What One May Come to Know

JOHAN VAN BENTHEM

1 Verificationism and Fitch's Paradox

The general verificationist thesis says that

What is true can be known or formally: $\phi \to \Diamond K\phi$ ***VT***

Fitch's argument trivializes this principle. It uses a weak modal epistemic logic to show that ***VT*** collapses truth and knowledge, by taking a clever substitution instance for ϕ:

$$P \wedge \neg KP \to \Diamond K(P \wedge \neg KP)$$

Then we have the following chain of three conditionals

(a) $\Diamond K(P \wedge \neg KP) \to \Diamond (KP \wedge K\neg KP)$

in the minimal modal logic for the knowledge operator K,

(b) $\Diamond (KP \wedge K\neg KP) \to \Diamond (KP \wedge \neg KP)$ in the modal logic T,

and finally (c) $\Diamond (KP \wedge \neg KP) \to \bot$ in the minimal modal logic for <>.

Thus, a contradiction follows from the assumption $P \wedge \neg KP$, and we have shown over-all that P implies KP, making truth and knowledge equivalent.

Proposed remedies for the Paradox fall mainly into two kinds (cf. Brogaard and Salerno 2002, Wansing 2002). Some weaken the logic in the argument still further. This is like tuning down the volume on your radio so as not to hear the bad news. You will not hear much good news either. Other remedies leave the logic untouched, but weaken the verificationist principle itself. This is like censoring the news: you hear things loud and clear, but they may not be so interesting. The proposal made below falls into the latter category, but using a different perspective from mere tinkering with proof rules or premises. We will emphasize positive reasons why ***VT*** can, and sometimes should fail, having to do with the ways in which we learn new information.

2 Knowable propositions and learning

Fitch's substitution instance exemplifies a much older problem about knowledge called *Moore's Paradox*. It consists in the observation that the statement

"*P*, but I don't know it"

can be true, but it cannot be known, as $K(P \ \& \ \neg KP)$ evidently implies a contradiction. Now, in an epistemic logic for a single agent, the possible knowledge of a proposition ϕ requires that $K\phi$ be satisfiable at some world in some model, and hence in all alternatives to that world. This differs from ordinary epistemic satisfiability, which just demands truth of ϕ at some world in some model. Tennant 2002 argues persuasively for the following restriction on the intended applications of ***VT*** to propositions ϕ:

$K\phi$ is consistent $\qquad\qquad$ $CK(\phi)$

In simple epistemic *S5*-models, this special requirement amounts to *global satisfiability* of ϕ: i.e., its truth throughout some model. Like ordinary satisfiability, the new notion is decidable for most modal logics (Note 1), and hence constraints of knowability can be formulated at least in an effective manner. But there is a bit more to the situation! Our first observation is that *CK* only partially captures the intuition behind ***VT***.

3 A dynamic shift: consistent update

Consider any epistemic model ***M***, *s*, with a designated world *s* for the actual situation. What might be known in this setting seems restricted to what might be known correctly *about that situation s*. We know already that it is one of the worlds in ***M***. What we might learn is that this model can be shrunk still further, zooming in on the location of *s*. In this dynamic sense, the verificationist principle that every true statement may be known amounts to stating that

What is true may come to be *known* $\qquad\qquad$ $\boldsymbol{VT}^*$

Clearly, $\boldsymbol{VT}^*$ only holds for propositions ϕ that satisfy CK. But it is more demanding. We need truth of $K\phi$ not in just any model, but in some submodel of the current one.

Fact $\quad CK(\phi)$ does not imply $\boldsymbol{VT}^*$ for all propositions ϕ.

Proof Let $\langle\rangle\phi$ be the existential dual of the operator *K*, standing for the epistemic (not the earlier modal!) notion of 'holding it possible that ϕ'. Now consider the statement

$$\phi = (P \ \& \ \langle\rangle\neg P) \vee K\neg P$$

This is knowable in the sense of CK, since $K((P \ \& \ <>\neg P) \lor K\neg P)$ is consistent: it holds in a model consisting of just one world with $\neg P$, where the agent knows that $\neg P$. But here is a two-world model ***M*** where ϕ holds in the actual world, even though there is no truthful announcement that would make us learn ϕ:

the actual world	....	some other world
P		$\neg P$

In the actual world, $(P \ \& \ <>\neg P) \lor K\neg P$ holds, but it fails in the other. Hence, $K((P \ \& \ <>\neg P) \lor K\neg P)$ fails in the actual world. The only truthful proper update of this model ***M*** just retains its actual world. But in the resulting one-world epistemic model with the proposition letter P true, $K((P \ \& \ <>\neg P) \lor K\neg P)$ fails. **QED**

