Logics for Information Update

a tutorial for *TARK Siena 2001*

Johan van Benthem
johan@{illc.uva.nl, csli.stanford.edu}
ILLC Amsterdam & CSLI Stanford

Abstract

Logical languages define propositions that describe states of the world, represented by some suitably chosen model. The typical format for this is a semantic truth definition

$\mathbf{M}, s \models \phi$ saying that formula ϕ is true at state s of model $\mathbf{M}$.

In terms of this schema, one can then define valid reasoning as all steps from formulas to formulas that preserve truth. The usual systems of epistemic logic follow this set-up, and the resulting account of reasoning is useful and well-known. But there is also a limitation to this perspective. It leaves out the 'logical dynamics' of many natural processes, such as communicating information. In that case, the basic phenomena are actions which change information states of speakers and hearers, and the locus of 'meaning' for a proposition is not its static content, but the dynamic *change* it induces from one model M, viewed as an information state for logical agents, to another:

announcing ϕ changes $\mathbf{M}, s$ into $\mathbf{N}, t$.

One now becomes interested in cognitive actions like informing, questioning, answering - with inference as an important, but by no means the only example. This Dynamic Turn has emerged in many areas: linguistics (Kamp, Heim, Groenendijk & Stokhof), philosophy and AI (Gärdenfors, Harman), computational linguistics (Gross & Sidner, Gabbay, Kempson & Meyer Viol) and computer science (Fagin/Halpern/Moses/Vardi, Abramsky, Reiter).

This tutorial presents an overview of update logics for communicative actions. The basic tool is: Hintikka-Kripke models for epistemic languages. The difficult questions are twofold. First, what precise updates are induced by various types of communicative action? One line in the literature has looked at examples of increasing complexity. Models and update procedures are simple for questions/answers, or public announcements, but they get much more complex as communication gets more 'private'. We follow one such line in current research, referring to work by Veltman, Groeneveld & Gerbrandy, and Baltag, Solecki & Moss on information update in conversation and games, whose current challenges include more complex linguistic expressions, defaults and cryptographic communication. This requires building sophisticated models in harmony with the observed phenomena.

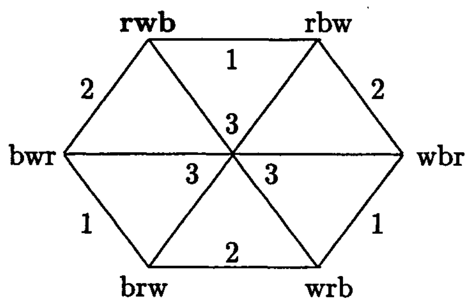
Our second topic concerns new logical issues in this setting. In a way, all relevant dynamics is contained in existing epistemic logics cum temporal models. But this observation is no more useful than saying we can do most of modern logic inside first-order logic, and all of it inside set theory. Instead, we survey new questions beyond the usual agenda. Examples are: (a) relating model-theoretic to syntactic views of update, (b) (non-)persistence of assertions under update, (c) calculi for short-term dynamic 'local inference', and connections with longer-term 'global inference'.

References for the general program (1) J. van Benthem, 1996, "Exploring Logical Dynamics", CSLI Publications, Stanford. (2) R. Muskens & A. Visser, 'Dynamics', in J. van Benthem & A. ter Meulen, eds., 1997, "Handbook of Logic & Language", Elsevier, Amsterdam. (3) J. van Benthem, 2000, 'Update Delights', http:// turing. wins. uva. nl/~johan/Update.Delights.ps, ILLC, University of Amsterdam – for proofs and details.

The rest of this paper is a brief tour of some technical issues in update logic, starting with a simple card game, which highlights some essential questions. The tutorial at TARK itself will be example-oriented.

