satisfy the same modal formulas*.

Any model has a bisimilar *unraveled tree model*, but also a smallest *bisimulation contraction* satisfying the same modal formulas. Another useful tool are modal analogues of Scott-sentences for infinitary logic (cf. [3]):

LEMMA 3.4 (State Definition Lemma). *For each model* M, s *there is an epistemic formula* β (*involving common knowledge*) *such that the following assertions are equivalent*:

(a) $N, t \models \beta$,
(b) N, t *has a bisimulation* $\equiv$ *with* M, s *such that* $s \equiv t$.

PROOF. The version and proof given here are from [21, 23]. Consider any finite multi-S5 model M, s. This falls into a number of maximal "zones" of worlds that satisfy the same epistemic formulas in our language.

CLAIM 3.5. There exists a finite set of formulas ϕ_i ($1 \leq i \leq k$) such that (a) each world satisfies one of them, (b) no world satisfies two of them (i.e., they define a partition of the model), and (c) if two worlds satisfy the same formula ϕ_i, then they agree on all epistemic formulas.

To show this, take any world s, and find difference formulas $\delta^{s,t}$ between it and any t which does not satisfy the same epistemic formulas, where s satisfies $\delta^{s,t}$ while t does not. The conjunction of all $\delta^{s,t}$ is a formula ϕ_i true only in s and the worlds sharing its epistemic theory. We may assume the ϕ_i also list all information about the proposition letters true and false throughout their partition zone. We also make a quick observation about uncertainty links between these zones:

\#   If any world satisfying ϕ_i is $\sim_a$-linked to a world satisfying ϕ_j, then all worlds satisfying ϕ_i also satisfy $\diamond_a \phi_j$.

Next take the following description $\beta_{M,s}$ of M, s:

(a)   all (negated) proposition letters true at s
        plus the unique ϕ_i true at M, s

  (b)   common knowledge for the whole group of
       (b1)   the disjunction of all ϕ_i,
       (b2)   all negations of conjunctions $\phi_i \wedge \phi_j$ $(i \neq j)$,
       (b3)   all implications $\phi_i \rightarrow \Diamond_a \phi_j$ for which situation # occurs,
       (b4)   all implications $\phi_i \rightarrow \Box_a \vee \phi_j$ where the disjunction runs
           over all situations listed in the previous clause.

CLAIM 3.6. $M, s \models \beta_{M,s}$

CLAIM 3.7. If $N, t \models \beta_{M,s}$, then there exists a bisimulation between N, t and M, s.

To prove Claim 3.7, let N, t be any model for $\beta_{M,s}$. The ϕ_i partition N into disjoint zones Z_i of worlds satisfying these formulas. Now relate all worlds in such a zone to all worlds that satisfy ϕ_i in the model M. In particular, t gets connected to s. We check that this connection is a bisimulation. The atomic clause is clear from an earlier remark. But also, the zigzag clauses follow from the given description. (a) Any $\sim_a$-successor step in M has been encoded in a formula $\phi_i \rightarrow \Diamond_a \phi_j$ which holds everywhere in N, producing the required successor there. (b) Conversely, if there is no $\sim_a$-successor in M, this shows up in the limitative formula $\phi_i \rightarrow \Box_a \vee \phi_j$, which also holds in N, so that there is no "excess" successor there either. This concludes the proof.   ⊣

The Invariance Lemma says that bisimulation has the right fit with the modal language. The State Definition Lemma says that each semantic state can be characterized by one epistemic formula. E.g., consider the two-world model for our question-answer episode. Here is an epistemic formula which defines its ϕ-state up to bisimulation:

$$\phi \wedge C_{\{Q,A\}} \left(\left(K_A \phi \vee K_A \neg \phi \right) \wedge \neg K_Q \phi \wedge \neg K_Q \neg \phi \right)$$

This allows us to switch, in principle, between semantic accounts of information states as models M, s and syntactic ones in terms of complete defining formulas. There is more to this than just technicality. For instance, syntactic approaches have been dominant in related areas like belief revision theory, where information states are not models but syntactic theories. It would be good to systematically relate syntactic and semantic approaches to update, but we shall stay semantic here.

Respectful and safe operations. The above also constrains epistemic update operations O. These should *respect bisimulation*:

  If M, s and N, t are bisimilar, so are their values $O(M, s)$ and $O(N, t)$

FACT 3.8. Public update respects bisimulation.

PROOF. Let $\equiv$ be a bisimulation between M, s and N, t. Consider their submodels $M|\phi, s$, $N|\phi, t$ after public update with ϕ. The restriction of $\equiv$ to these is still a bisimulation. Here is the zigzag clause. Suppose some world w has an $\sim_i$-successor v in $M|\phi, s$. This same v is still available in the other

model: it remained in M since it satisfied ϕ. But then v also satisfied ϕ in N, t, because of the Invariance Lemma for the bisimulation $\equiv$ — and so it stayed in the updated model $N|\phi, t$, too. ⊣

Many other proposed update operations respect bisimulations (cf. [8] on similar phenomena in process algebra). Finally, bisimulation also works for the full language of dynamic logic — but with a new twist ([20]). Intertwined with invariance for formulas ϕ, one must show that the zigzag clauses go through for all regular program constructions: not just the atomic R_a, but each transition relation $[[\pi]]$:

FACT 3.9. Let $\equiv$ be a bisimulation between two models M, M', with $s \equiv s'$. Then it holds that

 (i) s, s' verify the same formulas of propositional dynamic logic,
 (ii) if $s[[\pi]]^M t$, then there exists t' with $s'[[\pi]]^{M'} t'$ and $s' \equiv t'$.

This observation motivates a new invariance for program operations:

DEFINITION 3.10. An operation $O(R_1, \ldots, R_n)$ on programs is *safe for bisimulation* if, whenever $\equiv$ is a relation of bisimulation between two models for their transition relations $R_1, \ldots, R_n$, then it is also a bisimulation for the defined relation $O(R_1, \ldots, R_n)$.

The core of the above program induction is that the three regular operations; $\cup$ * of PDL are safe for bisimulation. By contrast, the operation of program *intersection* is not safe:
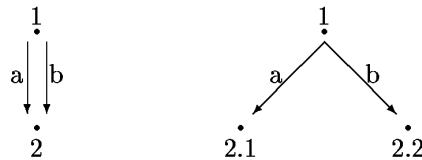
There is an obvious bisimulation here with respect to a, b — but the required zigzag clause fails for $R_a \cap R_b$.

After this technical extravaganza, we return to communication. In fact, the Muddy Children puzzle highlights a whole agenda of further questions. We already noted how its specific model sequence is characteristic for the field. But in addition, it raises many further central issues, such as

  (a) the benefits of internal group communication,
  (b) the role of iterated assertion,
  (c) the interplay of update and inference in reasoning.

We will look into these as we go. But we start with an issue already noted: the putative "learning principle" that was refuted by Muddy Children.

