# THE ART OF MODELING

Johan van Benthem, Amsterdam & Stanford, http://staff.science.uva.nl/~johan

*Abstract* 'Possible worlds semantics' for modal logic is a widely used term, sometimes with ominous metaphysical connotations, but what does this style of modeling involve today? We discuss three main issues, using epistemic logic as a running example, and drawing upon both mathematical results and practices in the expertise of working researchers. Our first question is a foundational one: how does one associate a type of model with a language, and what considerations affect that choice? Our focus is on invariance and definability results, familiar from the mathematical and computational tradition, though less so in philosophy. The second question is less deep, but maybe even more challenging in practice: once we have chosen a type of models for a language, how does one select and then maintain models appropriate to concrete scenarios of application? While there is a lot of 'art' to this in the literature, there is very little 'science' of model construction for modal logics. We show how this works in dynamic epistemic logics, and identify some current challenges for a true 'modeling theory' as opposed to the more abstract usual 'model theory'. Finally, we discuss the pervasive tension between 'thin' and 'thick' worlds in modal logic, using examples from game theory, and pointing out how the contrast can be made fruitful.

## 1 Possible worlds and modal logic

As a student I was taught that modal languages are about possible worlds, and that possible worlds are entities into which philosophers have deep insights with the naked mind's eye, not even a mental telescope is needed. Over time, 'possible worlds semantics' has come to stand for a wide variety of models, so much so that the objects inside them can be almost anything: points in time, states of a computer, mental states, outcomes of a game, etcetera. Indeed, Blackburn & van Benthem's 2006 introductory chapter in the *Handbook of Modal Logic* says that modal logic is about *directed graphs* $(W, R, V)$ of points $W$, arrows $R$, and a propositional valuation $V$, a mathematical structure like a grin of the Cheshire Cat which remains when we abstract from any particular feature of models that have been proposed in the tradition. [1] In what follows, I will approach matters at this formal level of generality, and even then there will be lots of interesting things left to say. Moreover, a history of growing abstraction is a well-known success pattern in science. Many theories in science that started with one thing in mind ended up with new ones, unintended by their inventors – and by the end of the road, they have become abstract structures in algebra or geometry.

In logic, there are even two driving forces for such diversity and need for abstraction. These correspond to the two natural directions in 'logical semantics' (cf. van Benthem 1983). One direction starts from some given language and maybe some associated deductive practice, and asks for models that would adequately illuminate, and perhaps validate the practice. This is done in many logical studies of philosophical argumentation, or in the semantics of natural language. But one can also start with an interest in some independently given kind of structure, say Time, develop ideas about its basic structure, and only then think about most appropriate languages for bringing out the important properties, making them subject to inference, or communication to others. Both directions will play a role in what follows, and of course, they can also work together in various phases of the development of a field.

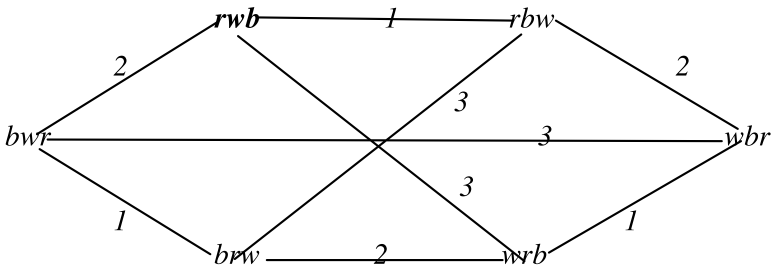
## 2 Modeling practice in epistemic logic

In what follows, we take epistemic logic as a source of examples, without any claims as to how it performs on its grander aims (cf. van Ditmarsch, van der Hoek & Kooi 2007 for a modern introduction to what follows). We all know the set-up. Let $\boldsymbol{M}$ be a possible worlds model representing the current information state of one or more agents. [2] Now, knowledge is a universal modality: an agent $i$ knows at world that $\varphi$ is true (written as $\boldsymbol{M}, s \models K_i\varphi$) if $\varphi$ is true in all worlds $t$ in $\boldsymbol{M}$ that are $R_i$–accessible from $s$. But what are those possible worlds in concrete scenarios? Here is a standard illustration showing how innocent they are:

[1] Indeed. the term 'possible worlds' seems to have lost most of its meaning – just as 'trust' did long ago with financial institutions, so that no one feels a problem with, say, '*anti-trust* legislation'.

[2] Think equivalence relations for the relations $R_i$ if you want, it does not matter for the issues here.

Consider a mini-game: three cards 'red', 'white', 'blue' are given to three players: *1, 2, 3*, one each. Each player can see her own card, but not that of the others. The real distribution over the players 1, 2, 3  is *red, white, blue* (***rwb***). Here is the resulting information state:



This pictures the 6 relevant states of the world (the 'hands', or distributions of the cards), with the appropriate accessibilities pictured by labeled 'uncertainty lines' between hands. The single 1-line between *rwb* and *rbw* indicates that player *1* cannot distinguish these two situations, while *2* and *3* can (they have different cards in them). The diagram says the following, intuitively. Though they are actually in ***rwb*** (as an outside observer might see), no player knows this. Of course, the game itself is a dynamic process yielding further information – and we will return to the natural resulting *changes* in this model shortly.

