EPISTEMIC LOGIC AND EPISTEMOLOGY: THE STATE OF THEIR AFFAIRS

Johan van Benthem

Amsterdam & Stanford, May 2005

1 Worlds apart?

Epistemology and epistemic logic At first sight, the modern agenda of epistemology has little to do with logic. Topics include different definitions of knowledge, its basic formal properties, debates between externalist and internalist positions, and above all: perennial encounters with sceptics lurking behind every street corner, especially in the US. The entry 'Epistemology' in the *Routledge Encyclopedia of Philosophy* (Klein 1993) and the anthology (Kim and Sosa 2000) give an up-to-date impression of the field. Now, epistemic logic started as a contribution to epistemology, or at least a tool in its modus operandi, with the seminal book *Knowledge and Belief* (Hintikka's 1962, 2005). Formulas like

$K_i\phi$ for "the agent *i* knows that ϕ"
$B_i\phi$ for "the agent *i* believes that ϕ"

provided logical forms for stating and analyzing philosophical propositions and arguments. And more than that, their model-theoretic semantics in terms of ranges of alternatives provided an appealing extensional way of thinking about what agents know or believe in a given situation. In particular, on Hintikka's view, an agent knows those propositions which are true in all situations compatible with what she knows about the actual world; i.e., her current range of uncertainty:

$\boldsymbol{M}, s \models K_i\phi$ iff for all $t \sim_i s$: $\boldsymbol{M}, t \models \phi$

These models for epistemic logic correspond to a widespread notion of *information* as a range of alternatives that are still open. New information will tend to shrink this range, perhaps until the actual world is all that is left: open to our scrutiny without a larger crowd of suspects to hide in. Since this is also a standard format for interpreting a universal modality, the laws of reasoning generated by this semantics are mostly familiar from modal logic. A typical example is the Distribution Axiom

$K_i(\phi \rightarrow \psi) \rightarrow (K_i\phi \rightarrow K_i\psi)$

Read epistemically, such axioms acquire a specific flavour, which has led to frequent philosophical debate concerning their plausibility. But debate, even including serious disagreement, implies potential communication – or so we shall claim later.

Epistemic logic going astray? So far, so good. But over time, epistemic logic has wandered into different fields. After lively initial concerns in the 1960s with philosophical issues of knowability, introspection into epistemic attitudes, or trans-world identification of objects, the field became a sleepy backwater in the 1970s. But around 1980, a Kiss of Life occurred when computer scientists became interested, because human-oriented metaphors of knowledge, ignorance, and communication turned out highly successful in understanding the behavior of complex interactive programs. In particular, the book *Reasoning about Knowledge* by the *IBM* Group (Fagin, Halpern, Moses & Vardi 1995), and the *TARK* conferences which these four authors initiated, made epistemic logic an interface theme between philosophy and computer science. Nowadays, one can go to computer science meetings on 'Agents', and hear scores of talks on knowledge, belief, desire, and intention. Another strand, and indeed an independent discovery, occurred in the 1970s when game theorists like Aumann in economics developed epistemic logics for representing what players know about actions by others (cf. Osborne & Rubinstein 1994), trying to provide logical underpinnings for the usual game-theoretic equilibrium concepts in terms of optimal behavior for rational agents. Thus – at least, in the eschatology of my own Dutch milieu of 'logical dynamics' – the agenda of epistemic logic has shifted to the study of information update, communication, and interaction among arbitrary agents, whether humans or machines – including the entangled virtual realities of Internet.

Or was there a Kiss of Death? If information engineers begin to like a subject, what can the nobility of the mind do except stand aloof? Epistemic logic plays a marginal role these days, even for many formal epistemologists. Just witness the lack of it in *Knowledge and the Flow of Information* (Dretske 1981) – or indeed, with Hintikka himself when discussing imperfect information of agents in his well-known games for independence-friendly logic (Hintikka & Sandu 1997, van Benthem 2005B). You may still find the occasional $K\phi$ operator in the literature, or discussions of the '*KK* Thesis', but these notations might just be the last vestiges of a passion long gone. For, many formal epistemologists started out as philosophical logicians! Was Churchill right this time: if you have not done epistemic logic before your thirties, you have no brain– but if you still do it after your thirties, you have no heart?

Persisting ties Maybe so. Personally, I find the lack of contact accidental, and the apparent distance magnified through ignorance. I see meeting-points all through the recent decades. For instance, the account of common knowledge in *Convention* (Lewis 1969) was an important contributions to epistemic logic, even though he did not formalize the notion – a task undertaken only by the computer scientists. Likewise, *Situations and Attitudes* (Barwise & Perry 1983), though critical of epistemic logic, eventually resorts to Hintikka-like modeling in terms of ranges of relevant situations, to get their account straight of attitude reports involving epistemic "seeing that". And the well-known monograph *Knowledge and its Limits* (Williamson 2000), though not a treatise on epistemic logic per se, definitely raises many logical issues concerning the valid laws of epistemic reasoning.

Indeed, it seems to me that many ongoing philosophical discussions about knowledge still show clear cultural influences from epistemic logic. It was there that issues of positive and negative introspection came to the fore in pregnant forms:

Is $K_i\phi \rightarrow K_i K_i\phi$ *valid?*
Is $\neg K_i\phi \rightarrow K_i \neg K_i\phi$ *valid?*

Epistemic logic would say that positions here depend on an analysis of the sort of 'access' that agents have to their range of indistinguishable alternatives. These can be either immediate, via transitive accessibility relations or even equivalence relations partitioning the logical space, or only in stages, in which case neither introspection principle holds. Likewise, the Distribution axiom validated by our modal semantics:

$$K_i(\phi \rightarrow \psi) \rightarrow (K_i\phi \rightarrow K_i\psi)$$

high-lighted, and I even suspect: engendered, the debate about logical omniscience. Is our knowledge closed under implications which we know to follow? Or at least, is our knowledge closed under its own logical consequences? Of course, the exact wording may change here, and many contemporary epistemologists are concerned with this logical schema in a very different guise, viz. the Sceptical Argument:

I know that I have two hands
I know that, if I have two hands, I am not a brain in a vat.
Therefore (?): I know that I am not a brain in a vat.