1 How information flows

Three cards *r* ("red"), *w* ("white") and *b* ("blue") are distributed over three players. This can happen in exactly 6 ways, which are the possible 'states of the world' in this scenario. Let us write *rwb* for: "1 has red, 2 has white, 3 has blue", and likewise for the other 5 distributions. The actual distribution in our story happens to be **rwb**. Now each player is allowed to look at his own card, but cannot see that of the others. Thus, no one can distinguish a situation from that with the cards of the other players interchanged. The resulting *collective information state* may be pictured as follows:



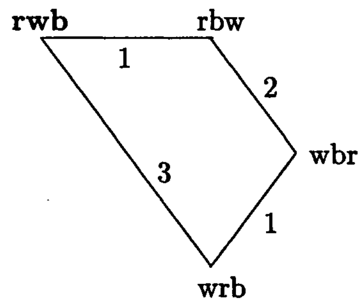
The lines indicate uncertainties for the relevant players. Note that, even though they are in **rwb**, no player knows this. Indeed, as they ponder what other players know or do not know about the cards and the others' information, they must take into account each of the 6 possible worlds. Now players start asking questions and giving answers. First, 2 asks 1

"Do you have the blue card?",

and the truthful answer is forthcoming:

"No"

As a result, each player now knows that 1 does not have the blue card. But apart from that, the effects are quite different. Player 2 has obtained enough information to figure out the exact situation: he has white, 3 must therefore have blue, and 1 must have red. Player 3 already knew that 1 did not have the blue card, so he learns less - and player 1 obviously does not learn much from telling the truth about himself. But this is just a first approximation. For simultaneously, players 3 now also understands that 2 has figured out where he is, which does represent an advance in 3's 'higher-order' knowledge. By contrast, 1 only knows that one of 2 or 3 must be the lucky knower - but cannot tell which one, because of the obvious symmetry. We can describe this episode in more and more subtle linguistic terms, but where to stop? A picture does better. 1's statement has ruled out all worlds where the first position is a 'b', and the players are left with this information state:



In the picture, looking at **rwb**, we see the truth of the above assertions at once. Moreover, we can make further predictions. For instance, only one player has the power now to make the actual distribution of cards into *common knowledge* by mentioning his own card, viz. player player 2. The resulting information state is just

rwb

By contrast, if 3 announces his card, this tells 2 the actual distribution of the cards - but 3 remains uncertain between *rwb* and *wrb*, while knowing the others are completely informed about the cards, (and 3's uncertainty).

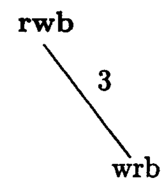
Information is passed each time people use language. Our opening story demonstrates a fairly typical episode of communication, where people offer and absorb information. Analyzing this flow of information in major types of communicative situation seems a typical task for logical analysis, on a par with the traditional concerns of 'inference' or 'meaning'. But do we really need logic here? Well it is not *easy* to capture the informational effects of simple speech acts completely in informal terms. For instance, the true effect of 1's assertion was the pictured update, and we have not even begun to give a complete 'linguistic' description of the group knowledge that answering a question is supposed to achieve. Well, but is not that easy after all? A Yes/No question asks whether some proposition A holds, and a positive answer "Yes" serves simply to make A into *common knowledge* for the group concerned: at least {Questioner, Answerer}. And the conversational law behind that must surely be this truism:

> *Public announcement of a statement A makes sure that A becomes* common knowledge *in the group: after all, the group just* learns *that A is the case.*

But this 'evident' principle is *wrong.* Suppose that after 1's initial assertion about not having the blue card, 3 makes the following true assertion:

"I don't know if 2 has the white card, and neither does 1"

The effect of this will include the following. 1 now learns that 3 was the player who was uncertain, and therefore 3 must have had the blue card. Thus, 1 learns the true distribution from 3's utterance, whence that the statement about his uncertainty is no longer true, and can never be common knowledge. In a picture, the update is this:



Here, what 3 said is *not* common knowledge, because the very ignorance announced enlightens at least one player involved in it! More generally, public announcement tells us how things *were* before the announcement, and the informational setting may change precisely as a result of it. This is a classical subtlety in dynamic logics of action involving changing situations. So, logical analysis helps spot pitfalls underneath communication.