For the moment, let us observe a few things about this elegant graph. First, there is no algorithm for producing it – but most people would agree that it fits the situation, and most students are quite capable of finding models like these with just a little training. This *art of modeling* is a cognitive abstraction skill that many people have, a serious fact in itself. [3] Also, in this setting, 'possible worlds' lose their doom-laden ring, since even non-logician audiences agree that the hands form the natural logical space here – and the 'actual world' is no big deal either: after the cards have been dealt, the players do find themselves in some particular initial physical situation, don't they? Next, the standard semantics of epistemic logic works in models like this, and allows us to read off immediately in a visual manner

[3] Of course, in more complex situations, it may require much more creativity and visual insight to find 'the right' epistemic model. I wish we would train students more in challenging practices like this, rather than the mind-numbing 'translation exercises' into natural language that give the wrong idea that 'applying' the formalisms is a matter for rote drill, and eventual automatization.

what agents know and do not know. By inspection of the diagram, we see at once that player *1* knows he has the red card in the actual world ***rwb***, but also that the others do not know, and that *1* knows that they do not know. Finally, this graph shows the attraction of models that represent a minimal structure for the concrete informational scenario at hand, while truth-conditional semantics (another high-brow term that turns concrete here) is just a manner of reading off information systematically from a geometrical structure.

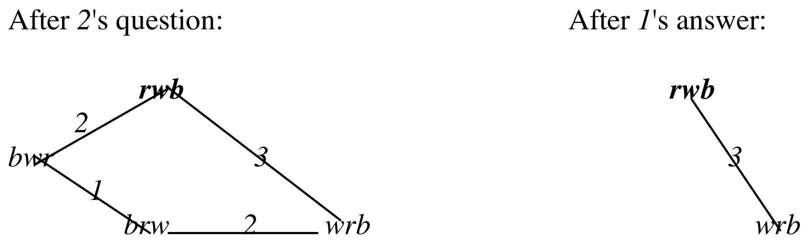
But choosing a model is not just a matter of its reflecting one static situation. The deeper test of a choice is how well it stands up to further things that may happen dynamically as information gets updated. Suppose that the following two conversational moves take place:

> *2 asks 1* *"Do you have the blue card?"*
>
> *1 answers truthfully* *"No".*

Who knows what then? Here is the effect in words:

> Assuming the question is sincere, *2* indicates that she does not know the answer, and so she cannot have the blue card. This tells *1* at once what the deal was. But *3* does not learn, since he already knew that *2* does not have blue. When *1* says she does not have blue, this now tells *2* the deal. *3* still does not know even then.

This may seem simple, but reasoning in this linguistic form can be tricky. We now give the updates in the above diagram, making all these considerations geometrically transparent. Here is a concrete 'update video' of the successive information states:


After *2*'s question:

After *1*'s answer:

We see at once in the final diagram that players *1, 2* know the initial deal now, as they have no uncertainty lines left. But *3* still does not know, given her remaining line, but she does know that *1, 2* know – and in fact, the latter is common knowledge. Similar analyses exist for other conversation scenarios, and for more complex puzzles and games. [4] The matching model constructions are studied in *dynamic epistemic* logic, to which we return below.

## 3 Structure = single worlds plus their interaction with other worlds

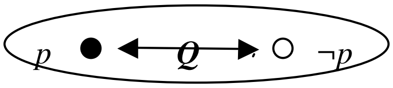
How plausible are the above models in some more intuitive sense? One often hears the following complaint: these 'worlds' do not encode anything about the mental states of the agents, by themselves and interactively with others, and hence they cannot do justice to the real complexities of a social communicative epistemic situation. This negative feeling may be strengthened by the common idea, which I still learnt as a student, that the accessibility relation between worlds is just a technical trick introduced around 1960 to 'weaken the logic' from modal *S5* to, say, *S4* or *K* – so that the above models are just formal devices.

But all this criticism seems to reflect a deep misunderstanding, whatever the intentions of the founding fathers. In the above models, the worlds are indeed very 'light' entities qua internal structure, spanning the ways the relevant part of the physical world might be. But the epistemic information consists crucially in the way these worlds are related through the accessibility relations. Thus, the real story of a world consists in some *minimal internal structure* plus, much more crucially, the sum total of its *interactions with other worlds*. And this approach is a perfectly intelligible, and widely used way of describing structure. An example is category theory in mathematics, where we specify objects largely through the morphisms that connect them to other objects in the category. Or more concretely, in mechanics, an object is described through its interactions with other moving objects.

The power and elegance of this geometrical modeling show when you compare the visual description of a world among its neighbours with its 'modal theory' in the formal language: the complete list of all epistemic statements true at that world. The latter is infinite and may be hard to understand, say because of convoluted iterated knowledge assertions of the form $K_1K_2\ldots\varphi$ that would be intelligible at once as simple $R_1$, $R_2$–cycles in the graph.

[4] In particular, one can also model the scenario where *2*'s question is not informative, with an alternative update video, whose description we leave to the reader.

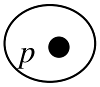
For an example, consider how a simple question answer episode might start (this is just one possible initial situation!). In the following diagram, reflexive arrows are presupposed, but not drawn. Intuitively, agent ***Q*** does not know whether *p*, but ***A*** is fully informed about it:



In the black world, the following are true (*K* for knowledge, *C* for common knowledge):

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

As for information flow, this is a good setting for ***Q*** to ask ***A*** if *p* is true, and the (positive) answer would update this model to one where *p* has become common knowledge:



In view of this semantic power, though there have been determined attempts at doing away with the relational models (cf. the ‘knowledge structures’ of Fagin, Halpern & Vardi 1991), they have survived all onslaughts. Of course, complex infinite possible worlds models may lose the visual advantage, and we revisit the duality with explicit linguistic description below. After all, one expects ‘language’ and ‘structure’ to be in some sort of harmony.

