A NOTE ON MODELING THEORIES

Johan van Benthem

Abstract We discuss formats for formal theories, from sets of models to more complex constructs with an epistemic slant, clarifying the issue of what it means to update a theory. Using properties of verisimilitude as a lead, we also provide some connections between formal calculus of theories in the philosophy of science and modal-epistemic logics. Throughout, we use this case study as a platform for discussing more general connections between logic and general methodology.

1. Logic and methodology of science

The border-line between logic and methodology of science is often hard to detect. Topics straddling it include the proper representation of information and patterns of reasoning. Theo Kuipers' ongoing work on theory structure, confirmation, truth-likeness, and theory change lies at this interface. It uses methods from logic, but also, logicians would be well-advised studying the general issues which it addresses! Instead of pursuing this interface in its proper depth, in this brief Note, I merely address one theme raised by Sjoerd Zwart in his piece for this Volume addressing Kuipers' view of *theories*. What is the appropriate format for formal theories, and for the ways they may change – if all goes well – toward some target theory: the Truth, if you like? Theories are basic building blocks in scientific reasoning, but just as well in the logic of daily practice. This leads to obvious formal analogies, as shown below, but it also reflects a material one. How people reason on 'gala occasions' in the pursuit of science is just an extension – and in some cases even a simplification – of our reasoning about the tasks of domestic daily life. Cognitive rationality is the same throughout.

2. Theories in logic

In logic, theories serve as information structures across a wide range: mathematical theories like Peano Arithmetic, knowledge bases in computer science, or the informal background knowledge and presuppositions that ordinary language users live by. Hence, computational demands of expressiveness and complexity in both scientific and everyday tasks affect the optimal definition of theories, when viewed as the longer- and shorter-term data structures involved in human reasoning.

Accordingly, theory structure in modern logic is a live topic, as new definitions keep appearing – even though this issue of 'logical architecture' is not yet an established agenda item, the way it has been in the philosophy of science. A real chronicle of this development goes beyond this Note, and what follows is just a sketch.

2.1. Theory formats

The basic view of theories in logic is quite austere. Theories T in a logical language are plain sets of syntactical sentences, or alternatively, they can be viewed semantically as the associated set of models

$$MOD(T) \quad = \quad \{ m \mid m \models T\}$$

The two viewpoints are inversely related. The bigger the syntactic theory (more sentences), the smaller the set of models (fewer semantic situations qualify). Either way, there are various types of motivation. A theory may be a stable set of axioms, as in the foundations of mathematics, but it may also be a more ephemeral set of premises in ordinary reasoning. Moreover, in the former setting, some axiomatic theories intend to describe one unique model, say the standard natural numbers, while others are meant to describe a class of mathematical systems, the more the better, as in group theory. In ordinary reasoning, these aspects may play together, witness the common distinction into a long-term 'background theory' constraining the class of situations one deals with, and a short-term 'local theory' of assertions describing one situation out of these, viz. the current topic of conversation.

For some reasoning tasks, this format is too poor, and more fine-structure is needed. The above two viewpoints already differ qua information. The model set approach abstracts away from 'details of syntax' – but in doing so, it loses the *packaging* into the chunks in which the information was presented. Each sentence ϕ in T corresponds to a set of models MOD(ϕ), and it is only their intersection which is recorded in MOD(T). Some modern theory definitions therefore work with

$$MOD^{*}(T) \quad = \quad \{ MOD(\phi) \mid \phi \in T\}$$

E.g., $\{A \vee B, \neg A\}$ is a different packaging for the theory $\{\neg(A\&B), B\}$, though the over-all set of models is the same, being just the valuation with V(B)=1, V(A)=0. Such details are crucial to counterfactual reasoning and theory revision, when seeking maximal sets of chunks consistent with new information that contradicts the theory as a whole (witness authors like Kratzer, Veltman, Segerberg; cf. also Gärdenfors 1987). And even this richer format suppresses information about *presentation*, as logically equivalent formulas defining chunks are identified. But logically equivalent assertions can have very different properties qua ease of understanding, or suggested inferences. In actual argumentation in everyday life, or science, presentation is all-important – but we will disregard such more proof-theoretic perspectives here, despite their relevance to general methodology.