The above graphs are a 'video' of successive information states, where each speech act transforms the current picture. But the separate 'movie frames' still come from standard logic: the above graphs are just the ordinary models for modal or epistemic logic, though now used in a 'dynamic mode'. This extension of the logical agenda may be found in [Gärdenfors87, Fagin et al.95, van Benthem96, Veltman96] and [Moss00, Baltag et al.99, Gerbrandy98], and [van Ditmarsch00] (from which we took the card example).

2 Public update

A typical building block of communication is a question/answer episode

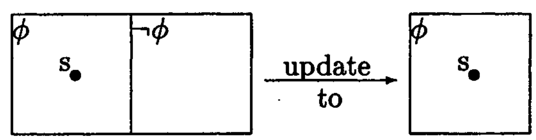
Q: ϕ?
A: Yes.

In a simple setting, this might happen in the following initial state, where ϕ is true, and **A** knows this, while **Q** only knows that **A** knows *if* ϕ:

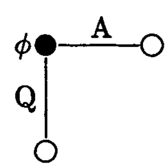
The black dot is the real world. **A**'s answer updates to the one-point model:

Epistemic logic describes the 'statics' of what is true at worlds in epistemic models. But communicative actions are 'dynamic': they *change* those models. Public announcement of proposition ϕ to a group changes the current information state via the following simple instruction:

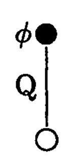
eliminate all worlds which currently do not satisfy ϕ



The idea here is that everyone understands that the eliminated worlds have been ruled out as candidates for the actual situation, and this update is uniform, with the same repercussions for every agent in the group. This simple procedure will not suffice for more delicate form of communication. Still it is the basic case on which all else rests - and we will see it has lots of surprising subtleties. Take again a question-answer episode, but this time, in this 3-world model. Questions themselves may convey information:



In the black state, **Q** thinks **A** might know the answer, as this is so in the bottom-most state. (Curiously, **Q** has this illusion about **A** on the basis of a factually wrong alternative ...) But in the right-most state **Q** knows that **A** does not know, and hence she would not ask there. The very question therefore tells **A** he must be in the black state, triggering an update to:



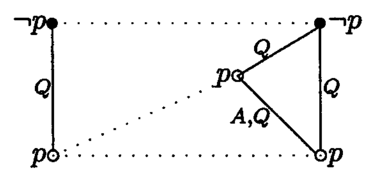
Now **A** knows, and updates with the correct answer to one single world!

3 Modal structures

Update logic is essentially modal or epistemic. We review some basics. First, expressive power of a language is usually measured by similarity between different models. For modal logic, this is 'bisimulation', a notion often motivated from an action-oriented process perspective - but which makes equals sense when comparing different information models.

Definition *Bisimulation*
A bisimulation between two models **M**, **N** is a binary relation E between their states m, n s.t. whenever $m\mathrm{E}n$, then (a) m, n satisfy the same proposition letters, (b1) if $m\mathrm{R}m'$, then there exists a world n' with $n\mathrm{R}n'$ and $m'\mathrm{E}n'$, (b2) the same 'zigzag clause' holds in the opposite direction.

Our question-answer example has a bisimulation with a 'variant', by the dotted lines in the following picture:



In a natural sense, these are two representations of the same group information state. This shows in the following basic results, which connect up bisimulation and satisfaction of modal or epistemic formulas. For convenience, we restrict attention to *finite models.* These can be general modal ones, or the usual multi-S5 models for epistemic languages.

Theorem *Modal Invariance*
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

Proof
See [van Benthem96, Blackburn et al.01]. □

Theorem *State Definition*
For each model **M**,s there exists an epistemic formula $\beta_{\mathbf{M},s}$ s.t. the following are equivalent:

(a) $\mathbf{N},t \models \beta_{\mathbf{M},s}$

(b) **N**,t is bisimilar to **M**,s

Proof
See [Barwise et al.97]. □

The Invariance Lemma says that bisimulation has exactly the discriminating power of the modal language. The State Definition Lemma says each semantic state can be characterized completely by one modal or epistemic formula. Except with single-agent S5-models, basic modal formulas do not suffice for this purpose. One needs an extension to dynamic logic, or an epistemic common knowledge operator - which is fine in our setting.