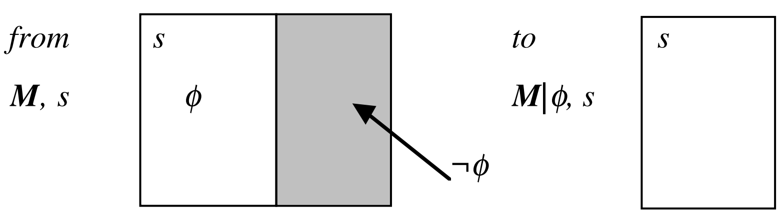
Thus, consistency of $K\phi$ need not be enough if we wish to learn that ϕ here and now in any model where it holds, as expressed by ***VT****. Our first point is then that

> *In a natural learning scenario, the Verificationist Thesis places stronger requirements on propositions than those stated so far in the literature.*

This observation suggests a closer look at the general dynamic viewpoint underpinning ***VT****. In a nutshell, what we know is the result of *actions of learning*.

4 Epistemic logic dynamified

The simplest way of learning is by being told through a true new proposition which prunes the current epistemic model. In particular, a *public announcement* $\phi!$ of assertion ϕ does not just evaluate ϕ truth-conditionally in the current model ***M***, s. It rather changes that model, by eliminating all those worlds from it which fail to satisfy ϕ:



This scenario works for simple questions and answers, but also for more intricate puzzles of knowledge and ignorance (van Benthem 2003; Note 2). Thus, the semantic setting for elementary forms of learning is a family of epistemic models as the relevant information states, related by a repertoire of actions of announcing propositions that increase information by moving from one model to another. Complete systems for this dynamified epistemics mix standard epistemic logics with dynamic logics of actions, with expressions describing what holds after an action was performed:

$[a]\psi$ ψ holds after every successful execution of action a

The expression a may be a computer program or some physical action, or a speech act. Now we can express statements about epistemic effects of successful communication:

$[\phi!]K_j\psi$ after a true public announcement of ϕ, j knows that ψ

There are complete and decidable calculi for this richer language. These include axioms relating statements about the result of an action to those that were true before. (Note 3.) In particular, epistemic logics for communication emphasize the multi-agent character of speakers, hearers, and audiences. Accordingly, the language can iterate knowledge assertions, as in $K_1\neg K_2P$: '*1* knows that *2* does not know that P'. There are also new notions for groups of agents G, such as common knowledge $C_G\phi$: everyone knows that ϕ, and they also know that the others know, and on to any finite depth of iteration. This point will return, as paradoxes for lonesome knowers may look brighter in groups.

5 The dynamic logic of learning

One open issue in dynamic epistemic logics concerns the generic effects of public announcement. Here is a putative, almost self-evident principle about the epistemic effects of a public statement that ϕ:

Learning Principle $[\phi!]C_G\phi$ ***LP***

Announcing ϕ publicly in a group G makes ϕ common knowledge

LP holds for atomic statements and many others. But it is false in general! E.g., if someone tells you truly "P, but you don't know it", the result is a model where P holds everywhere, and your ignorance has disappeared. Of course, this is Moore's Paradox again, but now in a dynamic epistemic setting. This very update would occur in the model used earlier to strengthen the Consistency of Knowledge principle *CK*.

Thus we have arrived at the second main observation of this paper:

*The 'paradoxical' behaviour of **VT** closely reflects that of **LP**.*

But this analogy also suggests another way of looking at the Verificationist Thesis. Upon reflection, the Learning Principle just seems an overly hasty assertion, and the given counter-example seems very natural. Indeed, announcements of ignorance are made frequently, and they can be very useful. E.g., in well-known puzzles like Muddy Children it is precisely public announcements of ignorance which drive the solution process toward common knowledge of the true state of affairs. Logicians have adapted to this situation, and turned a problem into an object of study. After all, what we see is merely that communication is more interesting than what *LP* thought. For instance, we can investigate what special syntactic forms of assertion *do* become common knowledge upon announcement. And this again suggests more general classifications. Statements of atomic facts are *self-fulfilling*: once announced, their common knowledge results. Moore's statement is *self-refuting*: once announced, its negation becomes common knowledge. But there are also wavering statements in between. (Note 4.) Given the analogy between ***VT*** and ***LP***, one might also develop an analogous enriched verificationist logic, distinguishing different roles for different types of statement.

In the remainder of this paper, we develop this technical theme a bit further. What does knowability or learnability of propositions amount to in a dynamic epistemic setting?