§**4. What do we learn from a statement?** *Specifying speech acts.* Update logic is a sequel to speech act theories, which originated in philosophy, and then partly migrated to computer science (cf. [43]). Earlier accounts of speech acts often consist in formal specifications of preconditions and postconditions of successful assertions, questions, or commands. Some of these insights are quite valuable, such as those concerning assertoric force of assertions. E.g., in what follows, we will assume, in line with that tradition, that normal cooperative speakers may only utter statements which they know to be true. Even so, what guarantees that the specifications are correct? E.g., it has been said that answers to questions typically produce common knowledge of the answer. But Muddy Children provided a counter-example to this naive "Learning Principle". Logical tools help us get clearer on pitfalls and solutions. The learning problem is a good example of this use.

Persistence under update. Public announcement of atomic facts p makes them common knowledge, and the same holds for other types of assertion. But, as we noted in Section 1, not all updates with ϕ result in common knowledge of ϕ! A simple counter-example is this. In our question-answer case, let A say truly

$$p \wedge \neg K_Q p \qquad \text{``}p, \text{ but you don't know it.''}$$

This utterance removes Q's lack of knowledge about the fact p, and thereby makes its own assertion false! Ordinary terminology is misleading here:

> *learning that* ϕ is ambiguous between: ϕ *was* the case, before the announcement, and ϕ *is* the case — after the announcement.

The explanation is that statements may change their truth value with update. For worlds surviving in the smaller model, factual properties do not change, but epistemic properties may. This raises a general logical issue of *persistence under update*:

> Which forms of epistemic assertion remain true at a world
> whenever other worlds are eliminated from the model?

These are epistemic assertions which, when publicly announced to a group, will always result in common knowledge. Examples are atomic facts p, and knowledge-free assertions generally, but also knowledge assertions $K_i p$ and ignorance assertions $\neg K_i p$.

New kinds of preservation result. Here is a relevant model-theoretic fact from modal logic (cf. [36]):

THEOREM 4.1. *The epistemic formulas* without common knowledge *that are preserved under submodels are precisely those definable using literals* p, $\neg p$, *conjunction, disjunction, and* K_i-*operators.*

Compare universal formulas in first-order logic, which are just those preserved under submodels. The obvious conjecture for the epistemic language

with common knowledge would just add arbitrary formulas $C\phi$ as persistent forms. But this result is still open, as lifting first-order model theory to such non-first-order modal fixed-point languages seems non-trivial, even on a universe of finite models.

QUESTION. Which formulas of the full epistemic language with common knowledge are preserved under submodels?

In any case, what we need is not really full preservation under submodels, but rather preservation under "self-defined submodels":

> When we *restrict* a model to those of its worlds which satisfy ϕ, then ϕ should hold throughout the remaining model, or in terms of an elegant validity: $\phi \rightarrow (\phi)^\phi$

QUESTION. Which epistemic formulas imply their self-relativization?

Indeed, which *first-order* formulas are preserved in this self-fulfilling sense? Model-theoretic preservation questions of this special form seem new.
A non-issue? Many people find this particular issue annoying. Non-persistence seems a side-effect of infelicitous wording. E.g., when A said "p, but you don't know it", she should just have said "p", keeping her mouth shut about my mental state. Now, the Muddy Children brand of non-persistence is not as blatant as this. And in any case, dangers in timing aspects of what was true before and is true after an update are no more exotic than the acknowledged danger in computer science of confusing states of a process. Dynamic logics were developed precisely to keep track of that. But let's stop fencing: *can* we reword any message to make the non-persistence go away? An epistemic assertion A defines a set of worlds in the current model M. Can we always find an equivalent persistent definition? This would be easy if each world has a simple unique factual description, as is the case with hands in card games. But even without assuming this there is a method that works, at least locally:

FACT 4.2. In each model, every public announcement has a persistent equivalent leading to the same update.

PROOF. Without loss of generality, assume we are working only with bisimulation-contracted models which are also totally connected: no isolated components. Let w be the current world in model M. Let j publicly announce A, updating to the submodel $M|A$ with domain $A^* = \{s \in M | M, s \models A\}$. If this is still M itself, then the announcement "True" is adequate, and persistent. Now suppose A^* is not the whole domain. Our persistent assertion consist of two disjuncts:

$$\Delta \vee \Sigma$$

First we make Δ. Using the proof of the State Definition Lemma of Section 3, this is an epistemic definition for A^* in M formed by describing each world in it up to bisimulation, and then taking the disjunction of these.

Now for Σ. Again using the mentioned proof, write a formula which describes $M|A$ up to bisimulation. For concreteness, this had a common knowledge operator over a plain epistemic formula describing the pattern of states and links, true everywhere in the model $M|A$. No specific world description is appended, however.

Clearly $\Delta \vee \Sigma$ is common knowledge in $M|A$, because Σ is. But it also picks out the right worlds in M. First, any world in A^* satisfies its own disjunct of Δ. Conversely, suppose any world t in M satisfies $\Delta \vee \Sigma$. If it satisfies some disjunct of Δ, then it must then be in A^* by the bisimulation-minimality of the model. Otherwise, M, t satisfies Σ. But then by connectedness, every world in M satisfies Σ, and in particular, given the construction of Σ, there must be a bisimulation between M and $M|A$. But this contradicts the fact that the update was genuine. $\dashv$

Of course, this recipe for phrasing your assertions is ugly, and not recommended! Moreover, it is local to one model, and does not work uniformly. Recall that, depending on group size, muddy children may have to repeat the same ignorance statement any number of times before knowledge dawns. If there were one uniform persistent equivalent for that statement, the latter's announcement would always lead to common knowledge after some fixed finite stage.

§**5. Internal communication in groups.** *The best we can.* The muddy children might just tell each other what they see, and common knowledge of their situation is reached at once. The same holds for card players telling each other their hands. Of course, life is civilized precisely because we do not "tell it like it is". Even so, there is an issue of principle of what agents in a group can achieve by maximal communication. Consider two epistemic agents that find themselves in some collective information state M, at some actual situation s. They can tell each other things they know, thereby cutting down the model to smaller sizes. Suppose they wish to be maximally cooperative:

> What is the best correct information they can give each other via successive updates — and what does the resulting collective information state look like?

E.g., what is the best that can be achieved in the following model?
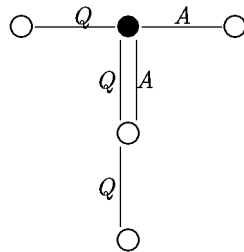
Geometrical intuition suggests that this must be:
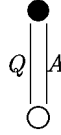
This is correct! First, any sequence of mutual updates in a finite model must terminate in some minimal domain which can no longer be reduced. This is reached when everything each agent knows is already common knowledge: i.e., it holds in every world. But what is more, this minimal model is *unique*, and we may call it the "communicative core" of the initial model. Here is an explicit description, proved in [27]:

THEOREM 5.1. *Each model has a communicative core, viz. the set of worlds that are reachable from the actual world via all uncertainty links.*

PROOF. For convenience, consider a model with two agents only. The case with more than two agents is an easy generalization of the same technique. First, agents can reach this special set of worlds as follows. Without loss of generality, let all states t in the model satisfy a unique defining formula δ_t as in Section 3 — or obtained by an ad-hoc argument. Agent *1* now communicates all he knows by stating the *disjunction* $\vee \delta_t$ for all worlds t he considers indistinguishable from the actual one. This initial move cuts the model down to the actual world plus all its $\sim_1$-alternatives. Now there is a small technicality. The resulting model need no longer satisfy the above unique definability property. The update may have removed worlds that distinguished between otherwise similar options. But this is easy to remedy by taking the *bisimulation contraction*. Next, let *2* make a similar strongest assertion available to her. This cuts the model down to those worlds that are also $\sim_2$-accessible from the actual one. After that, everything any agent knows is common knowledge, so further statements have no informative effect.