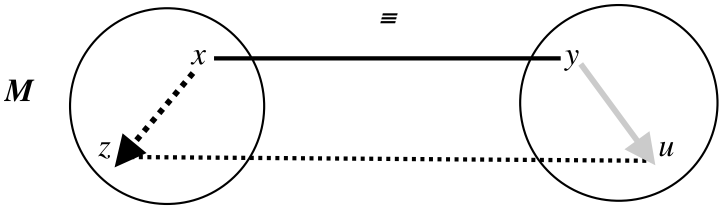
## 4 Structure needs transformations: bisimulation invariance

Our next question is foundational: have we found the right kind of structure in our models? A powerful way of thinking about this issue comes from mathematics and the sciences (van Benthem 2002 has a survey and discussion): well-designed languages reflect stable notions – and stability only shows with an independent account of *transformations* of the relevant system. The *invariants* of these transformations will then be candidates for stable properties that should drive a matching language of inference and communication. Moreover, we can note straightaway that there need not be one single legitimate choice for these notions.

Mathematical theories of Space include geometry, with 'rigid' Euclidean transformations of translation, rotation and reflection, Topology with 'rubbery' homeomorphisms – and other natural accounts of space and shape include the recent 'mathematical morphology'.

How to take this style of thinking with change and invariants to possible worlds models? Here is a notion of semantic invariance that fits modal languages like a Dior gown:

*Definition* (Bisimulation). A *bisimulation* between two models ***M***, ***N*** is a binary relation $\equiv$ between their states $s, t$ such that, whenever $s \equiv t$, then (a) $s, t$ satisfy the same proposition letters ('local harmony'), (b1) if $s\, R\, s'$, then there exists a world $t'$ with $t\, R\, t'$ and $s' \equiv t'$, and (b2) the same 'zigzag' or 'back-and-forth clause' holds in the opposite direction.

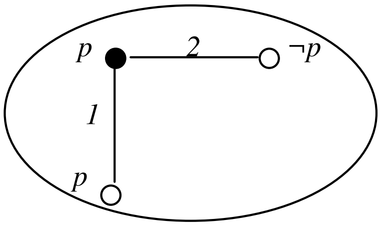


■

Clause (1) expresses 'factual harmony', the zigzag clauses (2) the dynamics of simulation. [5] Bisimulation is one answer to a fundamental semantic question about levels of structure that seldom gets asked in the philosophical literature, though it seems crucial:

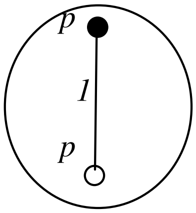
*When are two information structures the same?*

This issue comes up naturally in information update changing a current epistemic model. Suppose the initial model is like this, with the actual world indicated by the black dot:

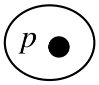


[5] Bisimulation is often given a 'procedural' spin: it identifies processes that run through similar states with similar local choices. Thus, it answers a foundational question: *When are two processes the same?* But there are other natural criteria of identity in process theory (van Benthem 1996).

All three worlds satisfy different epistemic formulas – and in particular, in the actual world, *1* does not know if *2* knows that *p*. But of course, in the actual world, agent *1* does know that *p*, and can say this – ruling out the right-hand world, and updating to the new model



But in this diagram, the two worlds seem redundant, and the following model represents *the same information state* for the agents, as witnessed by an obvious bisimulation:



■

The theory about bisimulation gives ways of contracting models to simplest equivalents, or blowing them up to geometrically perspicuous tree-like structures (Blackburn, de Rijke & Venema 2000). But most important are results tying this structural analysis of models to definability in matching modal languages, providing the dual perspective we are after:

## 5 Defining worlds in words: invariance and definability

Some basic model-theoretic results tie bisimulation closely to truth of modal formulas.

*Invariance Lemma* The following assertions are equivalent for finite graph models:

(a) ***M***, *s* and ***N***, *t* are connected by a bisimulation,

(b) ***M***, *s* and ***N***, *t* satisfy the same epistemic formulas. [6]

For infinite models, matters get more complicated, and sophisticated model theory arises. Still, there are also versions of the preceding result working analogously to *Ehrenfeucht-Fraïssé games* of structure comparison for first-order logic, that work on arbitrary models.

[6] The lemma even holds for arbitrary models, provided we take epistemic formulas from a language with arbitrary *infinite* conjunctions and disjunctions. We forego such technicalities here.

Incidentally, while the Invariance Lemma only needs formulas from the basic epistemic language, the widely used extended epistemic language with common knowledge is still invariant for modal bisimulation. This observation shows that, structurally, such extensions still fall within the same semantic realm as the original language of epistemic logic.

