Logique & Analyse **133-134** (1991), 5-14

REFLECTIONS ON EPISTEMIC LOGIC

Johan VAN BENTHEM

Traditionally, 'epistemic logic' has been a specialized area of general modal logic devoted to the epistemic modalities and their interaction with quantification, predication and identity. In this brief Note, we shall step back and see which broader forms of epistemic semantics exist, and discuss how they affect traditional concerns of the field.

1. *Standard Epistemic Logic*

Traditional epistemic logic started largely with the pioneering classic Hintikka 1962, which presented modal logics for the notions of a person knowing and believing a proposition, based upon standard possible worlds semantics with an accessibility relation of 'epistemic indistinguishability'. Roughly speaking, I know a proposition if that proposition holds in all worlds which I cannot distinguish from my current one, that is, if it holds across my whole 'range of uncertainty'. Different epistemic attitudes may then vary in the extent of that semantic range, or in the requirements imposed upon it. In a subsequent series of papers, Hintikka also developed an epistemic predicate logic, dealing with issues of individual identity across epistemic worlds and the interplay of knowledge and ('de re' versus 'de dicto') quantification. No extensive mathematical theory developed around this initial locus — the technical monograph Lenzen 1980 is a respectable exception — but the paradigm did find a number of active areas of application in the eighties, especially in computer science with the work of Joe Halpern and his colleagues at IBM San José, and in Artificial Intelligence by Bob Moore and some colleagues at SRI, Menlo Park. What the computational connection added, in particular, was a more concrete view of epistemic models derived from computational protocols or processes, where agents are well-defined and philosophical questions concerning the intentionality of human cognition can be safely side-stepped. Notions which then came to the fore were 'autoepistemics', concerning the equilibrium states that an agent can achieve in non-monotonic default reasoning (Moore 1987), and also various forms of collective epistemics expressed in the interplay of operators $K_a B$

('agent a knows that B') for various agents a, as well as 'common knowledge' for groups of agents (Halpern & Moses 1990). Moreover, these epistemic concerns were embedded in a larger environment of communication and physical action. As a result, there is now a flourishing community around the so-called TARK Conferences (cf. Halpern, ed., 1986, Vardi, ed., 1988 and subsequent volumes), which brings together philosophers, mathematicians, computer scientists, economists and linguists. In what follows, we shall take the basic core of epistemic logic for granted, while making only occasional reference to its computationally inspired extensions.

2. *Implicit versus Explicit Epistemics*

The above enterprise of epistemic logic has not been without its critics. In fact, many people have regarded it as a typical form of 'shallow analysis', describing epistemics by merely imposing some modal superstructure on top of an ordinary classical semantics, rather than reanalyzing the latter in depth. To introduce a distinction, the Hintikka-Halpern paradigm is one of 'extrinsic epistemics', which does not affect standard classical semantics, witness its key truth definition stating that, in any model, "K_aB is true at a world w iff B is true at all worlds v that are R_a-related to w". By contrast, a more radical 'intrinsic epistemics' would not take classical semantics for granted, and reanalyze the whole notion of truth in the light of epistemic considerations. The most prominent historical instance of such a radical enterprise is of course intuitionistic logic, where 'truth' recedes as the central logical notion in favour of 'provability' or 'assertability'. Thus, its guiding philosophy has been epistemically oriented from the start. (Compare the constructive theories of meaning based upon the 'Brouwer-Heyting-Kolmogoroff interpretation' that are surveyed in Troelstra & van Dalen 1988.) To be sure, intuitionistic logic, too, has its Kripke-style possible worlds semantics, but with an entirely different flavour. Worlds stand for information states, accessibility encodes possible informational growth, and truth at a world corresponds intuitively to epistemic 'forcing' by the available evidence there. Thus, the basic logical operations themselves become 'epistemically loaded'. But then, given an epistemic reinterpretation of what logic and semantics is about, where is the need for any separate 'epistemic logic'?