Next, suppose agents reach a state where further announcements have no effect. Then the following implications hold for all ϕ : $K_1\phi \rightarrow C_{\{1,2\}}\phi$, $K_2\phi \rightarrow C_{\{1,2\}}\phi$. Again using defining formulas, this means *1*, *2* have the same alternative worlds. So, these form a *subset* of the above core. But in fact, all of it is preserved. An agent can only make statements that hold in all of its worlds, as it is included in his information set. Therefore, the whole core survives each episode of public update, and by induction, it survives all of them.   ⊣

A corollary of the preceding proof is this:

FACT 5.2. Agents need only 2 communication rounds to get to the core.

In particular, there is no need for repetitions by agents. E.g., let *1* truly say *A* (something he knows in the actual world), note the induced public update, and then say *B* (which he knows in the new state). Then he might just as well

have asserted $A \wedge (B)^A$ straightaway: where $(B)^A$ is the relativization of B to A (cf. Section 6).

Incidentally, a two-step solution to the initial example of this section is the following rather existentialist conversation:

> Q sighs: "I don't know."
> A sighs: "I don't know either."

It does not matter if you forget the details here, because it also works in the opposite order.

The communicative core is the actual world plus every world connected to it by the intersection of all uncertainty relations. This is the range used in defining *implicit knowledge* for a group of agents in Section 2.1 Thus, maximal communication turns implicit knowledge of a group into common knowledge. As a slogan, this makes sense, but there are subtleties. It may be implicit knowledge that none of us know where the treasure is. But once the communicative core is all that is left, the location of the treasure may be common knowledge. Compare the difference between quantifier restriction and relativization. Implicit knowledge $I_G\phi$ looks only at worlds in the communicative core CC, but it then evaluates the formula ϕ from each world there in the whole model. By contrast, internal evaluation in the core is like evaluating totally relativized statements $(\phi)^{CC}$ in the model.

Another technicality is that the relevant *intersection* of relations, though keeping the logic decidable, is no longer safe for bisimulation in the sense of Section 4. Adding it to the language leads to a genuinely richer epistemic logic, for which some of the earlier model theory would have to be redone.

Planning assertions.   The preceding topic really shows a shift in interest. Update logics can be used to analyze given assertions, but they can also be used to plan assertions meeting certain specifications. An example is the following puzzle from a mathematical Olympiad in Moscow (cf. [41]):

> 7 cards are distributed among A, B, C. A gets 3, B 3, C 1. How should A, B communicate publicly, in hearing of 3, so that they find out the precise distribution of the cards while C does not?

There are solutions here — but their existence depends on the number of cards. This question may be seen as a generalization of the preceding one. How can a subgroup of all agents communicate maximally, while keeping the rest of the group as much in the dark as possible? Normally, this calls for hiding, but it is interesting to see — at least to politicians, or illicit lovers — that some of this can be achieved publicly. This sort of planning problem is only beginning to be studied.

Here we just observe an analogy with computer science. The dynamic epistemic logic of Section 2.3 is like well-known program logics manipulating

correctness assertions

$$\phi \to \Box_{A!}\, \psi \quad \text{if precondition } \phi \text{ holds, then saying } A \text{ always}$$
leads to a state where postcondition ψ holds.

Such triples may be looked at in different ways. Given an assertion, one can analyze its preconditions and postconditions, as we did for questions and answers. Or, given an assertion $A!$ plus a precondition ϕ we can look for their strongest postcondition ψ. E.g., with $\phi = \textit{True}$, and A atomic, it is easy to show that the strongest postcondition is the common knowledge of A. This is program analysis. But there is also program *synthesis*. Given a precondition ϕ and a postcondition ψ, we can look for an assertion $A!$ guaranteeing the transition. Here the postcondition ψ may specify which agents are supposed to learn what, using assertions of the forms $K_j \alpha$ and $\neg K_j \beta$. Conversation planning is like this, determining what to say so that only the right people get the right information.

§**6. Public update as relativization.** The following technical intermezzo ([25]) joins forces with standard logic.
Semantic and syntactic relativization. Here is a simple observation. Announcing A amounts to a logical operation of *semantic relativization*

from a model M, s to the definable submodel $M|A, s$.

This explains all behaviour so far — while raising new questions. For a start, in the new model, we can again evaluate formulas that express knowledge and ignorance of agents, in the standard format $M|A, s \models \phi$. In standard logic, this may also be described via *syntactic relativization* of the formula ϕ by the update assertion A:

LEMMA 6.1 (Relativization Lemma). $M|A, s \models \phi$ *iff* $M, s \models (\phi)^A$

This says we can either evaluate assertions in a relativized model, or equivalently, their relativized versions in the original model. For convenience, we assume that relativization is defined so that $(\phi)^A$ always implies A. For the basic epistemic language, this goes via the following recursion:

$$(p)^A = A \wedge p$$
$$(\neg\phi)^A = A \wedge \neg(\phi)^A$$
$$(\phi \vee \psi)^A = (\phi)^A \vee (\psi)^A$$
$$(K_i\phi)^A = A \wedge K_i\big(A \longrightarrow (\phi)^A\big)$$

In this definition, one immediately recognizes the above axioms for public update. Whether this works entirely within the language of epistemic announcements depends on its strength. E.g., relativization was less straightforward with common knowledge, as no syntactic prefix "$A \to \dots$" or "$A \wedge \dots$" on

absolute operators C_G does the job. But we saw how to extend epistemic logic with a binary restricted common knowledge operator. Actually, dynamic logic is better behaved in this respect.

FACT 6.2. Dynamic logic is closed under relativization.

PROOF. In line with the usual syntax of the system, we need a double recursion over formulas and programs. For formulas, the clauses are all as above, while we add

$$(\Box_\pi \phi)^A = \Box_{(\pi)^A}(\phi)^A$$

For programs, here are the recursive clauses that do the job:

$$(R; S)^A = (R)^A; (S)^A$$
$$(R \cup S)^A = (R)^A \cup (S)^A$$
$$((\phi)?)^A = (A)?; (\phi)^A?$$
$$(\pi^*)^A = ((A)?; (\pi)^A)^* \qquad \dashv$$

Now, common knowledge $C_G\phi$ may be viewed as a dynamic logic formula

$$\Box_{(\cup\{i|i\in G\})^*}\phi$$

Therefore, we can get a natural relativization for epistemic logic by the above Fact, by borrowing a little syntax from dynamic logic.

General logic of relativization. Stripped of its motivation, update logic is an axiomatization of one model-theoretic operation, viz. relativization. There is nothing specifically modal about this. One could ask for a complete logic of relativizations $(\phi)^A$ in first-order logic, as done for *substitutions* $\Box_{t/x}\phi$ in [10].

QUESTION. What is the complete logic of the relativization operator in a standard first-order language?