Also, their solutions to this particular problem need not be those of the logicians or computer scientists who have worried about omniscience. But see Section 2 below for a connection between logicians' attitudes and the contextualist way-out, which would claim that all three knowledge operators involved here come with their own contexts of

use, depending on the norms that we apply for 'knowledge' in each case. And such connections are enough as a basis for discussion. Indeed, every time I meet with epistemologists, concerns seem congenial to me, even though there is more to epistemology than just logic – and more to logic than epistemology. The rest of this paper is a brief discussion of issues of shared concern which arise when the agendas are put side by side. In the course of these issues, a more general view of knowledge emerges inspired by modern developments in logic, which I formulate at the end.

2 What is knowledge?

To arrange my list of topics, I start with the issue what knowledge really *is*.

Some famous accounts It is surely one of a philosophy student's most exciting experiences to read Plato's systematic discussion explaining why *true belief* is not what we should settle for – leading to the notion of knowledge as

justified true belief.

Of course, in a sadder but wiser post-Gettier world, this neat formula does not seem as satisfactory as it once was – but it can still serve us as a starting point for comparisons with logic. Just for a start, notice how Plato's account makes knowledge intertwined with other attitudes, viz. belief, while it also highlights the evidence for what we know, i.e., sources of knowledge and their certification. These issues will return below.

Now, the late 20th century has been exceptionally fertile in further conceptions with a similar broad force. Hintikka's work represents one such line, viewing knowledge as

truth throughout the logical space of possibilities

that the agent considers relevant. In modern terminology, this is the 'forcing view' (Hendricks 2005). By contrast, Dretske abandoned logic-oriented state-spaces in favor of mathematical information theory, and defined knowledge as

belief supported by reliable correlations

supporting genuine information flow. And the creative period of new basic definitions is not yet over, witness the intriguing idea in Nozick 1981 that knowledge of P involves a counterfactual aspect (written here in my own, simplified rendering):

true belief in P, while, if P had not been the case, I would have believed $\neg P$.

On the latter account, intriguingly, knowledge becomes interwined, not only with static beliefs, but with dynamic actions of belief *revision* underlying the counterfactual.

Some immediate connections with logic From my point of view, all these accounts state something of value and appeal, and they all raise intriguing issues without an obvious solution. For instance, all four definitions relate knowledge to further basic notions, such as truth, belief, information, and counterfactuals. This entanglement links epistemology to many parts of logic beyond pure 'epistemic logic' more narrowly construed. More specifically, Plato's definition requires dealing with justifications in some explicit manner, which is definitely beyond epistemic logic as it stands. Dretske's notion raises the old issue, already considered by Carnap, how the intuitive information concept in terms of logical state spaces relates to that underlying Shannon's information theory (cf. van Benthem 2004A). And Nozick's formula

$$K_i\phi \;\leftrightarrow \phi \;\&\; B_i\phi \;\&\; \neg\phi \;\Rightarrow B_i\neg\phi$$

is even an open challenge to logicians of a very concrete kind. It is easy to see that its adoption will block the above laws of standard epistemic logic, such as Distribution or Introspection. But are there any valid inference patterns left? Given some plausible background logic of belief and counterfactuals, *what is the complete set ofvalidities* of Nozick's *K*? Pittsburgh leads the way here. (Arlo Costa 2005) has a modal logic-style formulation in terms of neighborhood topology, (Kelly 2002) proposes a more computational account in terms of learning theory over a branching temporal universe.

The same immediate resonance on the part of a logician occurs with modern debates in epistemology, such as the *contextualist* line on the Skeptical Argument. The latter says that the notions of knowledge in the above argument are not all the same, and that one should specify which standards are applied. I know that I have two hands in some perceptual context c_1, I know that, if I have two hands, I am not a brain in a vat in some reflective context c_2, and perhaps the conclusion that I know that I am not a brain in a vat refers to yet another context c_3. Whether the inference is valid depends on how the three context are related, but it certainly will not hold in general. As it stands, this might seem like the age-old move of adding a parameter to block an undesired inference. More is needed philosophically: cf. Egré 2004 on the distinction between perceptual and reflective knowledge. But there is indeed an independent motivation! Contexts of various kinds are a powerful and intuitive device in explaining the workings of natural language, information update in computer science, and knowledge representation in *AI* (van Benthem & ter Meulen 1997, McCarthy 1993). Moreover, context change is essential to understanding how people communicate, or how information can travel from one location to another. In that light, the contextualist move in the debate on skepticism is not a trick, but a ticket to shared concerns with other academic communities. Note that my point here is not that there exists some

canonical formal theory of context structure and context change which epistemologists can now just adopt, and 'apply'. There is no such received theory across the areas which I mentioned. Perhaps we rather need the *philosophers* to look at what is going on in these areas around them, and create the right theory of context. But what does seem clear, even then, is that there is something substantial to talk about.