Information chunks may also be *ordered*, making theories consist of layers of less or more important parts, more or less prone to revision. Relative importance has many determinants: such as credibility, ease of computation, or elegance. Thus, preference orderings of components of theories, syntactic or semantic, come into play – as 'entrenchment relations' between assertions in the study of belief revision, or in accounts of data bases as 'ordered theories'. These modern ideas echo earlier ones in the philosophy of science about different importance of assertions involved in hypothetico-deductive explanation: core laws, facts, and auxiliary assumptions. Modern logics of default reasoning use refined entrenchment orderings, showing how these orderings may be updated through incoming information.

This completes our sketch, which is by no means complete. E.g., another major topic in the philosophy of science has been the *vocabulary* of theories, such as the distinction between observational and theoretical terms. Issues of language design and vocabulary update, too, are highly relevant to modern logical theory formats, even though they have been negelected so far – but we let it rest here.

2.2. Information update

A major interest in my own logical work is information update. Here, theories serve as information states about the relevant reality, which get updated as new information comes in. Changes range from straight addition to more drastic revision. I will stick with the former, for comparison with congenial themes in Kuipers' work. Typically, information update decreases the set of options for the actual situation.

> Ann, John, or Mary may or may not have been at the party.
> We want to find out, using the following information:
> (i) John comes if Mary or Ann comes
> (ii) Ann comes if Mary does not come.
> (iii) If Ann comes, then John stays away.

Initially, our information state consists of all possible party compositions

> {MAJ, MA-J, M-AJ, M-A-J, -MAJ, -MA-J, -M-AJ, -M-A-J}

The first assertion $(M \vee A) \to J$ removes three of these, and updates to

{MAJ, M-AJ, -MAJ, -M-AJ, -M-A-J}

The remaining updates are as follows:

$\neg M \to A$	{MAJ, M-AJ, -MAJ}
$A \to \neg J$	{M-AJ}

The final information state has just one situation M-AJ, representing the actual state of affairs with Mary and John present, while Ann is not. Thus, semantically,

information update proceeds by elimination of possibilities.

Dually, information grows syntactically by adding assertions to our current stock.

Now, the main thrust in logics of information update appears to differ from that in the formal philosophy of science. Update logics do not emphasize the statics of knowledge representation, but rather the *many-agent* dynamics of update in communication. Language use revolves around what people know about each other's information. For instance, I will normally try to say something informative, of which I believe that you do not know it yet. Thus, modern update logics will typically handle compound action–knowledge assertions such as the following

$[\,C!]\, K_{\mathbf{i}}\, \phi$ after public announcement of C, agent **i** knows that ϕ

E.g., before you answer my question, I do not know the answer; while afterwards, I do. This makes cognitive actions like announcements and agent-relative knowledge an explicit part of the logical system. Also, knowledge is not the only yardstick. Language use involves a rich repertoire of propositional attitudes toward assertions. Some are known, others believed, suspected, entertained, desired, or hoped for… The dynamics of communication also involves more private communication, gossip, white lies, and many other subtle cues on which we act in our daily lives. These skills may be more subtle than those in science, which is public and 'above board'. This may be the rationale behind Keith Devlin's recent provocative point (cf. http://www-csli.stanford.edu/~devlin/) that someone with the intellectual powers to follow the epistemic complexities of a soap opera on TV should have no problem whatsoever with something as simple as pure mathematics. Even so, it makes sense to compare epistemic perspectives across logic and philosophy of science.

3 Theories in Kuipers' work

Theo Kuipers' formal analysis of scientific theories has ranged widely through his work. (Here and henceforth, we refer to the relevant chapters in the monograph Kuipers 2000, 2001, and occasionally also the forthcoming paper Kuipers 2002. These go back to may earlier papers, not listed separately here.) At one extreme lies his interest in structuralist theories in Sneed's 'structuralist' tradition, with their sophisticated account of families of models, and expansions of vocabularies. In other settings, he sticks with much simpler formats to make his points. A noticeable example is his definition of relative truthlikeness.