Now, this connection between an epistemic language and bisimulation-invariant structure raises the issue if these are two sides of the same coin. The *modal theory* of a world $w$ in a model $\boldsymbol{M}$ is an explicit record of everything true internally at $w$ about the facts, agents' knowledge of these, and recursively, their knowledge of what others know. The following result (Barwise & Moss 1996) says that states in an epistemic model and maximally consistent epistemic theories are equivalent – where all preceding technical caveats apply:

*Definition Lemma* For each finite model $\boldsymbol{M}, s,$ there exists an epistemic formula $\beta$ (involving common knowledge) such that the following are equivalent:

(a) $\boldsymbol{N}, t \models \beta$

(b) $\boldsymbol{N}, t$ has a bisimulation $\equiv$ with $\boldsymbol{M}, s$ such that $s \equiv t$

Again this result extends to arbitrary models provided we are willing to use formulas from a language allowing arbitrary infinite conjunctions and disjunctions. Instead of a proof, here is an illustration. Consider the two-world model for our earlier basic question-answer episode. Here is an epistemic formula that defines its $\phi$-state up to bisimulation:

$$\phi \;\&\; C_{\{Q, A\}}\,((K_A\phi \vee K_A\neg\phi) \;\&\; \neg K_Q\phi \;\&\; \neg K_Q\neg\phi)$$

*Conclusion* The above results allow us to switch, in principle, between semantic and syntactic accounts of epistemic states. Syntactic approaches have been dominant in belief revision theory, and semantic ones in dynamic epistemic logics, and there is no general rule what suits best in practice. But also in theory, there is no need to view one level as being more fundamental than the other, now that we see their harmony as a desideratum. [7]

[7] Another contact of modal language and world structure occurs in Henkin models in completeness proofs, with maximally consistent sets being worlds. This model is 'universal': cf. Fine 1975. This perspective connects with the bisimulation analysis given here, but it would go too far to explain.

Despite all this pre-established harmony, I end with two disclaimers.

*Caveat 1* I have not claimed that bisimulation is the only natural invariance for modal logic. There can be different natural levels of information structure, and different useful matching epistemic languages. A case in point is the further notion of *distributed group knowledge* (cf. Fagin, Halpern, Moses & Vardi 1995), which is not bisimulation-invariant, and which requires a finer level of individuation for epistemic models.

*Caveat 2* The analogy between worlds and *complete* theories also suggests how a linguistic approach may sometimes be superior to a model-based one. In many epistemic scenarios, we are not interested in specific features, and taking one particular model as we did may be overly exhaustive. Often, we want to understand general generic features of the scenario, so that solutions that we find can be re-used elsewhere. For this purpose, incomplete linguistic descriptions often have an edge – and in line with this, for instance, Sadrzadeh & Cirstea 2006 have proposed algebraic versions of epistemic logic allowing for generic analysis. More can be said here, obviously, but in this paper, we just leave this issue as a tag.

## 6 Using a framework: transforming small models

These were the big issues, whose theory is well-understood. Now for the small issues in possible worlds semantics once we have chosen it, whose theory is much less-understood!

There is a general fact here. Logical frameworks are often judged on a priori grounds, since there is no further criterion of assessment: the formal systems are hardly ever put to use. Of course, we *say* that 'we use' epistemic logic 'to describe reasoning' by rational agents and the like, but who ever does? And the same is true for most other logical systems: the rhetoric does not match the reality. This is different from scientific theories, whose test is both a priori plausibility and deep experience with applications. Philosophers tend to say that Newtonian mechanics is based on the intuitive truth of its axioms – but its real success in physics is the large belt of successful applications built up over the 18th century. [8]

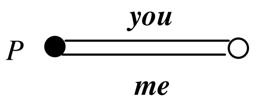
So, here are some practical questions about possible worlds models, again in the epistemic sphere. The issue is not just finding suitable epistemic models for situations on a case-by-case basis, but rather their *connections*: how to transform models as information flows. Then matters quickly become tricky. Here is a well-known example that we all use: *email*. Sending a message to another person is like a public announcement in the group consisting of the two of you, and simple world elimination of the earlier kind will do. Adding a *cc* with other recipients achieves public announcement in that larger group, still by the same mechanism. But what about the email button *bcc*, which we use so much and understand so little? It can mislead the people we communicate with in various ways. Can we say exactly which new epistemic models are created, and how these evolve over time? Information flow with secret communication channels is tricky, and can lead to embarrassing errors. [9]

One area where these issues have come to the fore is *dynamic epistemic logic*. Finding a systematic modeling for *bcc*-like 'private announcement to subgroups' was a long-standing challenge solved in Groeneveld 1995, and especially Gerbrandy 1999, which then led to the framework of Baltag, Moss & Solecki 1998 that is in wide use today, while van Ditmarsch 2000 proposed similar ideas based on parlour games like *Clue*. What follows is a sketch. Van Ditmarsch, van der Hoek & Kooi 2007, Baltag, van Ditmarsch & Moss 2008, and van Benthem 2008 are much more detailed expositions of this fast-growing paradigm.

*Example* (Doubtful signal). Here is a simplest *bcc*-like scenario. Initially, we are both ignorant whether *P* is true. A simplest epistemic model for this will look as follows:

[8] Likewise, in computer science, successful paradigms like Process Algebra have two basic aspects: fundamental perspicuous core ideas, but also, a successful and often surprisingly sophisticated practice of case studies, associated with applying the base notions to particular situations.

[9] Philosophers may find these issues irrelevant 'engineering issues', a far cry from understanding Truth and Knowledge. Personally, I think that intentional 'differential information flow' is so crucial to intelligent behaviour that understanding it better should also be a philosophical concern.