Diversity and relativism? But perhaps, all this richness of accounts is really poverty? Are we witnessing the disintegration of 'knowledge' into a family living apart together? There are different competing core definitions, and each of these might still be relativized to a potential infinity of contexts. Now, making distinctions can indeed be a two-edged device, creating consistency at the expense of coherence. But it does not necessarily mean dissolution. We could also say we are bringing to light necessary *parameters* for the application of the concept of knowledge, such as the role of evidence or context (Dretske 2004). There may still be enough family resemblance then between the various parametrized notions to justify a joint study. In particular, all of the above accounts of knowledge still share some undeniable family features. The known proposition is *true*, and the agent *believes* it on the basis of information picked up by certain actions. And more intangibly, but very crucially, there is a certain *robustness* to the situation. Knowledge is not attained lightly, and the way it was acquired guarantees a certain stability through changing situations. All this may not amount to a unique philosophical definition of knowledge. But it certainly delineates a genuine notion of wide range and appeal, well worth our attention.

We will return to this issue of unity in diversity toward the end of this paper. For the moment, we engage in a number of explorations, broadening the various interfaces with logic - 'epistemic' or not - briefly touched upon so far.

3 Clusters of epistemic attitudes, and epistemic actions

Knowledge and its neighbors The preceding definitions involve clusters of topics that have been studied in logic over the last decades, such as knowledge, belief, conditionals, and belief revision. Modern logics in computer science even deal with the 'Belief-Desire-Intention' framework of agency (Wooldridge 2002), where actions can only be explained through bringing together these various aspects in one system. On the general epistemological side, there is an issue then of the appropriate *cluster of notions* to be studied. Different definitions of knowledge bring together topics that may cross standard boundaries within philosophy. E.g., on Plato's and Hintikka's account, knowledge also involves belief, and the neglected historical gem Lenzen 1980 turned this into a rich and highly original study of a much richer cluster of attitudes, including being 'convinced', considering propositions 'probable', and yet others. Or, on

Nozick's account, we must deal with counterfactual conditionals, traditionally more of a topic in logic and the philosophy of science. During the 1980s and 1990s, it has become clear in logic and *AI* that conditionals are linked intimately with belief revision, and hence the 'epistemic cluster' even grows to include various cognitive actions.

But also independently, it seems obvious that we should study many epistemic concepts in combination. If knowledge is something like a gold standard, we only understand it by also analyzing less solid currencies, such as belief or 'understanding' – just as we learn vital facts about potential partners by taking a good look at their siblings and parents. Moreover, common parlance seems relevant. We would only say that someone 'knows' *P* if that person displays further expert behavior having to do with *P*. She should have learnt *P* on the basis of reliable procedures, but she should also be able to repeat the trick: be able to learn other things related to *P*, and very importantly, she should be able to *communicate* her knowledge to others. Whether this should go into the definition of knowledge may be debated (see Section 5 below), but all these features definitely form a natural agenda of things that belong together.

Logic combinations This pluralist view also reflects major trends in of modern logical research. Traditional philosophical logic was splintered in sometimes ridiculous ways, with subfields called 'modal', 'temporal', or even 'epistemic', 'doxastic', 'erotetic' or 'deontic' logic – the unfortunate result, one fears, of a desire to show off one's classical education. This led to niche formation which has been harmful to philosophical logic by and large. But the reality to-day is *combination of logics*, since about every meaningful task to be analyzed involves many of these things at once. Consider the simplest *conversation* about any topic. We cannot make sense of it logically unless we describe what people know, believe, desire, say, or do.

And if there is something which we as logicians can contribute to the philosophical discussion of the 'epistemic cluster' in its entirety, it is the growing awareness that combinations are not just a matter of putting ingredients together. In particular, the mathematical *complexity* of combined systems may remain simple, but it may also explode, depending in subtle ways on the *manner of combination*. For instance, consider epistemic agents that act on their world. Reasoning about this without any interactions between knowledge and action is about as simple as reasoning in existing separate modal logics of knowledge and action. But if we assume that the agents have Perfect Recall – i.e., for all actions *a*, they satisfy the axiom

$K[a]\phi \rightarrow [a]K\phi$ if you know the effect of your action beforehand, you will know afterwards that the effect holds .

then the logic can even become *undecidable* (Halpern & Vardi 1989)! Incidentally, Perfect Recall only holds for epistemically transparent actions *a*: it fails otherwise. I know that I am boring after drinking – but tragically, after drinking, I do not know that I am boring. Similar points about surprising emergent complexity can be made about combinations of modal logics for various epistemic attitudes (cf. Spaan 1993) .

Thus, current experience suggests that combining logics may lead to the emergence of new phenomena, depending on the mode of combination. Likewise, combined studies of clusters of epistemic notions may have many surprises in store.

4 Sources and evidence

Plato's formula for knowledge highlights the existence of a justification. This naturally leads us to consider the sources of knowledge. One traditional tripartition distinguishes three main sources for this purpose: *deduction*, *observation*, and *questioning*. If I want to know whether the Stanford campus has a beer outlet, I can try to deduce the answer to this from the immense amount of leaflets and brochures put at our disposal each year by a benevolent administration, or I can make a walk around the site checking each house or shed – or I can just ask a reliable authority: say, a student. It seems hard to dissociate the notion of knowledge as some sort of mental or world-attuned attitude per se from an account of these sources of evidence. And indeed, Plato builds justification into his account of knowledge, while Dretske insist on some informational correlation underpinning the knowledge.

∀ versus ∃ Even so, there is a curious 'non-homogeneous' feature here from a logic perspective. The heart of Hintikka's analysis was a *universal quantifier*: $K_i\phi$ says that

> ϕ *is true in* all *situations agent i considers as candidates for the current s.*

This universal quantifier explains basic features of the logic of knowledge such as the validity of modal Distribution and other laws: a universal quantifier distributes over implications, hence, so does a knowledge operator. But the quantifier in the other part of Plato's Formula is *existential*: it says that

> there exists *a justification*.