Example *Defining epistemic states*
Consider our standard 2-world question-answer example again. Here is an epistemic formula which defines its p-state up to bisimulation:

$$\mathrm{p} \,\&\, C_{\{Q,A\}}\ ((K_A\mathrm{p} \vee K_A\neg\mathrm{p}) \,\&\, \neg K_Q\mathrm{p} \,\&\, \neg K_Q\neg\mathrm{p})$$

The formula divides its models into two disjoint zones, defined by K_Ap and $K_A\neg$p. Moreover, unpacking some consequences, it states: "If you are in K_Ap , you can **A**-see K_Ap but not (of course) $K_A\neg$p ; and vice versa. Moreover, if you are in K_Ap, you can **Q**-see K_Ap but also $K_A\neg$p; and vice versa." This is all the information we need for a bisimulation. Atomic facts p, $\neg$p needed not be encoded explicitly, as the two leading formulas K_1p, $K_2\neg$p make it obvious how they lie.

Each information model is bisimilar to its *bisimulation contraction*, a smallest model modally equivalent to it. This setting also imposes technical conditions on update operations O that change models. A natural requirement on these is that they should *respect bisimulation* [van Benthem96]:

if **M**,s and **N**,t are bisimilar, so are their values O(**M**,s) and O(**N**,t)

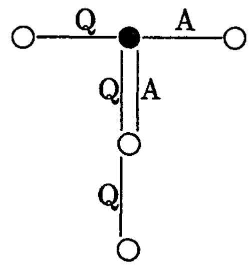
Theorem Public update respects bisimulation.

4 Communication: 'the best we can do'

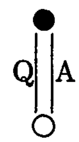
Models record update episodes. But they also suggest updates of their own. Consider two epistemic agents that find themselves in some collective information state **M**, at some actual situation s . They can tell each other things they know, thereby cutting down the model to smaller sizes. Suppose they wish to be maximally cooperative:

What is the best correct information they can give 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. This minimal model is *unique*, and serves as the 'communicative core' of the initial model. Here is an explicit description:

Theorem
The 'communicative core' of a model is the set of worlds that are reachable from the actual world via all uncertainty links.

Via this intersection of uncertainty relations, these worlds encode precisely what has been called 'implicit knowledge' of a group of agents [Fagin et al.95] Thus, communication converts implicit knowledge into common knowledge. This was also the point of the second question-answer in Section 2 above.

Corollary Agents need only 2 rounds of communication to get to the core.

In particular, there is no need for repetitions on the part of 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). He might have said A & $(B)^A$ straightaway: which $(B)^A$ the *relativization* of B to A.

5 Learning and persistence

It seems easy to say what public announcement of statement ϕ does. Everyone learns that ϕ is the case, which becomes common knowledge in the group. This is true for atomic facts p, and other types of assertion. But not all updates with ϕ result in common knowledge of ϕ!

Example *Biting one's tail*
In a question-answer update, let **A** make the true announcement:

$$p \ \& \neg K_Q p \quad \text{("p, but you don't know it")}$$

This very utterance removes **Q**'s lack of knowledge about the fact p , and thereby makes its own assertion false. This is not just a stupid side-effect of 'saying too much'. E.g., in the paradigmatic example of Muddy Children, it is crucial in the last round that the dirty children announce their ignorance once more, after which that very action removes it, and the dirty children know that they are dirty. 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 logical explanation is that statements may *change their truth value* when an epistemic model gets changed by an update. When worlds remain in the model after an update, their factual properties do not change, but their epistemic properties may. The issue here is *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:

> facts p, and knowledge-free assertions generally, knowledge assertions K_ip, ignorance assertions $\neg K_i$p

A relevant result from general modal logic is the following 'preservation theorem' from [Andréka et al.98]:

Theorem
The epistemic formulas *without common knowledge operators C* that are 'preserved under submodels' in the above sense are precisely those definable using literals p, ¬p, conjunction, disjunction, and K_i–operators.