Now you hear a public announcement that *P*, but you are not sure whether I have heard it, or I just thought it was a meaningless noise. In this case, intuitively, we need to keep two things around: one copy of the model where you think that nothing has happened, and one update copy for the information that I received. This requires at least *3* worlds, and hence we need a new update mechanism that can even increase the number of worlds. ■

Now, we can try to write the next model by hand, but the real modeling challenge is finding a *systematic mechanism* that transforms the current model into the new one. This requires a new idea, namely that agents learn from *events* to which they may have different access. For a start, the information models provided by epistemic logic have a natural companion, when we look at the events involved in scenarios of communication or interaction:

*Definition (*Event models). An *epistemic event model* $\boldsymbol{E} = (E, \{\sim_i\}_{i\in G}, \{Pre_e\}_{e\in E}, e)$ has a set of *events E*, uncertainty relations $\sim_I$ for each agent, [10] a map assigning *preconditions* $Pre_e$ to events *e,* stating just when these are executable, and finally, an *actual event e*. [11] ■

Agents' uncertainty relations encode which events they cannot distinguish, reflecting their powers of observation. An event model has no propositional valuation, but events come with *preconditions*. A public announcement *!P* presupposes truth of *P*, my asking a genuine question means I do not know the answer. Most events carry information about when they may occur. That is why they are informative! The following update mechanism describes how new worlds after update are pairs of old worlds with an event that has taken place:

*Definition* (Product Update). For any epistemic model $(\boldsymbol{M}, s)$ and event model $(\boldsymbol{E}, e)$, the *product model* $(\boldsymbol{M} \boldsymbol{x} \boldsymbol{E}, (s, e))$ has domain $\{(s, e) \mid s$ a world in $\boldsymbol{M}$, $e$ an event in $\boldsymbol{E}$, $(\boldsymbol{M}, s)|=$

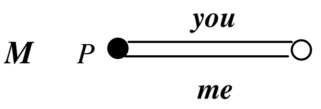
[10] Accessibilities will often be equivalence relations, but we also emphatically include event models with directed minimal accessibility for later scenarios of 'misleading' where knowledge gets lost.

[11] As with 'actual worlds' in epistemic models, we do assume that one event actually takes place.

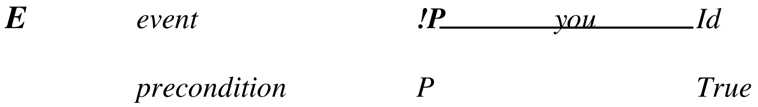
*$PRE_e$*}, while its accessibility relations satisfy the product rule *$(s, e) \sim_i (t, f)$* iff *both* $s \sim_i t$ *and* $e \sim_i f$. *T*he valuation for atoms *p* at *(s, e)* is the same as that at *s* in ***M***. [12] ■

This mechanism models a wide range of phenomena in observation and communication, keeping track of sometimes surprising changes in truth value for epistemic propositions:

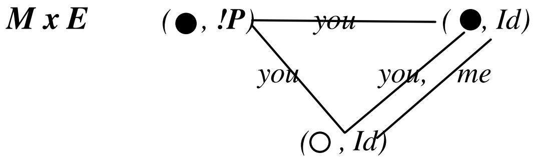
*Example* (Doubtful signal once more). Here is how model size can increase with product update, even in simple scenarios. Consider our earlier epistemic model



Now take an event model for the earlier scenario where I hear *!P*, but perhaps you merely experienced the trivial identity event *Id* which can happen anywhere:



This time, the product model ***M x E*** has *3* worlds, arranged as follows:



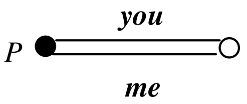
The reader can see how this satisfies our intuitive expectations described earlier. ■

[12] The new model is a Cartesian product of the old ***M*** and the event model ***E***, and this explains the possible growth in size. But some pairs are filtered out by the preconditions, and this elimination makes information flow. Moreover, the key epistemic *product rule* has the following motivation: we cannot distinguish two new worlds if we could not distinguish them before, and the new events cannot distinguish them either. To be more precise, we must still specify the language for the preconditions, which is often just the epistemic language itself: see the literature for extensions. Finally, the intuitive understanding is that the preconditions are common knowledge in the group. Many of these assumptions can be lifted eventually to make the framework much more general.

Current dynamic epistemic logics analyze event models explicitly in their formal language, and they turn out to extend to belief revision or preference change. Beliefs come into play when we allow for events making people mis-informed instead of just under-informed:

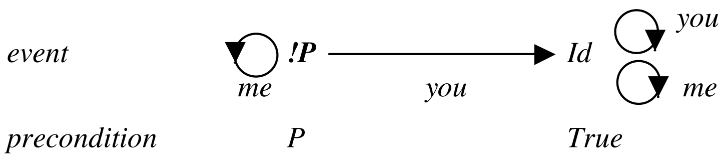
*Example* (Secret peep). We both do not know if *P*, but I secretly look and find out:

***M***



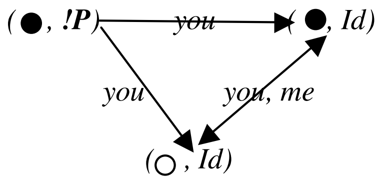
Here is the event model, with the pointed arrows read as just indicated:

***E***



We get an epistemic product model ***M x E*** with *3* worlds, ordered as indicated in the following picture. For convenience of drawing, we leave out a reflexive arrow for me in the actual world, and reflexive arrows for both of us in the other two:

***M x E***



In the actual world, I know exactly what has happened – while you still think, mistakenly, that we are in one of the other two worlds, where everything is just like before. ■

We need not go into details of this framework. I just want to say what all this means to me. Dynamic epistemic logic refines the usual assignment 'by hand' of epistemic models as static snapshots of an information process, via a mechanism for model transformation. Thus, even though there is still the 'art' of finding the initial epistemic model at the start, we get a systematic account of creating new subsequent models, something of importance in practice, as well as in theory. Attention now shifts to a structure that we have around but seldom investigate explicitly: the universe of all possible epistemic models, relations between these, and systematic *transitions* which create new models out of old. [13] Thus, we have the beginnings of a systematic 'modeling theory' behind possible worlds models in epistemic logic – and the same ideas will also work for other areas of modal logic.