In that light, knowledge consists in having strong evidence for a proposition, in the strongest case perhaps: a mathematical proof. Now, this co-existence of two views is not unheard of in logic. Consider the two main notions of logical validity. The semantic notion says that a proposition is universally valid: i.e., true on all domains under all interpretations. The syntactic notion says that there exists a proof for the proposition. And Gödel's *completeness theorem* established a harmony, at least for

first-order predicate logic: a formula satisfies the first condition if and only if it satisfies the second. But this is not quite what is at stake in Plato's formula, since there is no equivalence. The justification needs to come on top of the other features (say, a belief operator $B_i\phi$ read universally in Hintikka's semantic style) to have 'knowledge'. That is, for instance, why van Benthem 1993 proposed a merge of epistemic logic with a calculus of evidence, in order to do its epistemological job more properly.

Co-existence of known propositions and evidence And indeed, logic does provide instances of co-existence of knowledge and justification. In particular, proof-theoretic accounts of *intuitionistic logic* manipulate binary type-theoretic assertions of the form

x is a proof for ϕ.

Now, intuitionism is an implicitly epistemic philosophy, as knowledge of agents is not mentioned in the formalism, and only built into the intuitionist understanding of the logical operations. Still, proof-theoretic accounts of knowledge have been influential, e.g., in Dummett's verificationism (cf. Section 5 below for a modern take). Moreover, broader proof theories exists which handle binary statements where the above *x* can be any sort of evidence for ϕ: witness the 'labeled deductive systems' of Gabbay 1996.

But can we also combine epistemic logic with its explicit *K*-operators and explicit justifications directly? Here is a lead. Consider the *provability interpretation* of modal logic, where we read a necessity operator as saying that

$[]\phi$: *there is a proof for* ϕ

in some relevant proof calculus. The $\forall$ and $\exists$ accounts of $[]$ are in harmony to some extent in their basic logic. E.g., modal Distribution is still valid. When we have a proof for $\phi \rightarrow \psi$ and one for ϕ, then putting the two together, and adding one instance of Modus Ponens will produce a proof of ψ. And as for Introspection, if we have a proof for ϕ, then certainly any proof-checking algorithm will present us with a proof that the given sequence of formulas is a proof for ϕ.

But these very observations show how evidence might be made explicit. We must unpack the existential quantifier in "there exists a proof" to the specific items supporting it. For instance, the preceding justification of Distribution really provides more information than what is stated in the usual axiom $K_i(\phi \rightarrow \psi) \rightarrow (K_i\phi \rightarrow K_i\psi)$. Adding proof terms to the language, it really is the more informative statement that

$$[x](\phi \rightarrow \psi) \,\&\, [y]\phi \;\rightarrow\; [x\#y]\psi,$$

where # is some appropriate *sum operation* on proofs, or pieces of evidence generally. This idea has been developed in a sophisticated manner in the 'modal logic

of proofs' of Artemov 1994, 2005, which also includes operations of 'choice' and 'checking' on proofs. Notice that this same approach also makes sense in our earlier discussion of the contextualist view of knowledge. In that case the crucial law is

$$[c_1]\, K_i(\phi \to \psi) \,\&\, [c_2]\, K_i\phi \;\to\; [c_1 \# c_2]\, K_i\psi,$$

with # now an operation of context merge yet to be defined in precise logical terms. And so, we are back where we were: we know what to do, but it still needs to be done! In the following section, we turn to some developments that have already happened.

5 Logical dynamics: bring in the actions!

Much of epistemology has been concerned with what it means to possess knowledge, as a sort of 24-carat information that is available to us in reliable ways. Sophisticated accounts of such intuitions about quality are the backbone of the field. By contrast, much recent work in epistemic logic has concentrated on dynamic mechanisms that produce or modify knowledge and related epistemic attitudes like belief – such as speech acts, communication, observation, learning, or even more radical belief revision. This action-oriented trend shows an influence from computer science. For, one of the most powerful ideas in that field is the Tandem View:

representations and processes must always be studied together.

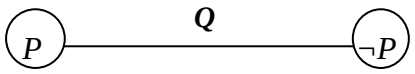
You cannot understand a process without thinking about the right data structures for it to work on, and you cannot design good data structures without having some process in mind that is going to use them. In logic since the 1980s, this insight has led to new theories of processes as first-class citizens such as information update, belief revision, natural language interpretation, and many others (Gärdenfors 1987, Kamp & Reyle 1993, Groenendijk & Stokhof 1991, Veltman 1996, van Benthem, Muskens & Visser 1997). The resulting program of uniformly treating logical propositions and actions on a par has been called *Logical Dynamics* in van Benthem 1996, van Benthem 2003.

Epistemic actions Taken to epistemology, the Tandem View would say that we need to treat knowledge and the actions producing and transforming it on a par. This view fits quite well with our earlier common sense observation that much of the 'quality' one associates with knowledge does not reside in some static relationship between a proposition, an agent and the world, but rather in some sustained dynamic behavior of being able to learn and communicate. In that light, the quality of what we *have* epistemically resides largely in what we *do* – individually, or socially in interaction with others. Once you start thinking that way, a wide space of phenomena unfolds. When I say "I see that ϕ" I really refer to an act of observation or comprehension;

when I ask a question, I tap into the knowledge of others. Knowledge involves learning, grasping, questioning, inferring, and so on. Without a grasp of this wider space of epistemic actions, just as solid as that of 'knowledge' per se, we do not know what to do with all that 24 carat gold stored in the vaults of our Mental Bank.

Kant once observed that every good idea turns out to have philosophical predecessors. There are surely philosophical precedents for the view taken here. In what follows, we give a few examples of the viability of this style of thinking allied to epistemic logic – which therefore still serves a 'laboratory' for inventing and elaborating new ideas.