These reflect the 'universal formulas' in first-order logic, which are just the ones *preserved under submodels*, by the Łos-Tarski theorem. The obvious conjecture for the language with common knowledge is this:

> *the above theorem also holds for the language with common knowledge, now also allowing C-operators*

This result is still open, as preservation theorems for this kind of language are not accessible to standard model-theoretic techniques. In any case, what we need for update analysis is not quite full preservation under submodels, but rather preservation under 'self-defines submodels':

When we *restrict* a model to those of its worlds which satisfy ϕ, then ϕ holds throughout the remaining model: $\phi \rightarrow (\phi)^{\phi}$

Model-theoretic preservation questions of this special form seem new.

6 Describing states and theory update

An update procedure is an algorithm for changing states in some computation device. What guarantees that the algorithm is 'correct'? SO far the pictures themselves are supposed to carry conviction. Another approach comes from computer science. One specifies in words what the algorithme must achieve globally, and then checks that it meets these specifications. This brings us to another general perspective on update. In much of the computational literature, 'information states' are identified with sets of assertions, data bases, or 'theories'. 'Update', in this setting, is addition of assertions, and subsequent closing under deductive consequences. The latter view is often taken to be simply dual to the more model-theoretic one:

eliminating possibilities corresponds to *adding assertions*

As a syntactic theory grows, its set of models shrinks, and vice versa. But in our setting, things are more complicated. The usual duality works for atomic facts, but it is problematic with ignorance: every time you learn something, you also lose something (if only a prejudice). Still in principle there always is a syntactic counterpart to semantic update. The *State Definition Lemma* of Section 3 assures us that modifying bisimulation-closed classes of models is like modifying suitable epistemic formulas.

Question
Find a perspicuous syntactic account of public update as theory modification.

Public update with epistemic assertions is not 'monotonic': one may have to *revise* the current theory, throwing away assertions. This is like Belief Revision Theory [Gärdenfors et al.95], but the exact connection is again open. The 'linguistic' approach of this section is sound in philosophical analyses of speech acts. It also occurs in work on communication like [Meyer et al. 95], which follows an alternative computer science view of 'dynamics', namely transformations of *preconditions* into *postconditions.* For instance, one may say that an answer serves to transform the precondition that *Q does not know if ϕ and holds it possible that A knows if ϕ* into the postcondition that *ϕ is common knowledge between A and Q.* (Of course, readers of Section 5 know that this is not a particularly bright specification because it does not do the right thing for some statements ϕ.) Still, there should be some systematic connection of the following sort, yet to be found:

The earlier semantic updates should be a *minimal model change* leading to the truth of the correctly stated postconditions.

7 Dynamic inference

Standard epistemic logics describe inference in fixed worlds in information models. In our setting, a more lively alternative would follow the dynamics of update more closely:

> Conclusion C *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 C.

That is, an update with the conclusion C leaves any information state 'pre-processed with the premises' unchanged. This dynamic notion of inference behaves rather differently from inference in standard logic:

- *order of presentation matters*
  Conclusions from A, B need not be the same as those from B, A: witness $\neg K$p, p (consistent) versus p, $\neg K$p (inconsistent)
- *multiplicity of occurrence matters*
  $\neg K$p & p has different update effects from ($\neg K$p & p) ; ($\neg K$p & p)
- *adding premises can disturb conclusions*
  $\neg K$p implies $\neg K$p, but $\neg K$p, p does not imply $\neg K$p.

None of these phenomena occur with classical logical reasoning, whose 'structural rules' say precisely that order, multiplicity, and overkill of premise information does not matter. So we ask:

How to describe dynamic inference systematically?

No general solution exists. But there is an analogy. Two variants of classical structural rules that do hold for the above dynamic *update-to-test inference* are given in [van Benthem96], chapter 7:

Left Monotonicity	$X \Rightarrow A$		implies	$B, X \Rightarrow A$	(LM)
Left Cut	$X \Rightarrow A$	$A, Y \Rightarrow B$	imply	$X, Y \Rightarrow B$	(LC)

There is also a 'representation result' for this kind of inference:

Theorem
Dynamic inference is axiomatized completely by LM and LC.