3.1. Verisimilitude and flat theories

Here is what it means for theory A to be at least as good a fit for theory T as B:

$$A \geq_T B \quad \text{if} \quad \text{(a)} \quad B \cap T \subseteq A \quad \text{and}$$
$$\text{(b)} \quad A - T \subseteq B$$

Thus, A does at least as well as B on recognizing T-instances, while doing no worse than B on non-T–instances. This Miller-Kuipers definition employs the bare logical view of theories, though with an original take on an important relation between them. One interesting issue here is the nature of T. If we think of approaching the Truth, in line with the above update story, we might expect T to be a singleton set consisting of just *the actual world*. But in Kuipers' work, T is often a larger set of models, standing for the *physically possible situations*. This is a structuralist way of thinking: physical theories describe whole families of possible situations, such as 'mechanical systems' obeying the laws of Newtonian mechanics. This is more like the above case of group theory, with many models for a theory being an advantage, as they stand for broad applicability. A similar point is made in Kuipers' recurring electrical circuit example with light bulbs and switches. Among all possible propositional combinations of atomic assertions 'switch **i** is open', 'bulb **j** is burning', only a certain subset corresponds to physically possible states of the circuit – reflecting the proper dependencies between the propositional variables. Scientific theorizing must then lead us toward this true set of physically possible systems.

From the perspective of the earlier update logics, having a set of worlds, rather than a single world as a target of an elimination sequence is quite reasonable. E.g., when a group of agents pools what they know, they need not expect one single world to emerge, but rather the smallest set of worlds representing their joint information. More generally, instead of zooming in on one object, the update process then becomes a way of learning the extension of a certain predicate, say, the *actual situations*. This is completely in line with many learning tasks, and with the corresponding epistemic logic of knowing which objects satisfy a certain predicate.

3.2. Theories in progress

But there may be more structure. Instead of identifying a theory with just the set of models where it holds, Kuipers also considers a more partial view with a pair of sets

$$R(t), S(t) \qquad \text{where } R(t) \subseteq S(t),$$

with R(t) the models shown to satisfy the theory so far at time t, and S(t) those satisfying all laws recognized so far as following from the theory. One can view this as what we know so far about the True theory, with R the result of experimental verifications (or analyses of empirical systems), while the information encoded in S might be the result of general deductions out of T, or even inductive leaps. More generally, this move replaces any theory T viewed as a set of models by pairs of an 'inner' and an 'outer approximation':

$$T_{down}, T^{up} \qquad \text{with} \quad T_{down} \subseteq T^{up}$$

This two-set view of theories seems the main source of Sjoerd Zwart's misgivings. The reason has to do with update. Suppose that we learn more about a partial theory T_{down}, T^{up}. Intuitively, this should mean that we get a new partial theory T', i.e., a situation T'_{down}, T'^{up} tightening the bounds to

$$T_{down} \subseteq T'_{down} \subseteq T'^{up} \subseteq T^{up}$$

This update from T to T' involves both decrease of a set of models (viz. the upper bound) and increase of another: the lower bound. Therefore, update is not quite a matter of eliminating models by adding more laws to be true: we also eliminate potential laws by adding models. Now, is this a flaw – or worse, an inconsistency? The heterogeneity seems just a feature of the more sophisticated notion of 'theory', though one must take care to use consistent terminology appropriate to the new setting. Indeed, the heterogeneity reflects a natural distinction. Sometimes we learn *necessary conditions* for a certain predicate, say the predicate 'T' defining the right models for our theory. These are constraints of the form

$$\forall x\, (Tx \rightarrow \phi(x))$$

where ϕ is some property of models – a 'law' in Kuipers' sense. This sharpens the upper bound. The counterpart of these are *sufficient conditions* of the form

$$\forall x\, (\psi(x) \rightarrow Tx)$$

Now, the prime examples for this direction were assertions 'model m is an instance of T'. These would rather have the atomic form Tm . But often, one does not verify that a single situation falls under a physical theory, but rather a whole class – say, all pendulums in a certain range. Then the sufficient condition format is more appropriate. Conversely, the law-format of necessary conditions also subsumes the special case of one negative instance, represented by an atomic assertion $\neg Tm$. For the latter is equivalent to the universal formula $\forall x\, (Tx \rightarrow x \neq m)$.