It is not so easy to adjudicate this debate. Certainly, the intuitionistic approach has generated more interesting mathematical theory, while also

being more influential in philosophical and computational circles. But there are many different causes for this course of events, not all of them having to do with the relative merits of implicit versus explicit epistemics. Moreover, as a matter of fact, there are also some advantages to the extrinsic approach. For instance, it can incorporate whatever sound analysis is provided in a classical semantics for a certain kind of propositions, without having to worry about philosophical compatibilité d'humeurs. By contrast, more intuitionistic proof-theoretic approaches have had difficulties making sense of logical operators such as generalized quantifiers that have perfectly straightforward model-theoretic explications. (But cf. van Lambalgen 1991 for a fresh start.) Moreover, the very explicitness of Hintikka's approach has encouraged an active search for new epistemic operators, such as the above ones referring to knowledge of multiple agents or groups which have no counterpart in intuitionistic logic. Even though mathematics is a social activity, too, where insights may depend essentially on cooperation of rational agents, the implicit intuitionist stance has not been very conducive so far to bringing this out formally.

Finally, the distinction between intrinsic and extrinsic epistemics is an 'intensional' one, which may be undercut at a more formal level of analysis. For instance, the well-known Gödel translation embeds intuitionistic logic faithfully into S4, the logic of choice for much of epistemic logic, thereby making it a kind of 'forward persistent' part of the latter approach. (Van Benthem 1990 provides further formal detail here. Incidentally, no converse embedding seems to work.) In this line, the explicit 'epistemic mathematics' of Shapiro 1985 may also be viewed as a natural extension of intrinsic intuitionistic mathematics. Conversely, the analyses of 'common knowledge' put forward in Barwise 1988 seem more 'intrinsic' analyses of this phenomenon, changing the underlying classical model theory to a form of information-oriented situation semantics. Generally speaking, then, there seems no problem of principle in combining both the agenda and the technical apparatus of intrinsic and extrinsic approaches to epistemic logic.

3. *Information-Based Semantics*

In recent years, the 'intrinsic' approach has gained ground in various new guises. Current semantics for natural languages and computation shows a trend away from truth conditions and correspondence with the world outside to explaining propositions in terms of their role in information processing

over models that are now viewed as 'information structures'. Here, the classical turn-stile becomes a notion of 'forcing' for statements by the available information. Examples of this trend are data semantics (Veltman 1985), various forms of partial modal logic (Thijsse 1992), constructive semantics in the style of Nelson (Jaspars 1993) or substructural semantics for relevant or categorial logics (van Benthem 1991, Wansing 1993). This development can be seen as epistemic logic in a broader sense, since much of cognition can in fact be subsumed under the heading of information processing. What this new phase adds, however, is an explicit concern with the nature of information states and possible updates over these effected by successive propositions. Much is still unclear at this stage, witness the persistent debate between 'eliminative' accounts of information processing (which proceed via successive elimination of epistemic possibilities) and 'constructive' ones (which build ever larger representation structures). Either way, these concerns do seem a natural addition to the foundations of epistemic logic.

Moreover, one clear trend can be observed which does not depend on the exact nature of epistemic states. Current information-oriented modellings suggest a richer and more systematic design of important epistemic operators. This phenomenon is illustrated by the basic case of Kripke models themselves. Initially, these were designed to model a particular epistemic language, whose operators were given independently. But then, one can also reverse the perspective and ask, given such models for information structure, which epistemic languages would best bring out their semantic potential, thereby redesigning the original language. For instance, as in temporal logic, there are two natural directions in the growth ordering of information states, both backward and forward. Knowledge may be mostly concerned with epistemic 'advance', but it has to do also with epistemic 'retreat', in contraction or revision of our information (cf. the symmetric theory of updates and contractions in Gärdenfors 1988). Thus, a more adequate epistemic logic should also incorporate operators reflecting these additional directions and their interplay, with 'forward knowledge' referring upward to all possible extensions of the current state, and 'backward knowledge' referring to the epistemic past. Their interplay will then generate different routes for epistemic revision ("if A had been found, then..."). Van Benthem 1990 explores the resulting hierarchy of operators over information models, using a framework of enriched modal logics. In particular, a next natural stage of epistemic expressiveness would involve operators reflecting the addition of information pieces (i.e., suprema in the ordering of possible

growth) as well as their downward counterparts (i.e., their infima). For instance, the binary epistemic modality $\phi+\psi$ would hold at those states which are the sum of a state verifying ϕ and a state verifying ψ. These will allow us, e.g., to refer to logical features of composite knowledge arising from the combination of various sources.