Questions and answers in epistemic logic Question-answer scenarios are among the most elementary actions in communicating knowledge. I ask you if *P* is the case, and you answer truly: "Yes". In normal Gricean circumstances, my question will tell you that I do not know if *P*, while I consider it possible that you do know. This would be the case, e.g., in the simple situation pictured by the following epistemic model, where the line between the two alternative worlds indicates my uncertainty:



Since you have no uncertainty lines in either world, you know whether *P* throughout, and I actually know that you know, as it is true in both of my alternatives. Now, your answer to my question changes this model, ruling out the ¬*P*-world. In modern jargon, an *update* takes place to the single-world model

where both you and I know that *P*, and we know this about each others, and so on: *P* has become common knowledge between the two of us. Standard epistemic logic can describe the various knowledge assertions involved in the separate stages of this process, including interactive ones concerning knowledge about others. Before the question was asked, in the initial model, the following assertions were true, expressing the two mentioned preconditions for asking a cooperative question:

$\neg K_Q P \;\&\; \neg K_Q \neg P$
$\langle \boldsymbol{Q} \rangle (K_A P \vee \boldsymbol{K_A} \neg P)$

After the *update* with *P*, the following assertions have become true:

$K_Q P,\ K_A K_Q P,\ K_Q K_A K_Q P, \ldots$ i.e., iterated mutual knowledge,
and even *common knowledge* $C_{\{Q, A\}} P$ in the group $\{\boldsymbol{Q}, \boldsymbol{A}\}$.

But this account of the question-answer episode still does not deal with the epistemic actions per se. For that, we borrow an idea from computer science, viz. the *dynamic logic* of programs and actions in general (Harel, Kozen & Tiuryn 2000).

Dynamic epistemic logic Let us introduce a modal operator stating effects of actions:

$[A!]\phi$ *after a truthful public announcement of proposition A,* ϕ *holds*

Typical uses of this notation in a communicative setting are

$[A!]K_i\phi$ after announcement of *A*, agent *i* knows that ϕ
$[A!]C_G\phi$ after announcement of *A*, ϕ is common knowledge

Logical formulas are often nuggets of history. With these few symbols, the above formulas combine ideas from philosophy, linguistics, and computer science! A decade of research on logics like this has shown that they can be developed just like standard epistemic logics, including intriguing properties of communicative actions (cf. the survey van Benthem 2002A, with an extensive list of credits and references). For instance, a key valid principle is the following 'Knowledge Prediction Axiom':

$$[A!]K_i\phi \quad \leftrightarrow \quad A \rightarrow K_i[A!]\phi$$

The interchange of knowledge and announcement actions displayed here reflects, amongst other features, the epistemic transparency of public announcements – resulting in Perfect Recall for epistemic agents (cf. Section 3). Nevertheless, the earlier threat of undecidability does not materialize: the basic dynamic-epistemic logic of announcements is decidable, and no more complex than Hintikka's base logic.

A new logic like this acts as a searchlight for finding new phenomena. Here is one example with mnay ramifications – and indeed, a pedigree in epistemology. We start with a question about epistemic intuitions. If actions are first-class citizens now, we should not just say what knowledge is, but also try to formulate with equal care what epistemic actions do. So, what is the effect of a public announcement? In particular,

Does public announcement of A always lead to common knowledge of A?

In terms of our dynamic-epistemic logic, the following axiom would then be valid:

$$[A!]C_G A$$

If nothing else, this at least enriches the set of typographical logical forms that epistemologists could argue about. Though intuitively plausible at first sight, this principle of epistemic action founders on *Moore-type* statements. Consider this case:

$\neg K_{you}P \ \& \ P$ "you do not know it, but *P*"

This may well be true now, but once uttered, it makes you know that *P*, and indeed, *P* becomes common knowledge, thereby invalidating $\neg K_{you}P \ \& P$ as a whole. It is not known yet precisely which forms of assertion produce common knowledge – an open problem known among dynamic-epistemic logicians as the 'Learning Problem'. Self-refuting assertions of the Moore type occur frequently to good effect in knowledge puzzles and games (van Benthem 2002B), so the example is no philosopher's fluke.

Verificationism and learnability The way back from technical developments in modern epistemic logic to issues in mainstream epistemology can be quite short. The following illustration is taken from van Benthem 2004B. Consider the simple but exasperating *Fitch Paradox* concerning the Dummett-style

Verificationist Thesis *What is true can be known,*

or in epistemic terms, plus some unspecified modality for the "can":

$\phi \rightarrow \Diamond K\phi$

Fitch gave the following simple argument showing that the Verificationist Thesis is inconsistent. The heart of the problem is again a Moore-style assertion:

$P \wedge \neg KP \ \rightarrow \ \Diamond K(P \wedge \neg KP) \rightarrow \ \Diamond (KP \wedge K\neg KP) \ \rightarrow \ \Diamond (KP \wedge \neg KP) \rightarrow \bot$

Therefore, we may conclude that

$P \rightarrow KP$, i.e., knowledge and truth collapse!

There is a booming literature on solving this paradox, but one obvious link is with epistemic actions. The Verificationist Thesis claims that we can come to know every true assertion – presumably by some epistemic action, hidden under the existential quantifier of the "can". Consider the simplest actions possible, viz. public announcements. Perhaps God will reveal all to us. Then, our earlier dynamic-epistemic observation about announcements invalidating themselves applies. The principle

What is true can be learnt through announcement
$\phi \rightarrow \exists A\colon [A!]K\phi$

is false in general. But the technical logical problem of finding out which assertions do produce their own common knowledge now translates into the perfectly respectable philosophical question *just which versions* of Verificationism are tenable.

