4

Johan van Benthem

Professor of Logic

University of Amsterdam & Stanford University

---

1. Why were you initially drawn to epistemic logic?

Initially, I was not 'drawn' to epistemic logic at all, but repelled by it. Like many of my generation of students around 1970, I considered epistemic logic an industrial variation on basic modal logic, the thing one can do for any sort of intensional operator in natural language, without any special redeeming interest. In particular, it was always a bit unclear whether epistemic logic had anything deep to say about the philosophical notion of knowledge, its professed motivation. And there was no technical compensation either to offset this lack of philosophical impact. No interesting mathematical questions arose from taking the epistemic focus: significantly, the first wave of technical research in the foundations of modal logic in the 1970s, where I was an eager young participant, bypassed it completely. This critical attitude is still common with some colleagues, especially when they go off the record. But for me, the negatives have eroded steadily, for reasons that I will explain, and by now epistemic logic has become a major interest of mine – one of those quiet but intense passions that gentlemen of advancing years are still capable of.

Let me first mention some external influences driving my change of heart and mind. The first was the TARK initiative around 1980, when it became clear that the multi-agent character of epistemic logic was a good fit with the then upcoming post-Turing challenge of understanding distributed computation by many processors, where the sparse abstraction level of a modal operator language seemed just right for bringing to light basic reasoning patterns about information and communication. Such surprising turns are of course frequent in intellectual history: a bad model for one phenomenon can turn out to be a very good model for another,

often in ways completely unintended by the originators. In particular, multi-agent aspects were absent from the original epistemological motivation, while they are now crucial to the success of epistemic logic. Indeed, in the hands of Joe Halpern and his colleagues, TARK has made epistemic logic and computation in a broad sense into a viable and respectable endeavour, vital and still expanding. This program was enhanced still further by the emergence of common knowledge as a fundamental notion, which started a new strand, the logical study of knowledge of groups as collective entities. Thus epistemic logic acquired operators far beyond the basic modal language, linking it to dynamic logics of actions, and generating new mathematical questions after all, some of them unresolved until today. As an aside, this same impulse that brought me closer to epistemic logic was the point of farewell for many philosophers, who felt that the subject was falling in with bad company (the greasy hands of programmers and engineers), straying far from what should be and remain its true agenda of Knowledge in the purest sense. This fateful parting of the ways is still with us, but more on that below.

Epistemic logic became even more noticeable to me in the mid 1980s, with the emergence of 'auto-epistemic logic' that dealt with deep issues in nonmonotonic reasoning, and the reflective equilibria underpinning it, as analyzed by Bob Stalnaker, someone who of course has also lifted epistemic logic to greater philosophical heights generally in his work through the years. Around the same time, Bob Moore wrote a famous essay on knowledge and action showing how combinations of epistemic and dynamic logic generated interesting perspicuous base logics of information-driven action, where one could pose fundamental questions about knowing what plans do, while getting the right information to achieve one's goals. This, too, was a combination of philosophical ideas with computational ones, be it in Artificial Intelligence, a buffer state between philosophy and computer science where philosophers did venture at the time. At least, they did in the early years of Stanford's Center for the Study of Language and Information, when disciplinary barriers went down for a while, and it seemed that the intellectual world was being redefined for good.

The third and final influence that made epistemic logic a reality in my thinking was when logicians like me started learning about the epistemic foundations of Game Theory, pioneered by Robert Aumann in the 1970s. I came to see that epistemic logic, even if limited in itself, created exciting creative mixtures when added to

other fields. That is also how I view the role of logic generally, as a catalyst for broader intellectual developments – or a spice, if you prefer cuisine. Moreover, this game-theoretic work changed my view of the position of logic in the university. In addition to the traditional allies of philosophy and mathematics, and in the second half of the 20^{th} century, linguistics and computer science, logic had now also reached out successfully to economics and the social sciences, a new alignment to me.

But respect is not yet love. Epistemic logic became vitally important to me personally only when I started developing my own 'Logical Dynamics' program around 1990, as a study of the informational dynamics that drives human agency – and it has become increasingly central to it over the years. But let me tell that story a bit later.