3.3. Theories in partial logic

More radically, we can even reverse the update direction altogether! Here is another view of partial theories. Instead of the lower and upper bound, we follow practice in three-valued logical semantics, when defining a partial predicate T of objects (in this case, 'actuality' of models). For this purpose, one specifies two disjoint sets: the *positive extension* consisting of the known instances and the *negative extension* consisting of the known counter-examples. This is encoded in an ordered pair

$$T^+, T^-$$

Here the two given sets are disjoint, leaving in general a third 'grey zone' of those objects which are neither definitely inside, nor outside of T . Of course, the two representations in 3.3 and 3.2 are equivalent, as we have the identities

$$A^+ = A_{down}, \quad A^- = -A^{up} \qquad \text{with } - \text{ for set-theoretic complement}$$

Update now amounts to increase *in both sets*: adding more positive instances and also adding more negative instances. This is indeed the way we learn the extension of many basic predicates of our ordinary language in daily life.

Partial logic poses interesting challenges when extending classical notions to such a new setting. A well-known case is implication between assertions. Classically, A logically implies B if B is true whenever A is. By contraposition, then also: A is false whenever B is. In partial logic, however, the latter clause does not follow from the former, which only implies that A is not true whenever B is not true. Typically, in a three-valued perspective, 'not true' is not the same as 'false'.

Indeed, if we want both implications to hold, we need so-called 'double-barreled consequence' (cf. the survey of partial logic in Blamey 1986):

(a) whenever A is true, B is true, and
(b) whenever B is false, A is false.

Similarly, it is not entirely clear how to extend the above definition of verisimilitude to partial theories. Such an extension makes sense, e.g., when comparing two partial theories versus the complete true theory, or two complete theories versus the best known partial approximation of the true theory. The question is now how to extend the definition given in 3.1. Our proposal would be to read the earlier implications in the double-barreled sense, making sure that all positive successes for B are also successes for A, while clear blunders for A are also blunders for B:

$A \geq_T B$ if (a) $T^+ \cap B^+ \subseteq A^+$ and $T^+ \cap A^- \subseteq B^-$
(b) $T^- \cap A^+ \subseteq B^+$ and $T^- \cap B^- \subseteq A^-$

In some ways, this is even clearer and more symmetric than what we had before. Moreover, when the partial theories are total (i.e., the given two parts exhaust the whole set of models), this stipulation reduces to the original one. Nevertheless, the three-valued partial logic perspective also allows other definitions of verisimilitude, and the general conclusions below do not depend on this particular choice.

4 Lifting one level

Does the partial view of theories contradict the original elimination view of update? The answer is negative, as we can *lift* the setting.

4.1. From partial models to sets of model with complete denotations

Let M be a model with a domain of objects satisfying a bunch of totally defined predicates **P**, plus some 'unknown' partial predicate T of objects. Now take all standard models M(T) over the same objects as M which leave the interpretation of all predicates in **P** unchanged, while providing complete two-valued interpretations for T extending the definite facts about membership of T already true in M. These models M(T) represent the definite ways the predicate T can turn out to lie on the basis of M. In particular, then, we can associate the pair T_{down} , T^{up} with the set of all those models M(T) whose T lies in between these two bounds. Thus, theories now become sets of sets, each representing a different option for the extension of T that is still open. In particular, an old standard theory T corresponds to a singleton set, as there is no variation possible. Closely related, an information state T in update semantics, which can still shrink arbitrarily, will correspond to the set of all subsets of its current T .

Clearly, the above 'increasing updates' adding new positive or negative instances now become eliminative ones after all, removing those complete extensions for the predicate T either lacking, or having the object mentioned in the update. Increase becomes decrease at the price of one extra level of sets. It just depends on how we want to model things. We can think of this more generally as shifting from updating in terms of objects and their properties to updating in terms of propositions about predicate extensions. Just as earlier, we eliminate all models M(T) that fail to satisfy the relevant assertion. But is this just a trick, or is there more to the new setting? I think there is.

4.2. Updating with arbitrary assertions

The richer setting allows for further updates, not covered by the format of learning a law, or a positive example. Consider a scenario with four people {1, 2, 3, 4}, two of which (2, 3) are women. Some people are criminals, but we do not know which. Now, successively, we learn the following about the gang T involved:

(i) 1 belongs to the gang
(ii) the gang has 2 members
(iii) the gang contains a woman

The first update might just lead to the partial theory {1}, ∅ . But the second one is of a different character: it gives a global property of the gang, without telling us about precise members. It rather puts a constraint on membership patterns. In terms of updates, the original information state has all 16 subsets of {1, 2, 3, 4} as possible extensions of T. The update for assertion (i) reduces this to the eight subsets containing person 1 . The update for assertion (ii) reduces this drastically to

{1, 2}, {1, 3}, {1, 4} ,

even though it does not add to our direct knowledge about individual gang members. Finally, assertion (iii) is again of a global type, reducing this information state to

{1, 2}, {1, 3}.