At least we may observe that there is more than the axioms listed in Section 2.3. For instance, the following additional fact is easy to prove:

Associativity $((A)^B)^C$ is logically equivalent to $A^{((B)^C)}$.

In our update logic, performing two relativizations matches performing two consecutive updates. Thus Associativity amounts to the validity of

$$\Box_{A!;B!}\phi \longleftrightarrow \Box_{(\Box_{A!}B)!}\phi$$

Why was this not on the earlier list of the complete axiom system? The answer is a subtlety. That axiom system does indeed derive every valid formula. But it does so without being *substitution-closed*. In particular, the above basic axiom for atoms

$$\Diamond_{A!}p \longleftrightarrow A \wedge p$$

fails for arbitrary formulas ϕ. Define the *substitution core* of update logic as those valid schemata all of whose substitution instances are valid formulas.

Associativity belongs to it, but it is not derivable schematically from the earlier axiom system.

QUESTION. Can one axiomatize the complete substitution core of public update logic?

There are also interestingly invalid principles, witness the discussion of persistence in Section 4. Announcing a true statement "p, but you don't know it" invalidates itself. More technically, even when $p \wedge \Diamond_1 \neg p$ holds, its self-relativization

$$(p \wedge \Diamond_1 \neg p)^{p \wedge \Diamond_1 \neg p} = p \wedge \Diamond_1 \neg p \wedge \Diamond_1 (p \wedge \Diamond_1 \neg p \wedge p)$$

is a contradiction. Thus some assertions are self-refuting when announced, and the earlier-mentioned typographically pleasing principle is not part of a general logic of relativization:

$$\phi \longrightarrow (\phi)^\phi \qquad \text{holds only for special assertions } \phi.$$

We look at further issues in the logic of relativization in Section 7, including iterated announcement and its connections with fixed-point operators.

Excursion: richer systems of update. In standard logic, relativization often occurs together with other operations, such as *translation of predicates* — e.g., in the notion of *relative interpretation* of one theory into another. Likewise, the above connection extends to more sophisticated forms of epistemic update (cf. Section 9). For instance, when a group hears that a question is asked and answered, but only a subgroup gets the precise answer, we must use a new operation of *arrow elimination*, rather than world elimination. More precisely,

all arrows are removed for all members of that subgroup between those zones of the model that reflect different exhaustive answers.

Arrow elimination involves substitution of new accessibility relations for the current ones. E.g., when the question "ϕ?" is asked and answered, the uncertainty relations $\sim_i$ for agents i in the informed subgroup are replaced by the union of relations

$$(\phi)?; \sim_i; (\phi)? \cup (\neg\phi)?; \sim_i; (\neg\phi)?$$

This is a translation of the old binary relation $\sim_i$ into a new definable one.

Next on this road, there are more complex "product updates" — which correspond to those interpretations between theories which involve construction of new definable objects, like when we embed the rationals into the integers using ordered pairs. Axioms for update logics will then still axiomatize parts of the meta-theory of such general logical operations. Thus, progressively more complex notions of update correspond to more sophisticated theory relations from standard logic.

Finally, relativization suggests a slightly different view of eliminative update. So far, we discarded old information states. But now, we can keep the

old information state, and perform "virtual update" via relativized assertions. Thus, the initial state already contains all possible future communicative developments. Another take on this at least keeps the old models around, doing updates with *memory*. There are also independent reasons for maintaining some past history in our logic, having to do with public updates which refer explicitly to the "epistemic past", such as:

"what you said, I knew already".

Section 9 below has more concrete examples.

§7. **Repeated announcement and limit behaviour.** *"Keep talking".* In the Muddy Children scenario, an assertion of ignorance was repeated until it could no longer be made truly. In the given model, the statement was *self-defeating*: when repeated iteratively, it reaches a stage where it is not true anywhere. Of course, self-defeating ignorance statements lead to something good for us, viz. knowledge. There is also a counterpart to this limit behaviour: iterated announcement of *self-fulfilling* statements makes them common knowledge. This happened in one step with factual assertions and others in Section 4. Technically, this means that the process of repeatedly announcing A reaches a *fixed-point*. More subtle cases are discussed in [29], viz. repeated joint assertions of rationality by players in a strategic game, saying that one will only choose actions that may be best possible responses to what the others do. These may decrease the set of available strategy profiles until a "best zone" is reached consisting of either a Nash equilibrium, or at least some rationalizable profiles to be in.

Limits and fixed-points. Repeated announcement of rationality by two players *1*, *2* has the following simple modal form, which we take for granted here without motivation:

$$JR : \qquad \Diamond_1 B_1 \wedge \Diamond_2 B_2$$

Here the proposition letter B_i says that i's action in the current world is a *best response for i* to what the opponent is playing here. It can be shown that any finite game matrix has entries (worlds in the corresponding epistemic model) in a loop with alternating assertions B_i:

$$x_1 \models B_1 \sim_2 x_2 \models B_2 \sim_1 x_3 \models B_1 \sim_2 \cdots \sim_1 x_1 \models B_1$$

Repeated announcement of joint rationality JR may keep removing worlds, as each round may remove worlds satisfying a B_i on which one conjunct depended. But clearly, whole loops of the kind described remain all the time, as they form a kind of mutual protection society. Thus, we have

FACT 7.1. Strong Rationality is self-fulfilling on finite game models.

The above suggests extending update logic with *fixed-point operators*. This is like extending modal logic to a μ-calculus, whose syntax defines smallest

fixed-points $\mu p \bullet \phi(p)$ and greatest ones $\nu p \bullet \phi(p)$. [19] has details on this, [5] on more general fixed-point logics. We explore this a bit, as there are delicate issues involved (cf. again [29]). For a start, we have this

FACT 7.2. The stable set of worlds reached via repeated announcement of *JR* is defined inside the original full game model by the greatest fixed-point formula $\nu p \bullet (\diamondsuit_E(B_E \wedge p) \wedge \diamondsuit_A(B_A \wedge p))$.

Iterated announcement in dynamic logic. In any model M, we can keep announcing any formula ϕ, until we reach a fixed-point, empty or not:

$$\#(\phi, M)$$

E.g., *self-fulfilling* formulas ϕ in M become common knowledge in $\#(\phi, M)$:

$$\phi \longrightarrow (C_G \phi)^{\#(\phi, M)}$$

What kind of fixed-point are we computing here? Technically $\#(\phi, M)$ arises by continued application of the following set function, taking intersections at limit ordinals:

$$F_{M,\phi}(X) = \{s \in X | M|X, s \models \phi\}$$
$$\text{with } M|X \text{ the restriction of } M \text{ to the set } X.$$

This map F is *not monotone*, and the usual theory of fixed-points does not apply. The reason is the earlier fact that statements ϕ may change truth value when passing from M to submodels $M|X$. In particular, we do not recompute stages inside one unchanging model, as in the normal semantics of greatest fixed-point formulas $\nu p \bullet \phi(p)$, but in ever smaller models, changing the range of the modal operators. Thus we mix fixed-point computation with *relativization* (cf. Section 6). Despite F's non-monotonicity, iterated announcement is a fixed-point procedure of sorts:

FACT 7.3. The iterated announcement limit is definable as a so-called inflationary fixed-point.