Of course, by this time, one will have left the traditional areas of the Theory of Knowledge which inspired Hintikka's initial enterprise. But that might also be considered a virtue, and one could certainly translate many of the newer technicalities back into genuine philosophical issues that might revitalize the somewhat fossilized agenda found in most philosophical textbooks.

4. *Adding Justifications*

The above type of enrichment may be seen as an extension of the modal viewpoint on epistemics, with a reinterpretation of its models in a Kripke-style intuitionistic spirit. But there is more to be learnt from a confrontation of extrinsic and intrinsic approaches. One conspicuous feature of contemporary intuitionistic logic and mathematics has been the development of type theories whose proof format rests essentially on binary assertions of the form $\pi : A$, meaning that function π is of type A, or that proof π establishes proposition A. Proof rules in the usual constructivist interpretation then typically build up compound conclusions in both components, witness cases like

$$\frac{\pi_1 : A \qquad \pi_2 : B}{(\pi_1, \pi_2) : A\&B} \qquad\qquad \frac{\pi_1 : A \to B \qquad \pi_2 : A}{\pi_1(\pi_2) : B}$$

Note how both assertions and their justifications are affected here. This binary logical format is gaining popularity these days for its greater perspicuity in bringing out combinations of linguistic propositions plus their underlying manifestations or justifications, where the latter need not always be explicitly linguistically encoded. This makes sense in mathematics and computation (cf. Barendregt 1993), but also more broadly in linguistics and Artificial Intelligence, witness the new research program of 'labeled deductive systems' put forward in Gabbay 1993.

Now here too, there is a very attractive move for epistemic logic. Much of the classical theory of knowledge and its initial logical formalizations seems hampered by the absence of any systematic way of bringing out the

justifications underlying our knowledge as first-class citizens. Put somewhat formally, saying that someone knows a proposition is an existential quantification stating that she has a justification for that proposition. But by keeping those justifications hidden in our logical framework, we create both technical and conceptual difficulties. For instance, much of Hintikka's own work on the Kantian notion of analyticity (cf. Hintikka 1973) has to wrestle with the fact that with Kant, 'analyticity' is a qualification of reasons or justifications as much as of statements, which makes its projection into standard logical systems somewhat problematic. An even more striking example is the so-called 'Problem of Omniscience', where knowledge of a proposition entails knowledge of all its logical consequences. This problem is highlighted in the standard epistemic Distribution Axiom $K(A \rightarrow B) \rightarrow (KA \rightarrow KB)$. By contrast, in the above binary format, this problem would not arise in the first place, because any logical inference will come with an explicit cost record in terms of a more complex justification. To see this, compare the above distribution principle with the corresponding binary type-theoretic inference from the two premises $\pi_1 : A \rightarrow B$ and $\pi_2 : A$ to the conclusion $\pi_1(\pi_2) : B$. Of course, one further question here is what systematic type-theoretic calculus will support explicit epistemic operators. There are various options to this effect, such as the following two introduction rules for epistemic operators:

$$\frac{\pi : A}{- : KA} \quad (\text{'forgetful'}) \qquad\qquad \frac{\pi : A}{*(\pi) : KA} \quad (\text{'reminiscent'})$$

For a more elaborate system of this kind, cf. the modal type theories in Borghuis 1993. (E.g., deriving the above epistemic distribution axiom will also involve suitable elimination rules for the K-operator.) The conceptual task remains to set up a complete plausible base theory for epistemic logic with an explicit calculus of justifications.

5. *Cognitive Action*

When model-theoretic semantics is reinterpreted as a theory of information structures, one further move becomes quite natural. Justifications and revisions are really examples of cognitive actions — and it would be quite appropriate then to embed our epistemic logic into an explicit dynamic logic. This move is in fact foreshadowed in the earlier-mentioned computational tradition in epistemic logic (cf. Moore 1984). For instance, the above binary

schema π : A has one further useful interpretation, stating that "program or action π achieves an effect described by proposition A". Examples of relevant cognitive actions are the earlier-mentioned 'updates' or 'contractions', but one can also think of a much richer repertoire of 'testing', 'querying', etcetera. Moreover, on top of this basic repertoire, one can describe complex cognitive actions or plans via the usual programming operations, such as sequential or parallel composition and choice. Evidently, human cognitive plans have compound structures not unlike those found in computer programs (cf. Morreau 1992). The precise extent of this analogy is an interesting issue by itself. For instance, can cognitive plans also display more infinitary structures like recursion?