2. What example(s) from your work, or work of others, illustrates the relevance of epistemic logic?

First of all, I do not think of epistemic logic as primarily a study of the philosophical, or even the psychological notion of knowledge. I would still submit that it has important things to say about knowledge, but that focus is not the best way of viewing what it really achieves. For a long time now, I have come to see epistemic logic as a theory of *semantic information*, in the sense of Carnap and Bar-Hillel, and indeed in the sense of most of the exact sciences that deal with semantic models. Of course, the models used are extremely simple (more on that below), but that is precisely why they allow us to see essential structures, and develop major themes in a perspicuous manner. To be sure, traditional epistemic logic is just a starting point here. Information comes in many kinds, from 'hard' to 'soft', and logic should, and can help us chart its basic varieties.

But epistemic logic only comes into its own when we make a further decisive step that was totally absent from the original development of the subject. This step, and a major interest of mine over the past decade is *information dynamics*, the explicit logical study of actions, events, and long-term processes that change information states of agents. This 'Dynamic Turn' has been my major interest over the last 20 years, witness my books *Language in Action* from 1991 (Elsevier Amsterdam & MIT Press Cambridge Mass.) and *Exploring Logical Dynamics* of 1996 (CSLI Publications Stanford). These books put the actions at centre stage that

drive the logical and linguistic behaviour of single agents, such as acts of inference, interpretation, and assertion. This move transcends the usual boundaries of logic, since on this view, an act of inference has no privileged position as a way of acquiring knowledge: an observation may do just as well (direct perception may even be more reliable than a mathematical proof, witness the role of experiment in the physical sciences), and so does communication from a trusted source. Thus, Logical Dynamics takes a broad view of sources of knowledge: the logic serves to describe them, but logical inference is not the only acceptable producer of knowledge.

And even beyond this, by now, I have come to see the heart of rational agency as lying in conversation, argumentation, and other forms of interaction between several agents, and the version of epistemic logic that really powers this is not the traditional zoo of epistemic logics with their 'capitalist' names (S4, S5, KD45). It is rather the much richer framework of *dynamic epistemic logic*, a simple yet powerful paradigm for studying information flow by observation and communication, that has emerged over the past 15 years. This is the shape in which epistemic logic becomes broadly applicable to a wide range of phenomena, throwing new light beyond the usual static analysis of knowledge and other cognitive attitudes. Dynamic epistemic logics define informational events, allowing us to talk explicitly about what creates and modifies knowledge, and they have brought to light valid axioms of this information flow that are every bit as intriguing as the traditional epistemic axioms of closure or introspection – perhaps even more so. Even this is still at the heart of epistemology, though, and one of my main philosophical claims in the past ('Epistemic Logic and Epistemology: the State of their Affairs', *Philosophical Studies* 128, 2006, 49–76) has been that we can only understand the robustness of knowledge in the setting of the natural dynamic processes that deal with it. Or in my favourite computer science terms, 'no representation without transformation'.

But knowledge and information flow cannot be the whole story of rational cognition, and they may not even be the most important part of it. To me, the most striking aspect of rationality is not the statics of 'correctness' or being right, but the dynamics of *correction*, i.e., the quality of what we do when we revise beliefs that have been refuted. Knowledge may be the pure gold of the philosopher, but belief is the currency that really makes the whole cognitive enterprise run. It is beliefs that guide our actions, from taking the train to choosing avenues for new math-

ematical research. And the highest quality of rationality is not following some standard that automatically takes us from knowledge to knowledge, but rather the ability to jump to new beliefs, and correct these as we keep learning. Thus, Hintikka was right in studying knowledge and belief together, but their interplay becomes even more crucial when we consider the dynamics: forming beliefs and revising them is the art of living dangerously, a much greater cognitive skill than slavish correctness. Fortunately, it has just been found that the dynamics of *belief revision* and learning through self-correction can be studied very well in the dynamic-epistemic style, making them wholly comparable to the dynamics of information flow ('Dynamic Logic of Belief Revision', *Journal of Applied Non-Classical Logics* 17:2, 2007, 129–155). Incidentally, doing this right also involves a much richer view of information, ranging from 'hard' to 'soft' varieties, and corresponding kinds of information-driven revision actions. (Incidentally, I wonder what would have happened in the study of belief revision in the 1980s, now a separate industry, if this sort of dynamic logical apparatus had been around, rather than staid static 'doxastic logic'.)