More logical dynamics Public announcements are just the tip of an epistemic iceberg. Other relevant actions include partial observation, hiding of information, or coded communication, all the way to the most sophisticated epistemic abilities such as lying and cheating. Moreover, these actions do not just concern knowledge, but also

belief, and belief revision (Gärdenfors & Rott 1995, Spohn 1988, Aucher 2003). Such actions, too, are studied in modern epistemic logic (cf. Baltag, Moss & Solecki 1998, van Benthem 2002A), and again, they seem relevant to our daily practice. We provide judicious mixtures of open and private information when using email with buttons like *cc* and *bcc*, and we keep track over time – as best as we can – of the resulting epistemic states of mutual knowledge or belief. And in parlor games, we expose ourselves to a repertoire of complicated informational moves, thereby testing our cognitive abilities in self-imposed epistemic 'laboratory conditions'. Again, one can think of this development as mere application of epistemic logic to computational engineering, far removed from true epistemology. But the more profitable way of thinking would be to see the challenge. Informational computational models like these are a rich source of thinking about human action, far beyond the bleak solipsism of the Turing Machines which *have* captured the philosophers' imagination.

6 Multi-agents, interactive epistemology, and groups

The Dynamic Turn puts actions at center stage as objects of logical study. Now these could still be individual actions of update and revision by agents on their own. But the earlier examples also point at the *social* character of cognition. This was already true in the original version of epistemic logic. Even though Hintikka emphasized the lonely Thinker and his knowledge, an equally exciting feature of the formal language was its ability to iterate assertions

$$K_i K_j \phi$$

describing agents' knowledge about what others know. Common knowledge

$$C_G \phi$$

then took this one step further to notions that essentially describe knowledge residing in a *group* of agents. Merged with logical dynamics, this brings us to the setting of group knowledge and group action. The most obvious instance of this are *games*.

Games Single update steps as described in Section 5 are just building blocks in longer conversations. We have plans for asking and answering, and more generally, saying things. And these plans serve a purpose. In particular, behind every question, there is a meta-question of *Why*? What are the intentions and goals of the speaker, and what are the dynamics of goals and desires as a conversation unfolds? The resulting interaction leads naturally to the setting of games. As we mentioned before, epistemic logic was discovered independently in the 1970s in game theory, and a confluence between research communities took place in the 1980s through the TARK

conferences. A driving force for epistemic analysis in game theory was the analysis of rational behavior, an enterprise very close to classical philosophical concerns (cf. De Bruin 2004). But by and large, it was only in the 1990s that the so-called *interactive epistemology* of games began to have serious impact in the hands of prominent philosophers (cf. Stalnaker 1999). This might also lead to the incorporation of *stategies* and plans as first-class logical citizens, taking logical dynamics to its next natural interactive phase. All this is beyond our story here, however. Instead we just make a few comments about two further relevant aspects of logical dynamics.

From single updates to learning The term 'learning' has been used in a very loose sense so far in our story of epistemic dynamics. More elaborate formal accounts of learning mechanisms and their epistemic relevance are found in (Kelly 1996) and Hendricks (2002, 2005). It is quite attractive to view learning theory as a natural continuation of dynamic-epistemic logic. It kicks in with long-term processes that require the larger arena of a branching temporal universe (van Benthem 2005A). A real unification of the two approaches might make for a very powerful coalition!

Group action and 'social' knowledge Epistemic logic has considered some notions of group knowledge, including not just common knowledge, but also 'distributed knowledge' which the group would have if agents pooled their information. But taking groups seriously as epistemic subjects in their own right would require a more sustained analysis of ways in which groups as plural subjects might be said to know propositions. In particular, such notions would require a structured account of ways in which group members can communicate, i.e., the *channels* in the group (Barwise & Seligman 1995). In parallel with this, one would also need an account of group action, perhaps beyond individual action – cf. the philosophical research on shared agency starting from Bratman 1993. There is some promising work on logics of powers of game-theoretic coalitions (Pauly 2001), but that is only the beginning – and it does not include epistemic considerations. One measure of the complexity of the logic of collective knowledge and action is our own vocabulary in natural language. The way we describe what "we" or "they" do together, or to "each other" in terms of collective predication is complex (Landman 1989, van der Does 1992), and no simple reduction to indidivual actions seems to work. Some linguists have suggested that we need to think of linguistic plural subjects as distributed information systems that can act, but again, this just shows that the linguists do not have a ready-made answer for us either.

But fact remains that we do switch easily from individual to plural perspectives in both knowledge and action – and perhaps epistemology should take this more seriously. Clerk Maxwell once quipped that, if a scientist says "We now think that such-and-

such", this just means "All people who thought otherwise are dead". But surely, there is more to the cognitive plural than this! Now, epistemic logic has nothing decisive to offer here so far. But for the purpose of this paper, it is enough to see that, in this area as in the earlier ones, it finds itself in the same boat with mainstream epistemology.

7 Conclusions

I conclude with a few general thoughts raised by writing this paper.

Omissions This paper has looked at some developments in modern epistemic logic that seem to run in parallel with epistemology in general. Many further illustrations could have been given, and many further issues could have been raised to broaden the interface. I have ignored (another epistemic action well worth studying!) longer-term epistemic behavior over time, evolution of cognitive practices (Skyrms 1990), learning theory (Kelly 1996**,** Hendricks 2002), as well as connections with other technical disciplines than logic that are relevant to epistemology, such as probability theory or information theory. True, but this does not weaken the points about fruitful contacts.