Speaking logically, these more general updates correspond to further assertions one might make about the predicate T. Thinking of a more definite language, one might think of a monadic predicate logic with predicates for existing structure in the model M plus a unary predicate T for the special objects we are interested in. In particular, updates on partial theories of the earlier special forms are so-called Horn clauses, whereas general assertions may involve *disjunctions*: either these two objects, or these, either this woman or that, etc. That is, assertions might have such forms as

$$(Tm \,\&\, \forall x\, (Tx \rightarrow \phi(x))) \vee (\neg Tn \,\&\, \forall x\, (\psi(x) \rightarrow Tx))$$

In ordinary reasoning, as we just saw, such disjunctive updates are not far-fetched. I think they also occur in science. It is quite possible to find that the models for our scientific theory must be either this class or that, without being able to determine which one. (Think of a theory given by differential equations allowing two types of solution, where the choice must come from further principles, yet unknown.)

4.3. Recovering the old inside the new

The question which does come up then is what makes the original partial theory format special in this broader world. For the information states corresponding to partial theories T_{down}, T^{up}, this is easy to see. We formulate two simple technical results comparing the two levels. Their only purpose here is to show that, instead of hunts for 'one right viewpoint' or inconsistencies, there are interesting *connections* to be observed and proved between different levels of looking at a 'theory'.

Fact A set U of models M(T), or the associated family of T-denotations, can be represented in the format $T_{down} \subseteq T^{up}$ iff the following conditions hold:

(a) The set U is closed under arbitrary unions and intersections

(b) U is convex: any set in between two of its members is in it.

We can think of these sets U as especially 'coherent' information states in the larger setting. Moreover, techniques from update logics (van Benthem 1996) characterize the above special updates. Take the case of adding a law, i.e., a necessary condition.

Fact The updates corresponding to satisfying a law are those maps Φ on families of sets U which satisfy the following three conditions:

(a) Introversion: $\Phi(U) \subseteq U$

(b) Continuity: $\Phi(\bigcup_{i \in I} U_i) = \bigcup_{i \in I} \Phi(U_i)$

(c) Tightness: $\{ X \mid \Phi(\{X\}) = \{X\}\}$ is a set-theoretic ideal.

Further questions of this sort arise concerning verisimilitude. We have some first results on relating the notion $A \geq_C B$ for partial theories to one for the corresponding families of complete models M(T) – but we will leave this tangential issue here.

4.4. Other instances of 'the shift' in logical modeling

The above is not an isolated discussion. Similar shifts in levels of representation occur in other areas. First, consider natural language, which involves many different updates. There is informational update, as we learn more facts – but e.g., there is also enrichment of the relevant *context*, as more individuals are introduced into the current narrative. Major systems of dynamic semantics treat this via a mix of representations. An assertion like "He came in" will enrich the current context with a link between the pronoun "he" and some object d , while at the same time, eliminating pairs <world, context> in which the assertion CAME-IN(d) is false. The literature calls this approach mixing of eliminative and 'constructive' update. The balance between the two will be dictated by considerations of parsimonious representation, and perhaps focus.

The contrast also shows in AI, with non-monotonic versus monotonic reasoning. E.g., in *predicate circumscription*, one takes only models of the premises so far with minimal extensions for the predicates, and valid conclusions are all assertions true in such models. Thus, saying that a and b are P implies that *only* a and b are P – but learning a new fact that c is P may subsequently invalidate this inference. The usual motivation for non-monotonic reasoning thinks of the given premises as *all we know*, using the small set of predicate-minimal models to draw 'eager conclusions'. Now, many critics find this unconvincing, as the additional fact does not contradict the earlier two – while we should only call a conclusion from premises valid if we are not going to be bothered by additional information. Again, this has to do with choosing a larger model set for the premises, encoding in advance what further information we might receive. As in the above discussion of theories, there seems to be no 'truth of the matter' between non-monotonic and monotonic approaches. It is up to us to choose if we want a small model set, supporting many inferences, but jolted by every new update into a drastic revision – or a larger model set allowing for lots of updates, through the smooth elimination of scenarios that will no longer occur.