PROOF. Take any ϕ, and relativize it to a fresh proposition letter p, yielding $(\phi)^p$. Here p need not always occur positively (it becomes negative when relativizing positive K-operators). Now the earlier epistemic Relativization Lemma says that

$$M, s \models (\phi)^p \quad \text{iff} \quad M|X, s \models \phi$$

Thus, the above definition of $F_{M,\phi}(X)$ as $\{s \in X | M|X, s \models \phi\}$ equals

$$\{s \in M | M, s \models (\phi)^p\} \cap X$$

And this computes a greatest *inflationary fixed-point* ([5]). ⊣

But then, why did iterated announcement of *JR* produce an ordinary greatest fixed-point in the epistemic μ-calculus? The above update map $F_{M,\phi}(X)$ is monotone with special sorts of formulas:

FACT 7.4. $F_{M,\phi}(X)$ is monotone for existential modal formulas.

The reason is that such formulas are preserved under model extensions, making their F monotone for set inclusion: cf. the related preservation issues in Section 4. As the epistemic μ-calculus is decidable, while its inflationary fixed-point extension is not, these issues of syntax matter to the complexity of reasoning about iterated announcement.

Excursion: comparing update sequences. Update logic is subtle, even here. What happens with different repeated announcements of rationality that players could make? [29] considers a weaker rationality assertion WR which follows from JR. Does this guarantee that their limits are included:

$$\#(SR, M) \subseteq \#(WR, M)?$$

The general answer is negative. Making weaker assertions repeatedly may lead to incomparable results. An example are the following formula ϕ and its consequence ψ:

$$\phi = p \wedge (\Diamond \neg p \longrightarrow \Diamond q)$$
$$\psi = \phi \vee (\neg p \wedge \neg q)$$

In the following epistemic model, the update sequence for ϕ stops in one step with the world *1*, whereas that for ψ runs: $\{1, 2, 3\}, \{1, 2\}, \{2\}$.

$$1 \overline{\qquad} 2 \overline{\qquad} 3$$
$$p, \neg q \quad \neg p, \neg q \quad \neg p, q$$

But sometimes, things work out.

FACT 7.5. If an *existential* ϕ implies ψ in M, then $\#(\phi, M) \subseteq \#(\psi, M)$.

PROOF. We always have the inclusion

$$T_\phi^\alpha(M) \subseteq T_\psi^\alpha(M)$$

The reason for this is the following implication:

$$\text{if } X \subseteq Y, \text{ then } F_{M,\phi}(X) \subseteq F_{M,\psi}(Y)$$

For, if $M|X, s \models \phi$ and $s \in X$, then $s \in Y$ and also $M|Y, s \models \phi$ — by the modal *existential* form of ϕ. But then we also have that $M|Y, s \models \psi$, by our valid implication. $\dashv$

One more type of fixed-point! Iterated announcement can be described by the finite iteration * of dynamic logic (cf. Section 2.2). This extension is studied in [2], which shows that formulas of the form

$$\Diamond_{(A!)^*} C_G A$$

are not definable in the modal μ-calculus. Still, it is known that formulas

$$\Box_{(A!)^*} \phi$$

with program iteration are definable with greatest fixed-point operators

$$\nu p \bullet \phi \wedge \square_{A!} p$$

But these cannot be analyzed in the earlier style, as they involve relativizing p to A, rather than the more tractable A to p, as happened in our analysis of repeated announcement.

§**8. Inference versus update.** *Dynamic inference.* Standard epistemic logic describes inference in unchanging information models. But the current literature also has a more lively notion following the dynamics of update (cf. [20]):

> Conclusion ϕ *follows dynamically* from premises $P_1, \ldots, P_k$ if after updating any information state with public announcements of the successive premises, all worlds in the end state satisfy ϕ.

Thus in dynamic-epistemic logic, the following implication must be valid:

$$\square_{P_1!;\ldots;P_k!} C_G \phi$$

This behaves differently from standard logic in its premise management:

> *Order of presentation matters*
> Conclusions from A, B need not be the same as from B, A:
> witness $\neg Kp, p$ (consistent) versus $p, \neg Kp$ (inconsistent).

> *Multiplicity of occurrence matters*
> $\neg Kp \wedge p$ has different update effects from $(\neg Kp \wedge p) \wedge (\neg Kp \wedge p)$.

> *Adding premises can disturb conclusions*
> $\neg Kp$ implies $\neg Kp$ — but $\neg Kp, p$ does not imply $\neg Kp$.

By contrast, the structural rules of classical logic say that order, multiplicity, and overkill does not matter. Nevertheless, we have a description.
Structural rules and representation. [32] has three modified structural rules that are valid for dynamic inference as defined above:

Left Monotonicity	$X \Rightarrow A$ implies $B, X \Rightarrow A$
Cautious Monotonicity	$X \Rightarrow A$ and $X, Y \Rightarrow B$ imply $X, A, Y \Rightarrow B$
Left Cut	$X \Rightarrow A$ and $X, A, Y \Rightarrow B$ imply $X, Y \Rightarrow B$

Moreover, the following completeness result holds:

THEOREM 8.1. *The structural properties of dynamic inference are axiomatized completely by the rules of Left Monotonicity, Cautious Monotonicity, and Left Cut.*

The core of the proof is a representation argument showing that any abstract finite tree model for modal logic can be represented up to bisimulation in the

following form:

> Worlds w go to a family of epistemic models M_w
> Basic actions a go to suitable epistemic announcements (ϕ_a)!

This suggests that public update is a quite general process, which can encode arbitrary processes in the form of "conversation games".

Inference versus update. Here is a more general moral. Logic has two different inferential processes. The first is *ordinary inference*, related to implications $A \rightarrow B$. This stays inside one fixed model. The second process is a model-jumping *inference under relativization*, related to the earlier formulas $\Box_{A!}\ B$. Both are important, and update logics mix them.

Even so, one empirical issue remains. Muddy children *deduce* a solution: they did not draw update diagrams. What is going on inside our heads?

§**9. The wider world.** As stated in Section 1, update analysis has two main directions: increased coverage by means of new models and mechanisms, and increased in-depth understanding of the logics that we already have. This paper has concentrated on the latter, hoping to show the interest of logical issues in communication. In this Section, the reader gets a lightning tour of what she has missed by taking this prong of the fork.

Keeping track of past assertions. Some puzzles involve reference to past states, with people saying things like "What you said did not surprise me" ([11]). This says they knew at the previous state, calling for a further update there. To accomplish this, we need to maintain a stack of past updates, instead of just performing them and trashing previous stages. In the limit, as earlier, this richer framework might also include protocol information about the sort of communication we are in.