My views on the importance of correction versus correctness are becoming quite radical. In the same light, I would now consider the traditional foundational search for consistency proofs of mathematical theories as misguided, since an essence of mathematical activity is the ability to construct new theories when old ones have gone bad. It should really be a relief to us all that Hilbert's Program failed for 'solving the foundational questions once and for all'. This is not to say that revision of mathematical theories is a well-understood phenomenon in my framework, and indeed, the dynamic-epistemic paradigm is still struggling with finding the best model for dealing with pure acts of reasoning.

My new monograph *Logical Dynamics of Information and Interaction* (Cambridge University Press, 2010) develops all this further into a logical study of information-driven agency, showing how epistemic logic in the dynamic sense is alive and thriving. I mention a few more themes from that book that represent major steps in extending the agenda of epistemic logic. One is incorporating more syntactic viewpoints on *fine-grained sorts of information*, and associated with this, acts of inference or introspection that make us explicitly aware of knowledge, beliefs, or even just of issues. A truly dynamic epistemic logic can deal with both syntactic and semantic information, and with acts that manipulate these – overcoming the traditional dichotomy between semantic

and proof-theoretic paradigms for studying knowledge. In particular, one more crucial action comes to centre stage then: *questions* that raise issues directing future inquiry. It has been claimed that the whole history of science could be written more informatively as one of successive questions, rather than answers. Even without going that far, a true dynamic epistemic logic should be able to deal with questions as well as answers. Another essential aspect in my current view is the role of *long-term informational processes* such as learning procedures, or information-driven strategic behaviour in games. The initial dynamic step from static instantaneous knowledge to single events that change information was not enough. We also need an account of the longer term, as learning theorists have long seen – and again, modern epistemic logic can deliver such connections. I have even claimed in recent work ('The Information in Intuitionistic Logic', *Synthese* 167:2, 2009, 251–270) that this generates a further notion of information that is sui generis and crucial to an epistemology that cares for a richer view of agents: 'procedural information' that determines what single information update steps 'mean', making sense of them in a longer history of events.

There may even be one more traditional border that cannot be maintained in all this. We all know that true communication and even the scientific search for truth is driven by *values*. Everything we say and do has a colouring of evaluation, and tied up with this, preferences and goals for action. While this mixture is often seen as a sort of Platonic impurity, I would rather see it as essential to truly rational behaviour. The rational mind is in balance between information and evaluation. Purely informational interaction fails if we do not build a sort of group resonance where engaging in a shared activity is a value per se. Can we keep this evaluative aspect distinct from the informational one? I am beginning to doubt it, and indeed, it is of interest to see that techniques developed for dynamic epistemic logics are now crossing over into the study of preference, preference change, and deontic logics of changing obligations and norms ('For Better of for Worse: Dynamic Logics of Preference', in T. Grüne-Yanoff & S-O Hansson, eds., 2009, *Preference Change*, Springer, Dordrecht, 57 – 84). Again, these studies involve a delicate interplay of semantic and syntactic features, since evaluation may depend on both content and presentation.

Some critics may find this expansion of epistemology toward goal-driven agency a case of unprincipled disregard for established

distinctions in philosophy. I myself would rather say that the program of knowledge in action outlined here represents a typical contribution that (dynamic) logic has to offer to philosophy, bringing to light common patterns in what are often considered disjoint fields of study.

The dynamic study of agency is not confined to philosophy, however. It applies equally well to (and draws on) the thriving areas of epistemic analysis of *agent systems* in computer science (the modern face of intelligent computation), and the epistemic-doxastic foundations of *game theory*. These examples also amply demonstrate what epistemic logic has to offer in its modern guise. Few areas of philosophical logic have had a similar impact.

3. What is the proper role of epistemic logic in relation to other disciplines, for instance mainstream epistemology, game theory, computer sciences or linguistics?

As I have said in earlier books in this series, I dislike this question, because 'proper role' smacks of unwarranted essentialism to me. But as I have noticed with successful Dutch colleagues in radio interviews, one should ignore the questions one dislikes, and just say what one wanted to say anyway when given the slightest chance. So, to me, epistemic logic is a general *logical information theory*, including both static information structure, the dynamic processes transforming this, and the changing attitudes of agents involved in all this. In particular, I think that the notion of information cannot be usefully analyzed without a matching logic of informational acts and processes, and hence what should be confronted with other disciplines is dynamic epistemic logic of one sort or another. Moreover, in this dynamic perspective, epistemic logic feeds naturally into a general account of information-driven *rational agency*, both in the common sense and in science. Thus, the border line with formal epistemology becomes fluid, but so does that with the philosophy of science.