Finally, the move to sets of sets occurs frequently with 'higher constraints' on possibilities. Suppose you are pondering the future course of a game played against me. For many purposes, this is adequately modelled by the well-known picture of a branching tree of all future worldlines from now on which you consider possible. Each worldline has moves by you and countermoves by me, and the total tree imposes restrictions on this interaction. But now, suppose you believe a 'uniformity' about my strategic behaviour. Either, I always chose the left-most move available to me in the game tree, or always the right-most move. Then one set of future world-lines no longer suffices, as you need to consider two separate future trees, one for each strategy of mine. As in the above, such a move up in sets also suggests a richer logical language with assertions exploiting the additional structure.

5 Logics and calculi

Typical aspects of a logic-style analysis are the use of formal languages, and the calculisation of valid inferences for notions of interest. Now, some key notions in Kuipers' methodology of science have the same flavour. Hence a precise logic connection would be of interest. Here is an example.

5.1. The logic of verisimilitude

Verisimilitude as defined in Section 3.1 satisfies a number of formal properties, making it function a bit like a logical conditional (van Benthem 1987). We list a few:

$A \geq_T B$, $B \geq_T C$	imply	$A \geq_T C$
$A \geq_T B$, $A \geq_{T'} B$	imply	$A \geq_{T \cup T'} B$
$A \geq_T B$, $A \geq_{T'} B$	imply	$A \geq_{T \cap T'} B$
$A \geq_T B$	implies	$A \geq_B T$

Tarski and the Polish logicians proposed a logical 'calculus of theories' in the 1930s – notions like this provide interesting extensions of its repertoire. Now, the above way of looking at theories provide a method for embedding issues of verisimilitude into standard logical systems. This accounts for the above formal inference patterns. Let the working language be monadic first-order logic with unary predicates A for sets, or equivalently, a basic *modal language* with operators [], <>, over an S5–style semantics, with proposition letters a corresponding to sets A.

Fact Reasoning with verisimilitude can be faithfully translated into modal S5.

Proof The above has inferences from some set of premises of the form $A \geq_T B$ to some such conclusion. More generally, these are Boolean inferences over atoms stating a relation of relative truthlikeness. Clearly, it suffices to translate these atoms into a modal setting. But this is easy, as an assertion $A \geq_T B$ interpreted as in 3.1 can be translated as a modal statement of the form

$$[]\ (((t\&b) \to a)\ \&\ ((\neg t\&a) \to b))) \qquad \blacksquare$$

As a consequence of this logical translation, the logic of verisimilitude is completely axiomatizable, and decidable at the same low complexity as S5, taking non-deterministic polynomial time (NP).

The translation even extends to cover truthlikeness of partial theories. For this purpose, a well-known trick suffices. For each partial theory A^+, A^-, take a pair of proposition letters a^+, a^-, and add a modal premise $[]\ \neg\ (a^+\&\ a^-)$. On top of this, reasonable definitions for verisimilitude between partial theories (like the one suggested above) will translate straightforwardly into the modal language.

5.2 Epistemic logic

From a logical point of view, there are really two languages involved in the above. The modal language with the universal box serves as a classification device: it provides a description of one particular model M(T), explaining the available sorts of objects (worlds, if you wish). But there is also the larger setting of a family of such models, which determines what we know about the extension of various predicates. For that we need an *epistemic language* with operators

$$K_{\mathbf{i}}\ \phi$$

stating that ϕ is true in all **i**–accessible members M(T) of the family, as seen from the actual model. A typical epistemic formula might be

$$K_{\mathbf{i}}\,[](a \to p)\ \&\ \neg K_{\mathbf{i}} <>(q\ \&\ a),$$

expressing that agent **i** knows that all A-situations satisfy law P, while she does not know that A-situations can co-occur with the property Q. And of course, the K–language will also allow for 'social' knowledge about other people's knowledge. In this semantics, over an information state for a partial theory, the objects in A^+ are precisely those for which people know that they belong to A, and those in A^- the ones they know not to belong to A. Moreover, one can show that the epistemic model supports no further 'hidden knowledge' on constraints governing the eventual filling of the grey zone. Thus, a partial theory A^+, A^- corresponds to an epistemic model concerning the real form of A which validates exactly the right assertions. (Cf. Fagin et al. 1995, and Van der Hoek & Meijer 1995 on the state of the art in modern epistemic logic, plus connections with partial logic.)