Privacy and hiding. The next stage of complexity beyond public communication involves hiding information, either willfully, or through partial observation. Here is about the simplest example, from [22]:

> *We have both just drawn a closed envelope. It is common knowledge between us that one envelope holds an invitation to a lecture on logic, the other to a wild night out in Amsterdam. We are both ignorant of the fate in store for us! Now I open my envelope, and read the contents, without showing them to you. Yours remains closed. Which information has passed exactly because of my action? I certainly know now which fate is in store for me. But you have also learnt something, viz.* that I know - *though not what I know. Likewise, I did not just learn what is in my envelope. I also learnt something about you, viz. that you know that I know. The latter fact has even become common knowledge between us. And so on. What is a general principle behind this?*
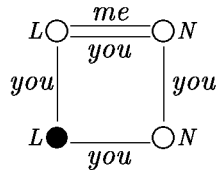
The initial information state is a familiar one of collective ignorance:

$$L\bullet\!\!=\!\!\!\overset{me}{\underset{you}{=\!\!\!=\!\!\!=}}\!\!\circ N$$
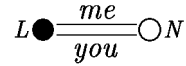
The intuitive update just removes my uncertainty link — while both worlds remain available, as they are needed to model your continued ignorance of the base fact:

$$L\bullet\!\!\underset{you}{\rule{2cm}{0.4pt}}\!\!\circ N$$
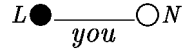
Such updates occur frequently in card games, when players publicly show cards to some others, but not to all. But card updates can also blow up the size of a model. Suppose I opened my envelope, but you cannot tell if I read the card in it or not. Let us say that in fact, I did look. In that case, the intuitive update is to the model

$$
\begin{array}{ccc}
L\circ & \overset{me}{\underset{you}{=\!\!\!=\!\!\!=}} & \circ N \\
{\scriptstyle you}\big| & & \big|{\scriptstyle you} \\
L\bullet & \underset{you}{\rule{1.5cm}{0.4pt}} & \circ N
\end{array}
$$

The road to this broader kind of update leads via the key publications [6, 14, 40]. The general idea is as follows. Complex communication involves two ingredients: a current information model M, and another epistemic model A of possible physical actions, which agents may not be able to distinguish. Moreover, these actions come with *preconditions* on their successful execution. E.g., truthful public announcement $A!$ can only happen in worlds where A holds. General update takes a *product* of the two models, giving a new information model MxA whose states (s, a) record new actions taken at s, provided the preconditions of a is satisfied in M, s. This may transform the old model M drastically. The basic epistemic stipulation is this. Uncertainty among new states can only come from existing uncertainty via indistinguishable actions:

$$(s, a) \sim_i (t, b) \quad \text{iff both } s \sim_i t \text{ and } a \sim_i b$$

In the first card example, the actions were "read lecture", "read night out". Taking the proper preconditions into account and then computing the new uncertainties gives the correct updated model

$$read\ lecture\ \bullet\!\!\underset{you}{\rule{1.5cm}{0.4pt}}\!\!\circ read\ night\ out$$
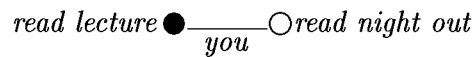
The second example involved a third action "do nothing" with precondition $\top$, which I can distinguish from the first two, but you cannot. Product update delivers a model with four worlds $(L, readL)$, $(L, do\ nothing)$,

$(N, read\ N), (N, do\ nothing)$ — with the agents' uncertainties as computed by product update exactly as those of the earlier intuitive diagram.

Clearly, truth values of epistemic propositions can change drastically in product update. Dynamic-epistemic logic now gets very exciting, involving combining epistemic formulas true in M and epistemic information about A expressed in a suitable language. Many of the concerns for public update in this paper will return in more sophisticated versions. Here is one intriguing example of a new issue concerning general update — for a much broader list of new open problems in update logic, cf. [33].

Evolution in the Update Universe. Unlike updates for public announcement, product update can blow up the size of the initial input model M. Here is a simple illustration of this possible blow-up of information models.

EXAMPLE 1. Suppose a public announcement of the true fact P takes place in a group $\{1, 2\}$, but 2 is not sure whether it was an announcement of P, or just some identity action Id which could happen anywhere. In that case, a two-world model with P and $\neg P$ turns into a three-world model with states *(p, "P!"), (p, Id)* and *(¬P, Id).*

A typical example of such blow-up occurs again in *games*. Players start from an initial situation M, say a deal of cards, and the action model A contains all possible moves that they have — with preconditions restricting when these are available at players' turns. The game tree consists of all possible evolutions through the nodes in a tree model. Such structures may be defined more precisely as follows:

DEFINITION 9.1. Let an initial epistemic model M be given, and also an action model A. Then the *update evolution model* $TREE(M, A)$ is the infinite epistemic model consisting of disjoint copies of all successive product update layers

$$M, M x A, (M x A) x A, \dots$$

[28] gives necessary and sufficient conditions on agents' abilities under which an arbitrary extensive game with imperfect information can be represented in the format $TREE(M, A)$. Next, as we have seen, the horizontal stages of $TREE(M, A)$ can grow in size all the time. This blow-up by product update is noticeable with sophisticated communication moves in parlour games such as "Cluedo" ([40]). But often, this infinity seems spurious — as happens in many games, where complexity of information can grow in mid-play, but then decrease again toward the end game. The counter-acting force is epistemic *bisimulation contraction*.
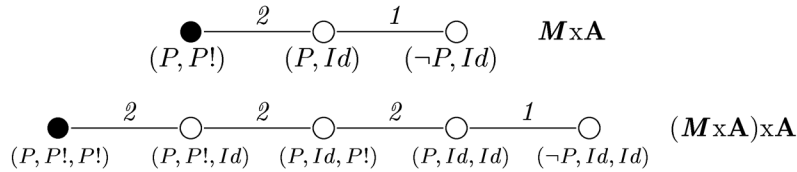
EXAMPLE 2. Stabilization under bisimulation. Consider a model with two worlds $P, \neg P$, between wich agent 1 is uncertain, though 2 is not. The actual world has P.

$$\bullet \xrightarrow{\quad 1 \quad} \circ \qquad \boldsymbol{M}$$
$$P \qquad\qquad \neg P$$
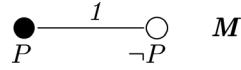
Now a true announcement of P takes place, and agent 1 hears this. But agent 2 thinks the announcement might just be a statement "True" which could hold anywhere. The action model for this scenario looks as follows:

$$\bullet \xrightarrow{\quad 2 \quad} \circ \qquad \boldsymbol{A}$$
$$(Precondition: P) \; P! \qquad Id \; (Precondition: T)$$
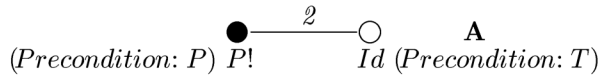
The next two levels of $TREE(A, A)$ then look as follows:

$$\bullet \xrightarrow{\quad 2 \quad} \circ \xrightarrow{\quad 1 \quad} \circ \qquad \boldsymbol{M}\mathrm{x}\boldsymbol{A}$$
$$(P, P!) \qquad (P, Id) \qquad (\neg P, Id)$$

$$\bullet \xrightarrow{\; 2 \;} \circ \xrightarrow{\; 2 \;} \circ \xrightarrow{\; 2 \;} \circ \xrightarrow{\; 1 \;} \circ \qquad (\boldsymbol{M}\mathrm{x}\boldsymbol{A})\mathrm{x}\boldsymbol{A}$$
$$(P, P!, P!) \quad (P, P!, Id) \quad (P, Id, P!) \quad (P, Id, Id) \quad (\neg P, Id, Id)$$

There is an *epistemic bisimulation* between the last two levels, linking the lower three worlds to the left with the single world *(P, P!)* in $\boldsymbol{M}\mathrm{x}\boldsymbol{A}$. Thus, $(\boldsymbol{M}\mathrm{x}\boldsymbol{A})\mathrm{x}\boldsymbol{A}$ is bisimilar with $\boldsymbol{M}\mathrm{x}\boldsymbol{A}$, and the potentially infinite iteration in the update evolution remains finite modulo bisimulation.