Using public updates, we just refuted all classical structural principles beyond these. Indeed, all abstract counter-examples in the proof given in [van Benthem96] seem available using suitable epistemic models:

Conjecture
Epistemic public update has only LM, LC as its structural rules.

Dynamic inferences may be viewed as a kind of of local reasoning that agents do on the basis of the actual real-time presentation of the information. By contrast, classical logic then describes longer-term reasoning when details of preservation are no longer present. There is no conflict here, but rather one more interesting open question:

How do classical epistemical logic and dynamic inferences cooperate - as they seem to do in human competence?

The following section broadens interfaces between updates and dynamic inference with existing logical systems, and raise more traditional questions.

8 Dynamic logic of update actions

Dynamic inference is the tip of an iceberg. Updates are actions taking us from models **M**,s to new models **N**, t. This suggests a more general view in terms of *dynamic logic* [Harel et al.99], which has expressions for

both *actions* and *propositions.*

Propositions have the usual Boolean operations, while actions may be combined using choice or sequential *composition.* Typically, propositions A can give rise to actions, such as a test (A)?, or more to the point for us, a public update (A)!. Also, one can form assertions $\Box_a\phi$ for any action a

$$\mathbf{M}, s \models \Box_a\phi \quad \text{iff} \quad \mathbf{M}, t \models \Box_a\phi \textit{ for every } t \textit{ such that } a \textit{ changes } s \textit{ into } t$$

A complete axiomatization for this logic is known, including axioms

$$\begin{array}{lcl} \Box_a(\phi \to \psi) & \to & (\Box_a\phi \to \Box_a\psi) \\ \Box_{a;b}\ \phi & \leftrightarrow & \Box_a\Box_b\phi \\ \Box_{a\cup b}\ \phi & \leftrightarrow & \Box_a\phi \ \&\ \Box_b\phi \end{array}$$

plus some axioms for iteration, expressing its smallest fixed-point character. This style of analysis also works nicely for public update [Groeneveld95]. Think of a kind of dynamic 'super-model' whose states are themselves epistemic models **M**,s that get modified via update actions. Here are some relevant axioms, suggested by [Gerbrandy98]:

$$\Box_{A!}(\phi \vee \psi) \quad \to \quad \Box_{A!}\phi \vee \Box_{A!}\psi$$

update is deterministic (at most one output state)

$$\Box_{A!}p \quad \leftrightarrow \quad p$$

update changes no atomic facts in the current state

$$\Box_{A!}K_i\phi \quad \leftrightarrow \quad K_i(A \to \Box_{A!}\phi)$$

One can also analyse the structural rules of dynamic inference in this dynamic logic. A complete system is found in [Moss00].

Remark *Truthful update*
Our notion of update assumes the implication $\Box_{A!}\phi \to A$. This does *not* state that A holds *after* the update: which would be $\Box_{A!}A$ (cf. Section 5). $\Box_{A!}\phi \to A$ says that A holds in the current state *before* the update.

This language is the proper setting for more complex 'epistemic programs' over information states. A nice example is again "Muddy Children". Here the father announces that at least one of the children is muddy, and then asks over and over again whether any children know their status. What we recognize here is a *standard program construction* of iteration:

"WHILE no one knows, REPEAT: 'state whether you know'"

The point is that the procedure terminates, in a state of common knowledge of who is what. The update universe serves as one computational model, where all the usual techniques of dynamic program logics apply. Its logic may also be understood technically as an explicit calculus of *relativization* [van Benthem00], where A-updates relativize assertions to the new universe.