I see this as a second general point about possible worlds models, neglected like the first. Our discussion of bisimulation invariance showed that we should take the *internal relations* between worlds inside one modal model very seriously. But now, we also saw that one should take *external relations* seriously, running between whole possible worlds models.

## 7 Coda: some challenges for modeling theory

Dynamic epistemic logic does not do it all: many natural scenarios call for new ideas. My point is there is a coherent area for study here, beyond current scattered observations – teaching which to students would greatly enhance the scope of possible worlds modeling.

Here are three challenges. First, dynamic epistemic logic itself has not yet managed the jump to another crucial area of differential information flow, viz. *security* and hiding information through *coding*. This will probably involve, not just simple modeling devices for scenarios like Internet sessions, but also a better foundational understanding of the relation between two basic notions found in logic: observation-based semantic information and deductive-computational information (cf. the survey van Benthem & Martinez 2008).

Next, many unresolved modeling issues arise with a current focus of research on agency, the formation of *groups* of agents, and what these can be said to know or believe. For instance, what happens when two agents meet, each with their own epistemic model, and these models need to be integrated into a plausible new model of the information for the

[13] Contrast this lively picture with the usual assignment of flat sets of models to theories, without dynamic interrelations. Actually, Barwise & van Benthem 1999 have proposed *'entailment along a relation'* as an account of logical consequence which lets one situation provide information about another. They use this 'model-crossing' setting, amongst other things, to prove new generalized interpolation theorems. But also beyond logic, the same dynamic picture of the universe of models makes sense in other areas, such as the construction of outcome spaces in probabilistic reasoning.

group that has now formed? Note that the agents may have different knowledge, as well as real disagreements in their beliefs, while even the language they use may be different. There has been work on 'belief merge' that is relevant here, with Andréka, Ryan & Schobbens 2002 as a sophisticated mechanism for merging individual preference relations into group relations, provided we get enough information about the epistemic-doxastic dominance structure of the group. One could also look for general product constructions in some category-theoretic sense. But my point is: we do not know yet, and it would be a typical example of a model transformation that would be of wide use once we have it.

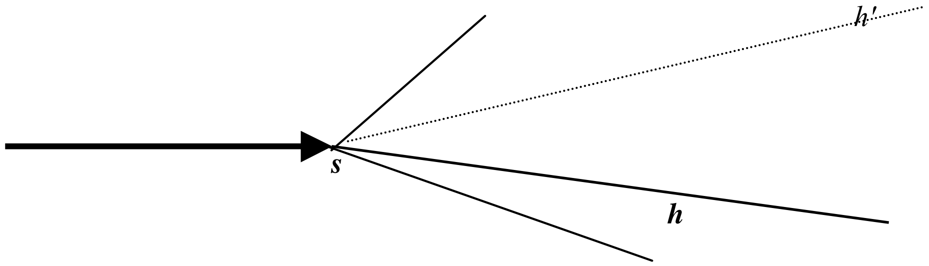
Finally, consider the role of the modal language: if you wish, the 'conceptual framework' used by agents. It is well-known from the philosophy of science that advances in science are often associated with changing the concepts, and thus, the language of our theories. But this is also true for daily practice, where we coin new phrases to resolve disputes, or re-interpret old phrases to restore consistency. Still, there is a general scarcity of 'modeling theory' in terms of language-changing constructions, even though we all know that one of the most frequent responses to, say, new evidence contradicting our beliefs is not rearrangement of beliefs in the old language, but *language change* (Weinberger 1965). [14]

## 8 A 'second opinion': from thin to thick models

Our final theme puts the preceding approach in perspective, by contrast with an alternative. Our perspective so far has been that of small or 'thin models', minimal for the task at hand. On the other hand, there is a tradition of 'thick models' that encode much more of the world. While this terminology is not exactly defined, it is easy to understand it from cases. For instance, in the philosophy of science, when modeling a theory like mechanics, thin models are concrete small mechanical systems of a few objects with mass and force functions, while large models are whole space-time universes with a total evolution for every particle. And in computer science, dynamic logics of programs describe transition relations to next stages of a computational process, while temporal logics describe total

[14] There are a few tricks in dynamic-epistemic logic that allow us to change the language in the course of product update, but so far, they have not been applied to realistic scenarios.

evolutions. In epistemic logics of agency, a prominent thick option are *branching temporal models* giving the complete possible evolutions of some system under study:



This is the Grand Stage view of agency, with histories as total runs of some information-driven process. It can be described by matching modal languages with temporal, epistemic and also doxastic operators that can be evaluated at pairs of histories *h* and stages *s* on them – the natural candidates for the 'worlds' of the models in many uses of this framework. [15]

A setting like this is very different from our earlier thin models. All possible informational events with their effects lie encoded in the branching structure: we do not need to construct any next stage. Of course, this encoding is by fiat, and one just assumes that the thick model has already solved these issues. To demonstrate the contrast, if you are at stage *s*, and an update is to take place with $\varphi$, you look at the sub-tree of continuations, and make an upward move to some closest state *s'* following *s* where $K\varphi$ has become true. Generally, there is a trade-off here. The more information one puts into the model and worlds inside it, the simpler one can keep actions to some next stage. This comes out starkly in game theory. Extensive games have an intuitive 'local dynamics' of what happens to players' knowledge and beliefs from one node to the next, as moves are played. But one can also design thick models of games where all these eventualities are encoded beforehand, including complete strategies for all players – and then playing the game is just a matter of simply observing which moves are played, since the thick world automatically produces all responses.