Indeed, the original version of this paper stated a *"Finite Evolution Conjecture"* saying that, starting from a given finite $\boldsymbol{M}$ and $\boldsymbol{A}$, the model $TREE(\boldsymbol{M}, \boldsymbol{A})$ always remains finite modulo bisimulation. This would imply that some levels in the tree must be bisimilar, say at depth k and $l > k$ — after which, infinite "oscillation" follows. But this was refuted in [18], which shows that the Conjecture fails in some models with two $S5$-agents. Nevertheless, the Conjecture holds for many special cases of such models, e.g., when the epistemic accessibility relations for all agents in the model $\boldsymbol{A}$ are linearly ordered by inclusion. [18] uses finite pebble games over $\boldsymbol{M}$ and $\boldsymbol{A}$ to determine when $TREE(\boldsymbol{M}, \boldsymbol{A})$ is finite modulo bisimulation. This is the case iff the "responding player" in the pebble game has a winning strategy. The computational complexity of having a finite update evolution modulo bisimulation is still unknown. Thus large classes of communicative scenarios might still satisfy the Finite Evolution Conjecture — e.g., those corresponding to most parlour games.

§**10. General communication.** *Complexity and diversity.* General tools like product update, single-step or iterated, can chart many varieties of communication, and their broad patterns. In particular, one then finds natural *thresholds* in types of communication. One such threshold occurs when passing from mere partial information to misleading, lying and cheating. In principle, product update also describes the latter, but there is a lot of fine-structure yet

to be understood. A more technological challenge in this realm are hiding mechanisms, such as security protocols on the Internet.

Another source of variation are people's limitations, such as bounded memory or limited attention span. Epistemic updates can be performed by ideal agents, but the reality is, of course, that agents may have *bounded rationality*. This diversity of agents calls for variants of the above product update mechanism, which are found in [35].

Games and social software. Again, there is not just analysis, but also synthesis. Communication involves planning what we say, for a purpose. The proper broader setting for this are *games*, which involve preferences and strategies (cf. [1, 26, 31]). This is one instance of what has been called "social software" recently: the design of mechanisms satisfying epistemic specifications such as who gets to know what (cf. [15, 16]). For a very concrete example of social software (and hardware), just think of your *email*! When you announce a fact to one friend, and send some *cc*'s, the effect is public announcement. But if you use the "blind carbon-copy" button *bcc*, the new information state requires product update which may increase complexity. Handling email and the many communication options it offers is a logically highly non-trivial task.

Communication channels. After all these sweeping vistas, we come down to earth with a simple example, again based on a puzzle. The 1998 "National Science Quiz" of the Dutch national research agency NWO ([42]) had the following question:

> Six people each know a secret. In one telephone call, two of them can share all secrets they have. What is the minimal number of calls that they have to make in order to ensure that all secrets become known to everyone?

The answers offered were: *7*, *8*, *9*. The correct one turns out to be *8*. For N people, $2N - 4$ calls turns out to be optimal, a result which is not deep but surprisingly difficult to prove. The best algorithm notes that four people can share all secrets that they have in four steps:

> *1 calls 2, 3 calls 4, 1 calls 3, 2 calls 4.*

So, single out any four people in the total group.

> First let the other $N - 4$ call one of them, then let the four people share all secrets that they have, then let the $N - 4$ people call back to their informant.

The total number of calls will be $2N - 4$. Now, this clearly raises a general question. What happens to update logic when we make a further semantic parameter explicit, viz. the *communication network*? Our running example of public announcement presupposed some public broadcast system. The gossip puzzle assumes two-by-two telephone connections without conference calls.

We can look for results linking up desired outcomes with properties of the network. E.g., it is easy to show that:

> Universal knowledge of secrets can be achieved if and only if the network is *connected*: every two people must be connectible by some sequence of calls.

But there are many other intriguing phenomena. Suppose three generals with armies on different hilltops are planning a joint attack on the adversary in the plain below. They have completely reliable two-way telephone lines. One of them possesses some piece of information *p* which has to become common knowledge among them in order to execute a successful coordinated attack. Can they achieve this common knowledge of *p*? The answer is that it depends on the scenario, or *protocol*. But this involves genuine extensions of the update logic framework to a temporal setting ([13, 30]) — as well as structured groups with *channels*.

If the generals only communicate secrets, even including information about all calls they made, then common knowledge is unattainable, just as in the more familiar two-generals problem with unreliable communication. Informally, there is always someone who is not sure that the last communication took place. More precisely, product update allowing for this uncertainty will leave at least one agent uncertainty chain from the actual world to a $\neg p$-world, preventing common knowledge. But what about general A phoning general B, sharing the information, and telling him that he will call C, tell him about this whole conversation, including the promise to call him? This is like mediators going back and forth between estranged parties. Can this produce common knowledge? Again, it depends on whether agents are sure that promises are carried out. If they are, then a scenario arises with actions and observations where product update will indeed deliver common knowledge.

We leave matters at this informal stage here. Our aim in this excursion has merely been to show that update logic fits naturally with other aspects of communication, such as the availability and reliability of channels.

§11. Logic and communication. Traditionally, logic is about reasoning. If I want to find out something, I sit back in my chair, close my eyes, and think. Of course, I might also just go out, and ask someone, but this seems like cheating. At the University of Groningen, we once did a seminar reading Newton's two great works: *Principia Mathematica*, and the *Optics*. The first was fine: pure deduction, and facts only admitted when they do not spoil the show. But the *Optics* was a shock, for being so terribly unprincipled! Its essential axioms even include some brute facts, for which you have to go out on a sunny day, and see what light does on when falling on prisms or films. For Newton, what we can observe is as hard a fact as what we can deduce. The same is true for ordinary life: questions to nature, or other knowledgeable sources such

as people, provide hard information. And the general point of this paper is that logic has a lot to say about this, too. One can see this as an extension of the agenda, and it certainly is. But eventually, it may also have repercussions for the original heartland. Say, what would be the crucial desirable meta-properties of first-order logic when we add the analysis of communication as one of its core tasks?

I will not elaborate on this, as this paper has already taken up too much of the reader's time. And in any case, the time for mere communication has passed. Now that you have told me about the road to the Colosseum, the time has come for me to use this knowledge, and act according to my soldier's pledge, made long ago.

§**12. Added in print.** This paper was written in 2002, when it appeared as an *ILLC* Tech Report in Amsterdam. In the meantime, various new developments have taken place, including a better insight into the Update Universe ([18]), solutions to some of the problems mentioned here (cf. the new survey [33]), uniform axiomatizations of dynamic epistemic logics with common knowledge in terms of reduction axioms ([37]), and a better view of fruitfull connections with temporal logic and game theory ([34]). Also, the first textbook on dynamic epistemic logic is preparation, cf. [38].