Pursued in this new way, epistemic logic addresses common themes across philosophy, computer science, game theory, and many other fields, including linguistics. Indeed, logic plays two roles here. One is as a messenger: a medium of communication allowing for flow of ideas between fields in jointly understood terms. The other is as a perspective that can help revitalize existing discussions. Take the vexed issue of the definition of 'knowledge' in the philosophers' sense. As I have said, I myself think that its

surplus over true belief is dynamic, having to do with robustness across many dynamic activities where information flows, by single agents, but especially also, in social settings of communication – the arena of Plato's *Dialogues* where the definition of knowledge first became an issue.

But is this merely 'interdisciplinarity' in the sense of aimless unpredictable wanderings? I have argued elsewhere that intellectual history shows continuities that only look discontinuous when one imposes the grid of standard disciplines ('Logic in Philosophy', in Dale Jacquette, ed., 2006, *Handbook of the Philosophy of Logic*, Elsevier, Amsterdam, 65–99). Epistemic logic itself is an example. Viewed in one way, it is a curiously segmented movie with disjoint parts in philosophy, computer science, and game theory. But if one thinks in terms of natural development of themes, a very different story emerges, of a continuous trend toward exploiting the potential of the semantic study of information in a multi-agent setting. I see it as one of the great missed opportunities in recent decades that philosophers have not followed these and other congenial developments beyond their borders, depriving themselves of the many inspiring conceptual insights that have come up in computer science, game theory, and other fields. Instead I often see defensive reactions, trying to explain why all this sort of work is 'shallow', 'irrelevant' and so on, at the same time when parts of analytical philosophy have reached heights of incrowd Scholasticism and elegant obscurity that not even the much-maligned Continental philosophers ever managed to achieve.

Against this background, some positive initiatives shine all the more. I mentioned the TARK conferences where congenial practitioners of many disciplines have met since the early 1980s, and one could cite more such events, such as the LOFT conferences, the recent Network on Rationality and Decision, or the new LORI conferences on Logic and Rational Interaction that have started in China. My impression is that a younger generation of logicians, epistemologists, and philosophers of science is forming these days that thinks much less in terms of traditional divisions, and that is open to external influences from other disciplines.

But now back to the role of epistemic logic in all this. Let me conclude with a caveat. One can have my positive view of epistemic logic as a public benefit between the disciplines mentioned in this question without claiming exclusive rights. One need not even claim exclusive rights as a flag-bearer for logic. For instance, one topic that intrigues me more and more is the dual existence

of two major approaches to information in logical theory and its applications. Epistemic logic is an *explicit account* of information, adding new operators for knowledge to classical languages that are taken for granted. Likewise, dynamic epistemic logic is a conservative extension of classical logics with explicit operators for informational activities. By contrast, say, intuitionistic logic is an *implicit account* of information, changing the semantics of former languages so as to incorporate informational aspects, and thereby generating 'non-standard logics' with different sets of validities. This contrast is all-pervasive. For instance, Amsterdam-style 'dynamic semantics' changes the meaning of natural language semantics to incorporate actions that change information, issues, or what have you. By contrast, an explicit approach would keep the old truth-conditional meanings as before, but add a superstructure of dynamic logic for pragmatic acts of language use. The two souls can live within one breast. Even Hintikka himself goes the implicit way when developing his 'game-theoretic semantics' for expressions in natural language, where information about moves is crucial but remains implicit – whereas I have proposed an explicit epistemic game logic version in terms of what players know, using Hintikka's old ideas to clarify his new ideas ('The Epistemic Logic of IF Games', in R. Auxier & L. Hahn, eds., 2006 *The Philosophy of Jaakko Hintikka*, Schilpp Series, Open Court Publishers, Chicago, 481–513). To some, the implicit approach is deeper, since it seems to affect the very foundations of logic, sending unpredictable complex ripples through the whole discipline, whereas the explicit approach is just one of steady expansion of coverage. To me, both are natural stances, and their relationship a major issue in the philosophy of logic.