[15] The Grand Stage view underlies many different paradigms in the current literature, such as Interpreted Systems (Fagin et al. 1995), Epistemic-Temporal Logic (Parikh & Ramanujam 2003), *STIT* (Belnap et al. 2001), or Process Algebra and Game Semantics (Abramsky 2008).

Thin views and thick views represent two temperaments of using possible worlds, and again this raises an issue of 'modeling theory'. How are the two related? Here is a concrete result showing that such questions can have definite answers (van Benthem & Liu 2004):

*Theorem* For epistemic temporal models ***H***, the following two conditions are equivalent:

(a) ***H*** is isomorphic to some branching epistemic tree generated from an initial epistemic model by successive epistemic product updates,

(b) the agents in ***H*** satisfy Perfect Memory and Uniform No Miracles, while the domains of executability of events are bisimulation-closed. [16]

Here the point is not the technicalities, but the general thrust. A result like this shows how transformations on thin models can be retrieved in the thick setting, and we find that the relevant conditions is having epistemic agents on the Grand Stage that are of the right kind. This result can be extended to beliefs over time, with conditions encoding Jeffrey-like revision tendencies on the part of agents (van Benthem & Dégrémont 2008).

Once thin and thick models are brought together in this precise manner, ideas can flow in both directions. Product update suggests explicit logics for constructing temporal models, moving implicit information about tree construction into an explicit formal language. Conversely, dynamic epistemic logic with its local model transformations can borrow an essential idea from the thick models, viz. the global *procedural information* about possible courses of the total process that lies encoded in a temporal tree. Cf. van Benthem, Gerbrandy, Hoshi & Pacuit 2007 on the resulting dynamic epistemic logics of 'protocols'.

The upshot is that possible worlds semantics come seven in two flavours, thin and thick – and that the two can enter into productive relationships. Many further examples of this co-existence may be found in the literature. For instance, van Benthem, Gerbrandy & Kooi 2006 consider intermediate kinds of entities in between thin physical events and thick

[16] Perfect Memory says that agents can only acquire new indistinguishability links based on old world links plus indistinguishable events, while No Miracles says that new distinctions can only arise because of old distinctions already in place, or a distinction between the observed events.

‘intensional’ events, such as “statement *P* as made by an agent of a particular type”:

*Example* (Liars versus Truth Tellers). You meet someone, but you do not know if she is a Liar or a Truth-Teller. You hear her say *P*, which you know to be true. Here is how product update tells you the person is a Truth-Teller. One forms *product events* of the form (agent type, proposition announced), with the following preconditions: $Pre_{\ (Truth\text{-}Teller,\ !P)} = P$, $Pre_{\ (Liar,\ !P)} = \neg P$. Only the first can happen in the actual world, and so you know that you are meeting a Truth-Teller. Of course, in general, more events will qualify, but the point is that we can encode precisely what we know about the agent types in events like this. ■

Thus, there seems no reason to choose between different takes on possible worlds. Indeed, ‘thin’ versus ‘thick’ is even too poor as a description of the natural levels that may occur in modeling one and the same phenomenon. Van Benthem 2008, Chapter 9, distinguishes at least three levels in the analysis of *games*. The thinnest version is to treat *nodes in an extensive game tree themselves* as worlds, which interpret modal languages with actions, preferences, and perhaps uncertainty. This suffices for many purposes, such as defining simple strategies, or defining the Backward Induction solution of a given game as a strategy profile. A less thin option models games as epistemic temporal trees, where the worlds are *complete histories*, or perhaps even pairs of a history plus some finite stage on it. This level can model many further assertions about the course of the game, such as the availability of appropriate responses by players over time. But this level is still too thin to model further phenomena, such as genuine uncertainty about the type of player one is up against. If you know that your opponent either has Perfect Memory, or is a memory-free automaton such as ‘Tit for Tat’, then you need very thick worlds which encode uncertainty between whole strategy profiles, and these are the usual *possible worlds models for games* from the game-theoretic tradition (cf. also Stalnaker 1999). And once could go even further in thickness, to a fourth level with scenarios where the game itself may change.

But our examples will already have made the point. Possible worlds models come in many degrees of thickness, and a true grasp of the phenomena requires understanding and exploiting this phenomenon, as well as the ability to jump between the levels as needed.

## 9 Conclusion

We have discussed three issues having to do with possible worlds models as a general semantic device, with their epistemic interpretation rather their original metaphysical guise as our running example. First, there was the foundational issue what kind of possible worlds structure we want in the first place. We have suggested that the harmony between modal languages and structural invariants should be our guide here, allowing us to see the duality of 'geometrical' structural description of phenomena and syntactic modal theories in formal languages. Next, we have looked at specific modeling issues within the usual possible worlds framework suggesting that there is a systematic 'modeling theory' behind current best practices in the art of logical modeling, which deserves further study, and perhaps even teaching. Finally, we have suggested that such models can live at different levels, with possible worlds of different 'thickness', and that true modeling skills require a view of how these levels are connected. Of course, more can be said here, and we have not exhausted all issues in the art of modeling. But we have provided a richer set of issues surrounding possible worlds semantics than is usually on people's minds. True, many of these further ideas come from mathematics and computation, rather than philosophy proper – but the mixture seems coherent and a worthy broadening of anyone's horizon.