9 More delicate update phenomena

In addition to public update, there are many more subtle forms of communication. Even with questions, there are other genres, say rhetorical, or hostile questions. More generally, there are many 'communicative modes' of language use, and we are rather good at picking up which one we are in. Update logic may help systematize these, and get clear on their differences. Next, there are more complex forms of communication, where not all is 'in the clear'. Examples are private update, or semi-public update. At the opposite extreme lie systematic forms of hiding information and secrecy. Then, there is a border-line between communicative conventions and personal defects, such as memory limitations, cheating, or other individual peculiarities that may affect information flow (cf. [Baltag99, Baltag et al.99]). Communication is not always public, and not all agents may get the same information. Examples abound in *games*, which stipulate different rights and duties for players. Here is one example of how this affects the above model. Consider the following episode in a semi-public game:

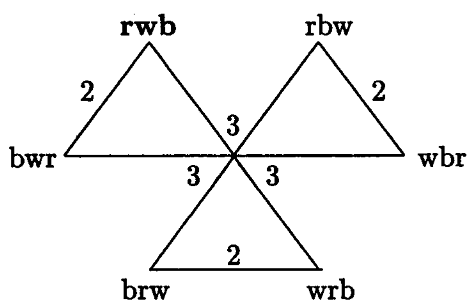
We have both just drawn a closed envelope. It is common knowledge between us that one envelope holds an invitation to a lecture on logic, the other to a wild night out in Amsterdam. Clearly, we are both ignorant of the fate in store for us. Now I open my envelope, and read the contents, without showing 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 this may not even be all. All these 'epistemic overtones' seem relevant to our further interactions. What is the general principle?

This story calls for updates that *eliminate links* instead of states. The initial state is one of collective ignorance, and the update removes my uncertainty link, while both worlds remain to model your factual ignorance:

This kind of update occurs in many *games.* [van Ditmarsch00] treats 'Cluedo' and defines other 'knowledge games' in this style. Returning to the card example in Section 1, suppose that the following happens:

Player 2 shows his card only to player 1

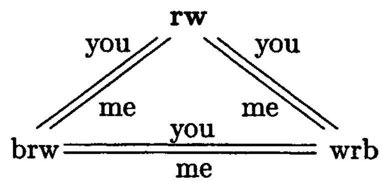
Again, this updates just some uncertainty links to:



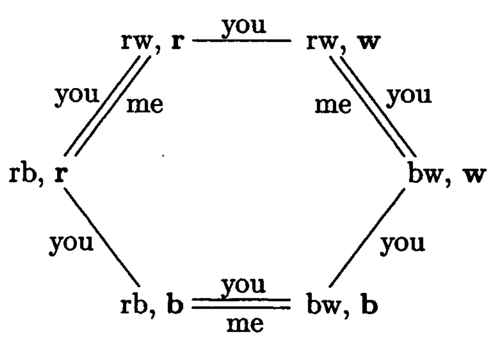
It is easy to vary much more in this simple picture, and get a feel for the various kinds of update that occur when knowledge is passed publicly, privately, wholly, or partially. Also, one can do update logic of 'link elimination' in much the same style as was done for 'world elimination' in Sections 3-8.

World and arrow elimination decrease the size of information models. But more general updates may also increase the size! This happens with communication that is public and explicit within some subgroup, but conveys only minimal information to 'the outsiders', the remaining group of agents.

Example *Cards again*
There are 3 cards on the table, red, white, and blue, but down. I draw 2, keep them face down, and put them in front of me. You see me do this. The initial information state is as follows (with, say, **rw** the true state of affairs):



Now, I pick up one of my cards, and read it, without showing you which one. As a result, I know know of at least one card that it is in my possession, and so do you (be it 'de dicto', while I know it 'de re'). For you, there are two possible card-reading actions of mine, which you cannot distinguish, though I can. The resulting picture duplicates the earlier one:



One easily checks that this diagram yields the intuitively correct outcomes.

A general treatment of this broader kind of update is found in [Groeneveld95, Baltag et al.99, van Ditmarsch00], [Gerbrandy98] (the basic modeling breakthrough) and [Baltag99] (who gave the most elegant system to date). One generates new states from old ones by 'tagging on' representations of the communicative action that has just taken place. This involves enriching standard epistemic logic with an account of *uncertainty between actions*:

(a) Abstract *update events* are defined separately from the usual uncertainty models, in a so-called 'action diagram'.

(b) Not every action need be succesfully executable in every state: e.g. 'telling the truth' requires that the assertion holds.