Stated in this way, we need a two-level system for combining epistemic statements with 'cognitive programs'. One possible logical architecture here is the 'dynamic modal logic' of van Benthem 1993, which has a relational repertoire of actions over information models interacting with a standard modal logic. Typical assertions in such a formalism will be a dynamic modality $[\pi]$ A, stating that "action π always achieves effect A", (compare the above binary schema), or modal iterations like $[\pi_1]<\pi_2>$ A stating that "action π_1 'enables' action π_2 to achieve effect A". The model theory and proof theory of this system are well-understood (cf. Harel 1984, de Rijke 1992). Moreover, its expressive power subsumes at least the better-known theory of belief revision in Gärdenfors 1988. But dynamic modal logic also supports more radical deviations from classical logic. For instance, van Benthem 1993 considers 'dynamic styles of inference' from premises $\pi_1,\ldots,$ π_n to conclusions C based on the idea that propositions are cognitive actions which are being processed in reasoning. Various options to this effect may be expressed in the above terms. For instance, one plausible dynamic style would state that 'sequential processing of the premises is a way of getting the conclusion', which may be expressed by the modal formula $[\pi_1]\ldots[\pi_n]$ C. Another typical dynamic style would rather state that 'processing the premises is a way of doing the conclusion' — which involves a stronger modal apparatus than that of standard dynamic logic (cf. Kanazawa 1993).

6. *Combinations and Conclusions*

The above Sections point at various attractive enrichments of standard epistemic logic. Putting all of these ideas together, however, raises some obvious further issues. For instance, which logical architecture would most

naturally combine statements, justifications and actions? A type-theoretic approach looks promising here. For instance, one of its key statements might be of the form $\pi : A \to B$, expressing that action π will always lead from states satisfying precondition A to states satisfying postcondition B. In such a format, one might describe the behaviour of compound cognitive actions or plans, mentioned in the above, with rules like the following:

$$\frac{\pi_1 : A \to B \qquad \pi_2 : B \to C}{\pi_1 ; \pi_2 : A \to C} \qquad \frac{\pi_1 : A \to B \qquad \pi_2 : A \to B}{\pi_1 \cup \pi_2 : A \to B}$$

Nevertheless, even this kind of generalization will still fail to capture some further essentials of cognition. First, we need logical systems that lift all of the above to many-person settings, which allow, amongst others, for updating common knowledge of groups via various acts of communication (cf. Jaspars 1993). But also, we would eventually need a theory of 'synchronization' between internal cognitive actions and external physical actions which change the world.

Even with all these open ends, our conclusions concerning epistemic logic will be clear. We would recommend expansion beyond Hintikka's original modal statement format, thereby effecting a junction with a much larger logical environment of epistemic import. Moreover, we think the effort would be well worth-while to systematically rethink much of traditional philosophical epistemology in this broader light.

University of Amsterdam

ACKNOWLEDGEMENT

This paper grew out of notes for two lectures. I wish to thank the organizers of TARK IV (Asilomar 1992), and especially Yoram Moses, for inviting me to their meeting. Moreover, I am grateful to Professor Paul Gochet for inviting me to his workshop on Epistemic Logic (Liege 1993).

REFERENCES

S. Abramsky, D. Gabbay & T. Maibaum, eds., 1993, *Handbook of Logic in Computer Science*, Oxford University Press, to appear.

H. Barendregt, 1993, 'Lambda Calculi With Types', to appear in S. Abramsky et al., eds.

J. Barwise, 1988, 'Three Views of Common Knowledge', in M. Vardi, ed., 365-379.

J. van Benthem, 1990, 'Modal Logic as a Theory of Information', CSLI Report 90-144, Center for the Study of Language and Information, Stanford University.

J. van Benthem, 1991, *Language in Action. Categories, Lambdas and Dynamic Logic*, North-Holland, Amsterdam, (Studies in Logic 130).

J. van Benthem, 1992, 'Abstract Proof Theory', manuscript, Institute for Logic, Language and Computation, University of Amsterdam.