## 10 References

S. Abramsky, 2008, 'Information, Processes and Games', to appear in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier Science Publishers, Amsterdam.

H. Andréka, M. Ryan, & P-Y Schobbens, 2002, 'Operators and Laws for Combining Preference Relations', *Journal of Logic and Computation* 12(1), 13 – 53.

A. Baltag, H. van Ditmarsch & L. Moss, 2008, 'Epistemic Logic and Information Update', to appear in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier Science Publishers, Amsterdam.

A. Baltag, L. Moss & S. Solecki, 1998, 'The Logic of Public Announcements, Common Knowledge and Private Suspicions', *Proceedings TARK 1998*, 43 – 56, Morgan Kaufmann Publishers, Los Altos.

J. Barwise & J. van Benthem, 1999, 'Interpolation, Preservation, and Pebble Games', *Journal of Symbolic Logic* 64, 881–903.

J. Barwise & L. Moss, 1996, *Vicious Circles: On the Mathematics of Non-Wellfounded Phenomena*, CSLI Publications, Stanford.

N. Belnap, M. Perloff & M. Xu, 2001, *Facing the Future*, Oxford University Press, Oxford.

J. van Benthem, 1983, *The Logic of Time*, Reidel/Kluwer, Dordrecht.

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

J. van Benthem, 2001, 'Games in Dynamic Epistemic Logic', *Bulletin of Economic Research* 53:4, 219 – 248.

J. van Benthem, 2002, 'Invariance and Definability: two faces of logical constants', in W. Sieg, R. Sommer, & C. Talcott, eds., *Reflections on the Foundations of Mathematics. Essays in Honor of Sol Feferman*, ASL Lecture Notes in Logic 15, 426–446.

J. van Benthem, 2008, *Logical Dynamics of Information and Interaction*, Manuscript, ILLC Amsterdam.

J. van Benthem & P. Blackburn, 2006, 'Modal Logic, a Semantic Perspective', in J. van Benthem, P. Blackburn & F. Wolter, eds., *Handbook of Modal Logic*, Elsevier, Amsterdam, 1 – 84.

J. van Benthem & C. Dégrémont, 2008, 'Multi-Agent Belief Dynamics: bridges between dynamic doxastic and doxastic temporal logics', Paper presented at *LOFT* Amsterdam & Workshop on Intelligent Interaction *ESSLLI* Hamburg.

J. van Benthem, J. Gerbrandy, T. Hoshi & E. Pacuit, 2007, 'Merging Frameworks for Interaction', *Proceedings TARK 2007*, University of Namur. To appear in the *Journal of Philosophical Logic*.

J. van Benthem, J. Gerbrandy & B. Kooi, 2006, 'Dynamic Update with Probabilities', ILLC Prepublication PP-2006-21, University of Amsterdam. Text of a paper presented at *LOFT*, University Liverpool, 2006.

J. van Benthem & F. Liu, 2004, 'Diversity of Logical Agents in Games', *Philosophia Scientiae 8:2,* 163 – 178.

J. van Benthem & M. Martinez, 2008, 'The Stories of Logic and Information', in P. Adriaans & J. van Benthem, eds., *Handbook of the Philosophy of Information*, Elsevier Science Publishers, Amsterdam.

P. Blackburn, M. de Rijke & Y. Venema, 2000, *Modal Logic*, Cambridge University Press, Cambridge.

H. van Ditmarsch, 2000, *Knowledge Games*, Dissertation, ILLC University of Amsterdam & Informatics, University of Groningen.

H. van Ditmarsch, W. van der Hoek & B. Kooi, 2007, *Dynamic Epistemic Logic*, Springer, Dordrecht.

R. Fagin, J. Halpern & M. Vardi, 1991, 'A Model-Theoretic Analysis of Knowledge', Journal of the ACM 38:2, 382 – 428.

R. Fagin, J. Halpern, Y. Moses & M. Vardi, 1995, *Reasoning about Knowledge*, The MIT Press, Cambridge (Mass.).

K. Fine, 1975, 'Some Connections between Elementary and Modal Logic', in S. Kanger, ed., *Proceedings Third Scandinavian Logic Symposium Uppsala 1973*, North-Holland, Amsterdam, 15 – 31.

J. Gerbrandy, 1999, *Bisimulations on Planet Kripke*, Dissertation, ILLC, University of Amsterdam.

W. Groeneveld, 1995, *Logical Investigations into Dynamic Semantics*, Dissertation, ILLC, University of Amsterdam.

R. Parikh & R. Ramanujam, 2003, 'A Knowledge-Based Semantics of Messages', *Journal of Logic, Language and Information* 12, 453 – 467.

M. Sadrzadeh & C. Cirstea, 2006, 'Relating Algebraic and Coalgebraic Logics of Knowledge and Update', in G. Bonanno & W. van der Hoek, eds., *Proceedings LOFT 2006*, Department of Computer Science, University of Liverpool.

R. Stalnaker, 1999, 'Extensive and Strategic Form: Games and Models for Games', *Research in Economics*, 53:293-291.

O. Weinberger, 1965, *Der Relativisierungsgrundsatz und der Reduktionsgrundsatz – zwei Prinzipien des dialektischen Denkens*, Nakladatelství Ceskoslovenské akademie Ved, Prague.