(c) Agents need not distinguish all update events, a feature which is again coded by uncertainty relations $\sim_a$, *but now between actions.*

(d) Updates of information states are computed as *products*: the new states are pairs (*old state, action applied to it*) , and the new uncertainty relations between these states are now computed as follows:

$$(s, A) \sim_a (s', A') \qquad \text{iff} \qquad s \sim_a s' \ \textit{and} \ A \sim_a A'$$

10 Conclusion

This survey has shown how information update in communication is a rich and rather TARK-ish area, where logic meets linguistics, philosophy, computer science and even experimental cognitive science. There are challenges on two fronts. Descriptively, we still need good update models for cryptographic protocols, and more general forms of hiding information from third parties. Moreover, we need to integrate updates with decision problems and game-theoretic strategic behaviour. On the more technical side, even the simpler 'well-understood' systems are full of open logical questions, once one pursues the update perspective more systematically. Our discussion has not even been complete in this respect, as there are also interesting computational issues, e.g. concerning complexity of single and repeated updates. Finally, some preliminary experience shows that the above is easy to teach - and may provide one more entry into the world of logical analysis.

References

[Andréka et al.98] Andréka, van Benthem, Németi, 1998, 'Modal Logics and Bounded Fragments of Predicate Logic', Journal of Philosophical Logic 27:3, 1998, 217-274.

[Baltag99] A. Baltag, 1999, 'A Logic for Communication', Centre for Mathematics and Computer Science, Amsterdam.

[Baltag et al.99] A. Baltag, L. Moss & Solecki, 1999, 'The Logic of Public Announcements, Common Knowledge, and Private Suspicions', with Alexandru Baltag and Slawomir Solecki. Proceedings TARK 1999, Chicago. Department of mathematics, Indiana University, Bloomington.

[Barwise et al.97] J. Barwise & L. Moss, 1997, Vicious Circles, CSLI Publications, Stanford.

[van Benthem96] J. van Benthem, 1996, Exploring Logical Dynamics, CSLI Publications, Stanford.

[van Benthem00] J. van Benthem, 2000, 'Update as Relativization', Tech report, ILLC Amsterdam.

[Blackburn et al.01] P. Blackburn, M. de Rijke & Y. Venema, 2001, Modal Logic, Cambridge University Press.

[Dekker00] P. Dekker, 2000, 'Optimality Theory and Game Theory: some parallels', manuscript, ILLC Amsterdam.

[van Ditmarsch00] H. van Ditmarsch, 2000, Knowledge Games, Ph.D. thesis, Department of computer science, Groningen, and Institute for Logic, Language and Computation, Amsterdam.

[Fagin et al.95] R. Fagin, J. Halpern, Y. Moses & M. Vardi, 1995, Reasoning about Knowledge, MIT Press, Cambridge (Mass.). P.

[Gärdenfors87] Gärdenfors, 1987, Knowledge in Flux, MIT Press, Cambridge (Mass.).

[Gärdenfors et al.95] P. Gärdenfors & H. Rott, 1995, 'Belief Revision', in Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 4, ed. by D. M. Gabbay, C. J. Hogger, and J. A. Robinson, Oxford University Press, Oxford 1995.

[Gerbrandy98] J. Gerbrandy, 1998, Bisimulations on Planet Kripke, Ph.D. thesis, ILLC Amsterdam.

[Groeneveld95] W. Groeneveld, 1995, Logical Investigations into Dynamic Semantics, Ph.D. thesis, ILLC Amsterdam.

[Harel et al.99] D. Harel, D. Kozen & J. Tiuryn, 1999, Dynamic Logic, MIT Press, Cambridge (Mass.).

[Meyer et al. 95] J-J Meyer & W. van der Hoek, 1995, Epistemic Logic for Computer Science and Artificial Intelligence, Cambridge University Press.

[Moss00] L. Moss, 2000, "Logics of Announcements: Survey and Open Problems", Department of mathematics, Indiana University, Bloomington.

[Veltman96] Veltman, 1996, 'Defaults in Update Semantics', Journal of Philosophical Logic 25, 221-261.