6 Exploring learnable propositions

As in the usual discussion of the Knower paradox, consider the case of a single agent. Suggestions for the case of more agents will follow later. Define a *learnable proposition* ϕ as one whose truth can always become known by announcement of some suitable true formula A. I.e., the following implication is valid:

$$\phi \rightarrow \exists A \ \langle A! \rangle K\phi \qquad \textit{Learnability}$$

The existential modality $\langle A! \rangle K\phi$, dual to $[A!]K\phi$, says that a truthful announcement of A is possible, leading to knowledge of ϕ. The consequent $\exists A \ \langle A! \rangle K\phi$ is shorthand for an infinite disjunction over all formulas A of our language.

Fact Learnability is decidable in *S5*.

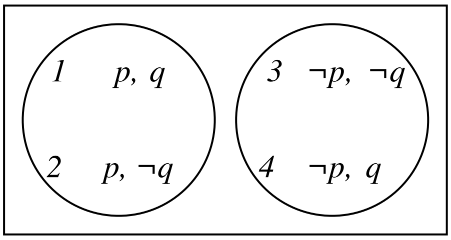
Proof All models of an *S5*-language for a finite set of proposition letters can be enumerated, as the only options that matter are which propositional valuations to represent in the set of worlds. Thus, for each epistemic formula ϕ, we can enumerate all models $\boldsymbol{M}, s \models \phi$. Now, for ϕ to be learnable in the above sense, each of the latter models must have a submodel $\boldsymbol{N}$ containing s with ϕ true in every world of that submodel. But this can be checked effectively in the finite list. **QED**

A stricter form of learnability demands that there be some *finite* set of announcements A one of which must lead to knowledge of ϕ in any given model of ϕ. This learnability by finite cases is equivalent to the above version, however, by the compactness theorem for *S5* – or more simply, by the above enumeration argument. A truly stronger version is the existence of one single assertion A such that the following formula is valid:

$\phi \to \langle A! \rangle K\phi$ *Uniform learnability*

Fact Uniform learnability is stronger than learnability.

Proof Consider the following 4-world model $\boldsymbol{M}$:



Let ϕ be an *S5*-formula which is only true in the following minimal situations

(a) in the pictured model $\boldsymbol{M}$: at the worlds *1, 3*, and no others
(b) in the two models indicated by the ellipses: at both worlds.

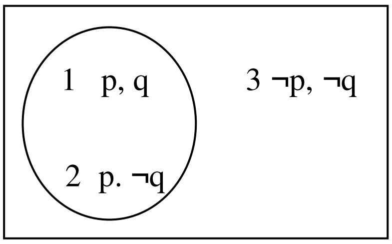
It is easy to write down such a formula explicitly. According to the above description by model enumeration, ϕ is learnable: some update reveals it whenever it is true. But is clear no single formula A does this job uniformly, since the selection of the submodel in $\boldsymbol{M}$ has to depend on which of the two ϕ–worlds is our point of departure. **QED**

Still stronger is the case where announcing ϕ itself produces its knowledge. This is the earlier learning situation of self-fulfilling assertions, restricted to the single-agent case:

$\phi \rightarrow \langle\phi!\rangle K\phi$ Self-fulfillment

Fact Statements can be uniformly learnable without being self-fulfilling.

Proof Consider the following model ***M***, in the same style as the preceding one:



Let ϕ hold only in

(a) in model ***M***: at world *1*,
(b) in ***M***'s oval two-world submodel: at both worlds.

Uniform learnability is satisfied here. In every model ***M****, s* where this formula ϕ holds, announcing the true atomic fact *p* makes ϕ true throughout the resulting model. But, announcing the true statement ϕ itself in the *3*-world initial model *(**M**, 1)* would leave just the *1*-world submodel *p, q,* where ϕ fails by its definition. **QED**

Thus, ***VP*** and ***LP*** are analogous to some extent, but the two putative principles do not coincide. On the way to this outcome, we have seen the flexibility of the dynamic framework in phrasing different versions of learnability. This concludes our account of the single agent setting for epistemic update and learning. Our third observation is that

> ***VT****,* ***VT****** *and* ***LP*** *point toward an interesting general logic of knowledge assertions, announcements, and learning actions.*

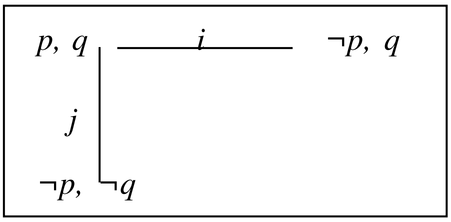
Looking at some possible extensions adds still further detail to this perspective.