Logic in philosophy Even so, many people would still doubt the relevance of bringing logic and philosophy together along topics of shared concern. What good will the encounter do? Well, it is not exactly a blind date. We do know what logic-philosophy interfaces have achieved in the past. Let's consider the benefits for logic first. I would just say philosophers have more open minds toward the delicacies and complexities of many subjects, and that virtue can help keep logicians more open-minded and less internally driven by the needs of proving theorems in some ritualized industrial mode. But I think the story of this paper also demonstrates clear benefits to philosophy. Epistemic logic has shown its use as a tool for clarifying philosophical notions and arguments, a sort of notation entering into a creative interplay with its subject matter (I owe this wonderful point to Paul Egré). Indeed, this can take existing philosophical debates to new depths, witness the 'search-light function' of dynamic-epistemic logic in probing the viability of verificationism, or to mention another field involved here, speech act theory. Even more ambitiously, logical systems also provide ways of developing new philosophical views, the way Carnap used them (Leitgeb 2004). This fits the idea in Smullyan 1997 of 'crazy philosophers', people using logical tools as a means of creating new worlds and phantasies: much cheaper, and much less dangerous than mind-blowing drugs. And if one still feels that logicians sacrifice too much with the simplifications needed for their systems, I will quote a leading Dutch thinker who once said: "Any fool can see that the world is rich, beautiful, and complex. But it takes a genius to make one good simplification".

A view of knowledge In the course of this paper, a certain view of knowledge has emerged which I did not have when I started. As observed earlier, many philosophical views of knowledge try to get at its *robustness* or *stability*. I share that intuition. But the more I think of it, the more I see the robustness of knowledge, not as an isolated feature of a single agents and single propositions. It is rather something which can only be explained in a setting of further epistemic attitudes, further epistemic agents, and a rich repertoire of epistemic actions. The robustness of knowledge lies in its successful functioning in a complex epistemic environment: the one we live in. And therefore, both logic and epistemology need to set their sights accordingly.

Bridges This volume is about bridges, even *seven bridges* in the original invitation. That is a metaphor. And a rich one. I could not help thinking of the 'Königsberger Brücken', which Euler used in graph theory, and where Kant must have walked:

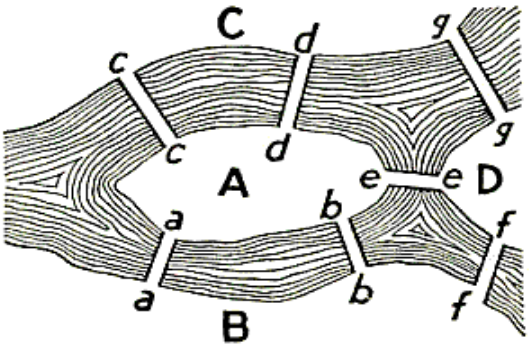


FIGURE 98. *Geographic Map: The Königsberg Bridges.*

Is logic one of the pieces of mainland here, to be connected by a bridge to the philosophical territory on the other side? Is it one of the bridges facilitating traffic between different fields: philosophy, linguistics, computer science? Or is logic the *island*? I cannot say. But I do know that bridges illustrate the main concerns of this paper. One must know where they are, they are made for dynamic actions of crossing, preferably by many agents, but groups should beware, and not march too much in step.

References

H. Arlo-Costa, 2005, 'First-Order Modal Logic', to appear in V. Hendricks & S. A. Pedersen, eds., '40 Years of Possible Worlds', special issue of *Studia Logica*.

S. Artemov, 1994, 'Logic of Proofs', *Annals of Pure and Applied Logic* 67, 29–59.

S. Artemov, 2005, 'Evidence-Based Common Knowledge', CUNY Graduate Center, New York.

G. Aucher, 2003, *A Joint System of Update Logic and Belief Revision*, Master of Logic Thesis, ILLC University of Amsterdam.

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

J. Barwise & J. Perry, 1983, *Situations and Attitudes*, The MIT Press, Cambridge (Mass.).

J. Barwise & J. Seligman, 1995, *Logic and Information Flow*, Cambridge University Press, Cambridge.

J. van Benthem, 1993,'Reflections on Epistemic Logic', *Logique et Analyse* 34, 5-14 (*TARK* lecture).

J. van Benthem, 2002A, 'One is a Lonely Number: on the Logic of Communication', Report PP-2002-27, ILLC Amsterdam. To appear in P. Koepke et al., eds., *Colloquium Logicum*, AMS Publications, Providence.

J. van Benthem, 2002B, 'Rational Dynamics', LOGAMAS workshop, department of computer science, University of Liverpool. In S. Vannucci, ed., *Logic, Game Theory and Social Choice III*, University of Siena, department of political economy, 19–23. To appear in *International Journal of Game Theory*.

J. van Benthem, 2003, 'Logic and the Dynamics of Information', in L. Floridi, ed., *Minds and Machines* 13:4, 503–519.

J. van Benthem, 2004A, 'Information as Correlation versus Information as Range', manuscript, ILLC, University of Amsterdam. To appear in L. Moss., ed., *Memorial Volume for Jon Barwise*.

J. van Benthem, 2004B, 'What One What One May Come to Know', *Analysis* 64 (282), 95–105.

J. van Benthem, 2005A, 'Open Problems in Logical Dynamics', ILLC Preprint:, Institute for Logic, Language and Information, University of Amsterdam, DARE electronic archive 148382. To appear in D. Gabbay, S. Goncharov & M. Zakharyashev, eds., *Mathematical Problems from Applied Logic*, Russian Academy of Sciences, Novosibirsk & Springer, New York.

J. van Benthem, 2005B, 'The Epistemic Logic of *IF* Games', in L. Hahn, ed., *Jaakko Hintikka, Library of Living Philosophers*, Southern Illinois University, Carus Publishers.