But we can also make finer distinctions now, such as one between $A \geq_T B$ and the stronger $K_i A \geq_T B$. In addition, knowledge assertions may be specified for individuals, or even for common knowledge $C_G \phi$ in groups G. Thus, we plug in to the existing, and fast-growing body of research on epistemic logic. This includes dynamic logics for information update, telling us, e.g., how relations of truthlikeness can change as new information comes in, through assertions of the form

$[P!]\, C_G\, \phi$ after a public announcement of P (say, in some journal publication), community G has common knowledge of ϕ

Even more expressively than this, one might consider the size of the total domain of objects as subject to ignorance. In that case, one might add dynamic operators referring to addition or removal of objects from the current domain – as studied by Theo's Groningen colleague Gerard Renardel in his recent dynamic logics with 'creation and destruction' of objects in the current domain.

As stated before, this setting imports special concerns from dynamic-epistemic logic with procedural details of communication and other forms of information passing, and that between different agents forming different groups. This may be too fine-grained for much methodology of science, but it might be an interesting aspect of method in scientific communities as well – especially, if one could show that their communicative conventions differ systematically from those found in daily life. E.g., many puzzles of actual communication involve learning useful factual things from knowledge of *ignorance* by others – with card games as a clear example. Is there any counterpart to this in the methodology of science, unless our aim is to describe the mechanics of competition?

6 Logic and methodology of science

This Note has taken rather long to make some simple points about representation of theories, update, and knowledge. Nevertheless, it may also serve more broadly to show how methodology of science and logic meet in significant themes. The above story is only a beginning. One might take ideas across in either direction. E.g., the methodologists' idea of one theory being a better approximation of the truth than another makes good sense in epistemic logic, too – as do even more general 'quality comparisons' between model-theoretic constructs vis–à–vis some given standard. Likewise, scientific practice has speech acts like 'proposing' or 'entertaining' a theory which might enrich the repertoire of communicative actions in logical semantics. Vice versa, powerful research trends in modern logic such as belief revision, or many-agent perspectives, seem of immediate relevance to understanding the workings of scientific methodology.

At the editors' request, I conclude with a few general reflections on logic and the methodology of science. In the years around 1980, I did some work at this interface – with van Benthem 1982 as a fair example. My expectation then was that the model theory of formal theories would blossom into a flourishing new field in between the philosophy of science and mathematical logic. These hopes have not really been fulfilled. The two research communities seem rather disjoint, despite occasional journal issues or conference sessions where practitioners are thrown together in hopes of sudden love and understanding. In particular, logicians have looked elsewhere for affection, taking AI, computer science, and linguistics as sources of inspiration for flourishing new interface communities. Despite all this, and even in this wider modern landscape, I still feel that logic and methodology of science share many common causes.

First, qua subject matter, I see hardly any difference at all. Pioneering 19th century logicians like Bolzano, Mill, or Peirce, were also innovative methodologists of reasoning in the broadest sense. The fact that general methodology moved out of logic textbooks is largely a historical accident of the agenda-contracting turn toward the foundations of mathematics in the early 20th century. But a blend of logic and general methodology still persists with influential later authors like Lewis, Hintikka, van Fraassen, or Gabbay – to name just a few. It would be tedious to draw a sharp boundary in their work between logic proper and general methodological concerns. But also less anecdotically, it is a matter of simple observation in the literature that key topics in general methodology such as confirmation, explanation, theory structure, or causality, also occur in logic, AI, or linguistics. A good example of a happy marriage of this sort is the theory of belief revision in Gärdenfors 1987, mixing sources from both logic and the philosophy of science in the most natural fashion. A similar blend may be found in the formal account of events and causal reasoning in Shoham 1985. More recent examples are modern studies in the logic of time and space (surveyed in van Benthem 1995), or investigations of abduction and other forms of non-monotonic inference (Flach & Kakas 2000, as well as Aliseda-Llera's contribution to this volume), where logic runs into philosophy of science, computer science, and general argumentation theory. And of course, as said before, Theo Kuipers' own work provides lots of pointers for fruitful confluence.