J. van Benthem, 1993, 'Logic and the Flow of Information', in D. Prawitz et al., eds.

T. Borghuis, 1993, 'Interpreting Modal Natural Deduction in Type Theory', in M. de Rijke, ed., 67-102.

J. van der Does & J. van Eyck, eds., 1991, *Generalized Quantifier Theory and Applications*, Dutch Network for Language, Logic and Information, Amsterdam, (to appear with CSLI Lecture Notes, Chicago University Press).

D. Gabbay, 1993, *Labeled Deductive Systems*, Department of Computing, Imperial College, London, (to appear with Oxford University Press).

D. Gabbay & F. Guenthner, eds., 1984, *Handbook of Philosophical Logic, volume II*, Reidel, Dordrecht.

P. Gärdenfors, 1988, *Knowledge in Flux. Modeling the Dynamics of Epistemic States*, The MIT Press, Cambridge (Mass.).

M. Ginzburg, ed., 1987, *Readings on Non-Monotonic Reasoning*, Morgan Kaufmann Publishers, Los Altos.

J. Halpern, ed., 1986, *Proceedings First Conference on Theoretical Aspects of Reasoning About Knowledge*, Morgan Kaufmann Publishers, Los Altos.

J. Halpern & Y. Moses, 1990, 'Knowledge and Common Knowledge in a Distributed Environment', *Journal of the Association for Computing Machinery* 37, 549-587.

D. Harel, 1984, 'Dynamic Logic', in D. Gabbay & F. Guenthner, eds., 497-604.

J. Hintikka, 1962, *Knowledge and Belief*, Cornell University Press, Ithaca (New York).

J. Hintikka, 1973, *Logic, Language Games and Information*, Clarendon Press, Oxford.

W. van der Hoek, 1992, *Modalities for Reasoning About Knowledge and Quantities*, dissertation, Department of Computer Science, Free University, Amsterdam.

J. Jaspars, 1993, *Logics for Communicative Action*, dissertation, Institute

for Language and Knowledge Technology, University of Brabant, Tilburg.

M. Kanazawa, 1993, 'Completeness and Decidability of the Mixed Style of Inference with Composition', Center for the Study of Language and Information, Stanford University.

M. van Lambalgen, 1991, 'Natural Deduction for Generalized Quantifiers', in J. van der Does & J. van Eyck, eds., 143-154.

W. Lenzen, 1980, *Glauben, Wissen und Wahrscheinlichkeit*, Springer, Vienna.

R. Moore, 1984, 'A Formal Theory of Knowledge and Action', CSLI Report 31, Center for the Study of Language and Information, Stanford University.

R. Moore, 1987, 'Possible Worlds Semantics for Auto-Epistemic Logic', in M. Ginzburg, ed., 137-142.

M. Morreau, 1992, *Conditionals in Philosophy and Artificial Intelligence*, dissertation, Department of Philosophy, University of Amsterdam.

D. Prawitz, B. Skyrms & D. Westerståhl, eds., 1993, *Proceedings 9th International Congress of Logic, Methodology and Philosophy of Science. Uppsala 1991*, North-Holland, Amsterdam.

M. de Rijke, 1992, 'A System of Dynamic Modal Logic', Report LP-92-08, Institute for Logic, Language and Computation, University of Amsterdam.

M. de Rijke, 1993, *Extending Modal Logic*, dissertation, Institute for Logic, Language and Computation, University of Amsterdam.

M. de Rijke, ed., 1993, *Diamonds and Defaults*, Kluwer, Dordrecht.

S. Shapiro, ed., 1985, *Intensional Mathematics*, North-Holland, Amsterdam.

E. Thijsse, 1992, *Partial Logic and Knowledge Representation*, dissertation, Institute for Language and Knowledge Technology, University of Brabant, Tilburg.

A. Troelstra & D. van Dalen, 1988, *Constructive Mathematics*, North-Holland, Amsterdam, (Studies in Logic 121, 123).

M. Vardi, ed., 1988, *Proceedings Second Conference on Theoretical Aspects of Reasoning About Knowledge*, Morgan Kaufmann Publishers, Los Altos.

F. Veltman, 1985, *Logics for Conditionals*, dissertation, Department of Philosophy, University of Amsterdam.

H. Wansing, 1993, *The Logic of Information Structures*, Springer, Berlin, (Lecture Notes in Artificial Intelligence 681).