4. Which topics and/or contributions should have had more attention in late 20th century epistemic logic?

Question 4 is just the emotional version of Question 5. I will not deign it with a direct answer. But on the analogy of the preceding question, let me state another view of mine that seems relevant. It is a commonplace to say that epistemic logic leaves out various important features of real knowledge, cognition, and agency. While many people seem to be impressed by such criticisms, I find them thoroughly facile and predictable. Pointing out that Reality is more complex than some proposed formal model is easy, anyone can do it. Finding extreme simplifications that still leave us with

important insights is hard, and only open to the most creative minds. Think of a conference where you are listening to a lecture. If you do not have the time or inclination to really think about what is being said, you can always raise a question of coverage and omitted aspects at the end: no deep thought required, and it never fails to impress the audience. But asking a truly perceptive question, on the other hand, requires real thought, and real immersion in the system that has been proposed. My sympathies are squarely with the small system builders and real engagement with what they do. Only when we understand the fruitful simplifications, I would go for disciplined extension toward the broader functioning of information and cognition, the way I have explained in the above.

And also, please recall my earlier point about the virtues of poverty. Simplified models may lead to surprising new applications and research directions, far beyond their initial territory and the intentions of their originators.

5. What are the most important open problems in epistemic logic and what are the prospects for progress?

I mentioned quite a few issues already in my answers so far. Dynamic epistemic logic is a research program in full swing that will generate new questions for quite a while to come. For instance, much remains to be understood about the logical interplay of various sorts of information (hard versus soft, semantic versus syntactic), and the many attitudes beyond knowledge and belief that agents can have in tune with these. There is every reason to assume that dynamic epistemic logic needs a much richer theory of agency than we have right now. Also, I feel that we are only beginning with the serious logical study of how information-driven knowledge works intertwined with revision, self-correction, and *learning* – and perhaps, if you go along with my earlier statements, even with the dynamics of *evaluation*. I have already indicated that these developments will probably lead to stronger connections between epistemic logic with the theory of agency in computer science, but also with learning theory in philosophy (as pioneered by Kelly and Hendricks), and between epistemic logic and game theory, where more attractive new models for rational agency may come about by endowing games with some epistemic agent structure. We seem to be moving toward an epistemic 'Theory of Play' rather than mere theory of games.

But more will have to happen, and the time seems ripe. I have already identified the mysterious interplay of external and internal views of knowledge of information, as exemplified by epistemic logic and intuitionistic logic. We need to understand this duality much better, and I am sure that we will. One further urgent desideratum is reintegrating basic themes from traditional philosophical logic, in particular *predication and quantification* over objects, merging knowledge of propositions with knowledge of objects. And once on this path from knowledge of propositions to that of objects, I would go even one step further. Think of learning and teaching. Most often, the precise propositional knowledge that we impart is only a means towards a more important end: knowledge of *methods*, acquiring skills that can work under different circumstances. You truly show that you know something when you can apply it, not in the same setting as the original, but in different ones. That is where we should go eventually.

At a higher abstraction level, I foresee several new alignments between fields triggered by modern developments in epistemic logic. I have already indicated some natural confluences inside philosophy between epistemic logic and learning theory, or between epistemic logic and logical theories of information. And I have pointed at the flourishing links between epistemic logic with computer science and game theory that keep inspiring 'joint ventures'. But there is more. A clear next goal down the line is forging links with *probability theory*, the major alternative paradigm in science and philosophy for analyzing information and agency. Probability crops up naturally and frequently even inside epistemic logic, once you start thinking about degrees of belief, global memory structure about the cognitive past, or mixed strategies for multi-agent behaviour in societies. In particular, there is no need to force a choice between 'competing paradigms' here: we should rather understand and cultivate the interface. And finally, I foresee new alliances between epistemic logic and cognitive science, looking at the intricacies of social cognition and intelligent interaction far beyond the usual studies of single-agent inference that have dominated the interface so far.

To conclude, please recall that in my view, all these strands are not disjoint from philosophy, as they form a natural intellectual whole, including attractive agenda extensions for traditional epistemology and related fields. So let me return to the beginning of this interview. Epistemic logic started as a tool for analyzing the philosophical notion of knowledge and its standard problems. It

may not have done too good a job at that. But what it has to offer these days is an enticing view of what epistemology *could become*, provided that philosophers are willing to loosen up, and engage in the same sort of agenda dynamics that has taken epistemic logic to its current vitality and scope.