Moreover, these concerns are shared for a reason. As stated at the start of this Note, I believe in the continuity of human cognition, from daily life to more exalted intellectual discipline, when practicing, law, theology, or science. This does not mean that no differences exist at all between science and common sense – but the specialized intellectual activities of science are probably best understood as special 'parameter settings' of one general cognitive functioning. As a student, I was made to memorize Ernest Nagel's list of supposedly crucial differences between 'science' and 'common sense' at the beginning of his classic book *The Structure of Science*. Nowadays, I find that all points mentioned there are matters of degree – and that one cannot even understand the success of science without seeing it as an extrapolation of human common sense behaviour. Instead of arguing for this view here, let me just stick out my neck for it. I would be happy to defend the position that virtually every issue in understanding science has its 'domestic' counterparts in common sense cognition. Of course, philosophers of science have much to say about larger aggregation levels of knowledge, and longer-term historical phenomena across generations of scientists, which do not normally surface in logic textbooks. But eventually, even these broad-band perspectives seem equally crucial to the functioning of language and common sense reasoning in daily life.

I do not pretend that I forecast all the above commonality around 1980. To the contrary, I find many things I wrote in those days remarkably un-visionary. Let me list a few predictive failures. First, the more fruitful logic interface for methodology has not been model theory and mathematical logic, but rather philosophical logic and logic systems of a more computational slant. Also, the computational turn which has shaken up research agendas in both logic and philosophy of science is conspicuously absent. And finally, I now find much of what I wrote about philosophers 'just stating definitions' patronizing and misleading. There may indeed be a historical division of labour between mainstream logicians and philosophers of science, in that the former take definitions of formal systems and proving technical theorems as their standard, while the latter look for interesting formal models explicating phenomena in the (empirical) sciences. But difficult issues of modeling and concept explication are also crucial to logic itself – and indeed, some of the breakthrough achievements in logical semantics of natural language, or recent dynamic logics of information update, have been precisely of the latter sort. Delicate modeling skills may even be a rarer talent among students than hard-hitting theorem-proving powers.

It would take a much longer story to substantiate the above claims. But even with the broad outline I have painted, one clear message of my piece is this. Despite the distance between our communities, logicians should find trips to the work of Theo Kuipers and his colleagues in the philosophy of science highly rewarding!

References

van Benthem, J. (1982). The Logical Study of Science. *Synthese*, 51, 431-472.

van Benthem, J. (1987). Verisimilitude and Conditionals. In Th. Kuipers, ed., *What is Closer-to-the-Truth ?*, Amsterdam: Rodopi. Pp. 103-128.

van Benthem, J. (1995). Temporal Logic. In D. Gabbay, C. Hoggar, J. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming*, Volume 4, Oxford: Oxford University Press. Pp. 241-350.

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

van Benthem, J. (2001). Language, Logic, and Communication. In *Logic in Action*, Amsterdam: ILLC Spinoza project, ILLC.

Blamey, S.1(986). Partial Logic. In D. Gabbay, F. Guenthner, eds., *Handbook of Philosophical Logic*, Vol. III. Dordrecht: Kluwer. Pp. 1-70.

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

Flach, P., A. Kakas, A., eds. (2000). *Abduction and Induction, their Relation and Integration*. Dordrecht: Kluwer.

PGärdenfors, P. (1987). *Knowledge in Flux*. Cambridge (Mass.): MIT Press.

Cambridge: Cambridge University Press.

Kuipers, Th. (2000). *From Instrumentalism to Constructive Realism*. Dordrecht: Kluwer.

Kuipers, Th. (2001). *Structures in Science. Heuristic patterns based on cognitive structures*. Synthese Library 301. Dordrecht: Kluwer.

Kuipers, Th. (2002). Inductive Aspects of Confirmation, Information, and Content.

To appear in the Schilpp-volume *The Philosophy of Jaakko Hintikka*.

Shoham, Y. (1985). *Reasoning about Change: Time and Causation from the Standpoint of AI.* Cambridge (Mass.): MIT Press.

University of Amsterdam & Stanford University

Institute for Logic, Language and Computation

Plantage Muidergracht 24, 1018 TV AMSTERDAM, The Netherlands

johan@science.uva.nl, http://staff.science.uva.nl/~johan/