REFERENCES

[1] A. Baltag, *Logics for insecure communication*, **Proceedings TARK VIII** (Los Altos), Morgan Kaufmann, 2001, pp. 111–121.

[2] A. Baltag, L. Moss, and S. Solecki, *Updated version of "The Logic of Public Announcements, Common Knowledge and Private Suspicions, Proceedings TARK 1998"*, 2003, Department of Cognitive Science, Indiana University, Bloomington and Department of Computing, Oxford University.

[3] J. Barwise and L. Moss, **Vicious Circles**, CSLI Publications, Stanford, 1997.

[4] P. Blackburn, M. de Rijke, and Y. Venema, **Modal Logic**, Cambridge University Press, Cambridge, 2001.

[5] H.-D. Ebbinghaus and J. Flum, **Finite Model Theory**, Springer, Berlin, 1995.

[6] J. Gerbrandy, *Bisimulations on Planet Kripke*, Technical Report DS-1999-01, Institute for Logic, Language and Computation, University of Amsterdam, 1999.

[7] J. Halpern and M. Vardi, *'the complexity of reasoning about knowledge and time'*, **Journal of Computer and Systems Science**, vol. 38 (1989), no. 1, pp. 195–237.

[8] M. Hollenberg, **Logic and Bisimulation**, vol. XIV, Publications Zeno Institute of Philosophy, Zeno Institute of Philosophy, University of Utrecht, 1998.

[9] D. Kozen, D. Harel, and J. Tiuryn, **Dynamic Logic**, MIT Press, Cambridge (Mass.), 2000.

[10] M. Marx and Y. Venema, **Multi-Dimensional Modal Logic**, Kluwer Academic Publishers, Dordrecht, 1997.

[11] J. McCarthy, *Two Puzzles About Knowledge*, 2002.

[12] B. Moore, *A Formal Theory of Knowledge and Action*, research report, SRI International, Menlo Park, 1985.

[13] Y. Moses, M. Vardi, R. Fagin, and J. Halpern, *Reasoning About Knowledge*, MIT Press, Cambridge (Mass.), 1995.

[14] L. Moss, S. Solecki, and A. Baltag, *The logic of public announcements, common knowledge and private suspicions*, **Proceedings TARK 1998** (Los Altos), Morgan Kaufmann Publishers, 1998, pp. 43–56.

[15] R. Parikh, *Social software*, **Synthese**, vol. 132 (2002), pp. 187–211.

[16] M. Pauly, *Logic for Social Software*, Technical Report DS-2001-10, Institute for Logic, Language and Computation, University of Amsterdam, 2001.

[17] J. Plaza, *Logics of public announcements*, **Proceedings 4th International Symposium on Methodologies for Intelligent Systems**, 1989.

[18] T. Sadzik, *Dynamic Epistemic Update System*, 2004.

[19] C. Stirling, *Bisimulation, modal logic, and model checking games*, **Logic Journal of the IGPL. Interest Group in Pure and Applied Logics**, vol. 7 (1999), no. 1, pp. 103–124, Special issue on Temporal Logic.

[20] J. van Benthem, **Exploring Logical Dynamics**, CSLI Publications, Stanford, 1996.

[21] ———, *Dynamic Bits and Pieces*, Report LP-97-01, Institute for Logic, Language and Computation, University of Amsterdam, 1997.

[22] ———, *Annual Report Spinoza Award Project "logic in action"*, Technical report, Institute for Logic, Language and Computation, University of Amsterdam, 1998.

[23] ———, *Dynamic Odds and Ends*, Report ML-98-08, Institute for Logic, Language and Computation, University of Amsterdam, 1998.

[24] ———, *Radical Epistemic Dynamic Logic*, note for course "logic in games", Institute for Logic, Language and Computation, University of Amsterdam, 1999.

[25] ———, *Update as Relativization*, manuscript, Institute for Logic, Language and Computation, University of Amsterdam, 1999.

[26] ———, *Logic in Games*, lecture notes, Institute for Logic, Language and Computation, University of Amsterdam, 1999–2002, http://staff.science.uva.nl/~johan.

[27] ———, *Update Delights*, invited lecture, esslli summer school, birmingham, and manuscript, Institute for Logic, Language and Computation, University of Amsterdam, 2000.

[28] ———, *Games in dynamic epistemic logic*, **Bulletin of Economic Research**, vol. 53 (2001), no. 4, pp. 219–248.

[29] ———, *Rational Dynamics*, Invited lecture, logamas workshop, department of computer science, university of liverpool, Institute for Logic, Language and Computation, University of Amsterdam, 2002.

[30] ———, *Logic and The Dynamics of Information*, 2003.

[31] ———, *Rational dynamics and epistemic logic in games*, **Logic, Game Theory and Social Choice** (S. Vannucci, editor), vol. III, University of Siena, Department of Political Economy, 2003, pp. 19–23.

[32] ———, *Structural properties of dynamic reasoning*, **The Dynamic Turn** (Amsterdam) (J. Peregrin, editor), Kluwer Academic Publishers, 2003, pp. 15–31.

[33] ———, *Open problems in update logic*, **Mathematical Problems from Applied Logic** (Russian Academy of Sciences, Novosibirsk) (N. Rozhkowska, editor), Plenum Publishers, 2005.

[34] ———, *Update and Revision in Games, paper for APA-ASL Meeting, San Francisco. Continuation of "A Mini-Guide to Logic in Action"*, 2005, Philosophical Researches 2004, Supplement, 21–30, Chinese Academy of Sciences, Beijing.

[35] J. van Benthem and F. Liu, *Epistemic Update for Resource-Bounded Agents*, working paper, Institute for Logic, Language and Computation, University of Amsterdam, 2004.

[36] J. van Benthem, I. Németi, and H. Andréka, *Modal logics and bounded fragments of predicate logic*, **Journal of Philosophical Logic**, vol. 27 (1998), no. 3, pp. 217–274.

[37] J. van Benthem, J. van Eijck, and B. Kooi, *A Logic for Communication and Change*, Technical report, ILLC & CWI Amsterdam, Philosophy Department, Groningen, 2005.

[38] W. van der Hoek, B. Kooi, and H. van Ditmarsch, *Dynamic Epistemic Logic*, Kluwer-Springer Academic Publishers, Dordrecht, to appear.

[39] R. van der Meyden, *Common knowledge and update in finite environments*, **Information and Computation**, vol. 140 (1998), no. 2, pp. 115–157.

[40] H. vanDitmarsch, *Knowledge Games*, Technical Report DS-2000-06, Institute for Logic, Language and Computation, University of Amsterdam, 2000.

[41] ———, *Keeping Secrets With Public Communication*, Technical report, Department of Computer Science, University of Otago, 2002.

[42] Nationale Wetenschapsquiz, *Gossip Puzzle*, 1998, http://www.nwo.nl/.

[43] M. Wooldridge, **An Introduction to Multi-Agent Systems**, John Wiley, Colchester, 2002.

INSTITUTE FOR LOGIC
LANGUAGE AND COMPUTATION
UNIVERSITY OF AMSTERDAM
PLANTAGE MUIDERGRACHT 24
1018 TV AMSTERDAM, NL
and
DEPARTMENT OF PHILOSOPHY
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305
USA
E-mail: johan@science.uva.nl