J. van Benthem & A. ter Meulen, eds., 1997, *Handbook of Logic and Language*, Elsevier, Amsterdam.

J. van Benthem, R. Muskens & A. Visser, 1997, 'Dynamics', in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier Science Publishers, Amsterdam, 587-648.

M. Bratman, 1992, 'Shared Cooperative Activity', *The Philosophical Review*101:2, 327–341.

B. de Bruin, 2004, *Explaining Games. On the logic of game-theoretic explanations*, Dissertation DS 2004-03, Institute for Logic, Language and Computation, University of Amsterdam.

J. van der Does, 1992, *Applied Quantifier Logics. Collectives and Naked Infinitives*, Dissertation, Institute for Logic, Language and Computation, University of Amsterdam

F. Dretske, 1981, *Knowledge and the Flow of Information*, Chicago University Press, Chicago.

F. Dretske, 2004, 'Externalism and Modest Contextualism', *Erkenntnis* 61, 173-186.

P. Egré, 2004, *Attitudes propositionnelles et paradoxes épistémiques* Thèse de Doctorat, Université Paris 1 Panthéon-Sorbonne, IHPST.

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

D. Gabbay, 1996, *Labelling Deductive Systems (Vol.1)*, Clarendon, Oxford.

P. Gärdenfors, 1987, *Knowledge in Flux*, The MIT Press, Cambridge (Mass.).

P. Gärdenfors & H. Rott, 1995, 'Belief Revision', in D. M. Gabbay, C. J. Hogger & J. A. Robinson, eds., *Handbook of Logic in Artificial Intelligence and Logic Programming* 4, Oxford University Press, Oxford 1995.

J. Groenendijk & M. Stokhof, 1991, 'Dynamic Predicate Logic', *Linguistics andPhilosophy* 14(1), 39-100.

J. Halpern & M. Vardi, 1989, 'The Complexity of Reasoning about Knowledge and Time', *Journal of Computer and Systems Science* 38:1,195-237.

D. Harel, D. Kozen & J. Tiuryn, 2000, *Dynamic Logic*, The MIT Press, Cambridge (Mass.).

V. Hendricks, 2002, 'Active Agents', PHILOG Newsletter, Roskilde. In J. van Benthem & R. van Rooy, eds., special issue on Information Theories, *Journal of Logic, Language and Information* 12:4, 469–495.

V. Hendricks, 2005, *Mainstream and Formal Epistemology*, Cambridge University Press, New York.

J. Hintikka, 1962, *Knowledge and Belief*, Cornell University Press, Ithaca. Reprint 2005, V. Hendricks & J. Symons, eds., King's College Publications, London.

J. Hintikka & G. Sandu, 1997, ‘Game-Theoretical Semantics’, in J. van Benthem & A. ter Meulen, eds., *Handbook of Logic and Language*, Elsevier, Amsterdam, 361–410.

H. Kamp & U. Reyle, 1993, *From Logic to Discourse*, Kluwer, Dordrecht.

K. Kelly, 1996, *The Logic of Reliable Inquiry*, Oxford University Press, Oxford.

K. Kelly, 2002, 'Knowledge as Reliable Inferred Stable True Belief', Department of Philosophy, Carnegie Mellon University, Pittsburgh.

J. Kim & E. Sosa, eds., 2000, *Epistemology: An Anthology*, Blackwell, Malden (Mass.).

P. Klein, 1993, 'Epistemology', *Routledge Encyclopedia of Philosophy* , Routledge, London.

D. Kozen, D. Harel & J. Tiuryn, 2000, *Dynamic Logic*, The MIT Press, Cambridge (Mass.).

F. Landman, 1989, 'Groups I, Groups II', *Linguistics and Philosophy* 12, 559 - 605, and 723 - 744.

H. Leitgeb, 2004, 'Carnap's Logischer Aufbau Revisited', Department of Philosophy, University of Salzburg.

W. Lenzen, 1980, *Glauben, Wissen und Wahrscheinlichkeit*. Springer Verlag, Wien, Library of Exact Philosophy.

D. Lewis, 1969, *Convention*, Blackwell, Oxford.

J. McCarthy, 1993, 'Notes on Formalizing Context', *Proceedings of the 13th International Joint Conference in Artificial Intelligence* (IJCAI'93).

R. Nozick, 1981, *Philosophical Explanations*, Harvard University Press, Cambridge (Mass.).

M. Osborne & A. Rubinstein, 1994, *A Course in Game Theory*, The MIT Press, Cambridge (Mass.).

M. Pauly, 2001, *Logic for Social Software*, dissertation DS-2001-10, Institute for Logic, Language and Computation, University of Amsterdam.

B. Skyrms, 1990, *The Dynamics of Rational Deliberation*, Harvard University Press, Cambridge, (Mass.).

R. Smullyan, 1997, *The Tao is Silent*, Harper & Row, New York.

E. Spaan, 1993, *Complexity of Pure and Applied Intensional Logics*, Dissertation,, Institute for Logic, Language and Computation, University of Amsterdam.

W. Spohn, 1988, 'Ordinal Conditional Functions. A Dynamic Theory of Epistemic States', in: W.L. Harper, B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics*, vol. II, Kluwer, Dordrecht 1988, 105-134.

R. Stalnaker, 1999, 'Extensive and Strategic Form: Games and Models for Games', *Research in Economics* 53:2, 93-291.

F. Veltman, 1996, 'Defaults in Update Semantics', *Journal of Philosophical Logic* 25, 221–261.

T. Williamson, 2000, *Knowledge and its Limits*, Oxford University Press, Oxford.

M. Wooldridge, 2002, *An Introduction to Multi-Agent Systems*, John Wiley, Colchester.