7 Refining the issues

Our analysis has looked at the verificationist principle and the Paradox of the Knower in terms of epistemic actions. This does not solve the original problem, but places it in a broader setting of interaction between many agents. In particular, the original formulation of the Paradox of the Knower now becomes a special case in several senses. For a start, even with a single agent, two different senses of learnability emerged: either by means of fixed assertions leading to knowledge, or by context-dependent assertions. But also, the multi-agent setting suggests further refinements. With a single agent, the only candidate for the required knowledge level was $K\phi$. But now one can require knowledge for *other* agents as well: some, or all. E.g., Moore's Paradox disappears with some other agent *2* learning that "*P*, and *1* does not know it", as $K_2(P \ \& \ \neg K_1 P)$ may quite well become true. Also, with groups, we can strengthen the original knowledge condition of ***VT*** as follows:

If ϕ is true, then it is possible that it becomes *common knowledge* in the group.

Perhaps each member finds out part of the truth, and by pooling this information, they arrive at $C_G\phi$. E.g., consider the following model ***M*** with actual world *p, q*:



Announcing *q* will make *j* know the Moore statement that "*p* and *i* does not know it". But this can never become common knowledge in the group *{i, j}*. What can become common knowledge, however, is *p & q*, when *i* announces that *q*, and *j* then says *p*. Many further distinctions can be made in interactive versions of knowability. Moreover, more delicate learning scenarios involving secrecy, hiding, and even misleading, occur in epistemic update logic, with only special subgroups getting complete information about the true facts: cf. Baltag, Moss & Solecki 1998. (Note 5.)

Finally, our perspective also adds a dimension. In recent jargon, the phrase "knowable" suffers from the common disease called '∃–sickness'. This means using an existentially quantified notion in a situation where more explicit information is available, whose logic could be brought out. Common symptoms are frequent uses of "-ility"'s. Compare: provability versus a concrete proof, past tense as 'once upon a time' versus some specific past episode, solvability versus producing an algorithm, winnability of a game versus a concrete winning strategy, etc. Dynamic epistemic logic would cure the sickness in this particular case of 'knowability' by making learning actions and their properties an explicit part of the logic of verificationism, however construed. There may be a price for this expressive power, however, in that general results about learnability for many agents may become harder to formulate and prove (Note 6.)

8 Conclusion

We have shown that knowability of a proposition involves more than consistency of its being known, by placing the Paradox of the Knower in a dynamic setting where learning involves changing the current epistemic model. The Verificationist Thesis then turns out related to the Learning Principle for public announcement. Elaborating this analogy, we found different versions of knowability in an update setting, plus interesting extensions to multi-agent learning. This twist in perspective also reflects a change in mood. Much of the literature on the Fitch Paradox seems concerned with averting a disaster, and saving as large a chunk of verificationism as possible from the clutches of inconsistency. In our perspective, there is no saving ***VT*** – but there is also no such gloom. For in losing a principle, we gain a general logical study of knowledge and learning actions, and their subtle properties. The failure of naïve verificationism just highlights the intriguing ways in which human communication works.

9 Notes

1 The computational complexity of global decidability may go up, as we are now adding a so-called 'universal modality'. Cf. Blackburn, de Rijke & Venema 2001.

2 More sophisticated 'product update' formats, beyond simple elimination, model complex forms of communication mixing public actions and information hiding. This covers much of what happens in games and more realistic communicative settings.

3 As an illustration, one valid principle reduces knowledge after communication to 'relativized knowledge' to be true before it: $[A!]\, K_i\, \phi \leftrightarrow (A \rightarrow K_i\, (A \rightarrow [A!]\, \phi))$.

4 All *universal* modal formulas are self-fulfilling. These are the ones constructed using (negations of) atoms, conjunction, disjunction, K_i and C_G. But there are other self-fulfilling types of statement as well, such as $\langle\rangle p$. A complete syntactic characterization of the self-fulfilling patterns has been one of the open model-theoretic problems of epistemic dynamic logic since the mid 1990s.

5 A multi-agent epistemic language with a common knowledge modality is not like plain *S5*. Simple arguments like that for decidability of single-agent learnability no longer hold, and the same is true for other model-theoretic techniques. We do not know how our earlier results fare in this setting.

6 As a side benefit, our proposal also enriches dynamic epistemic logic. Our observations about single-agent *S5* show that learnability assertions are definable there, and do not add anything new to the language. But now consider public announcements $A!$ in a first-order language, where a formula $A = A(x)$ restricts the individuals to the definable subdomain of those satisfying A. Now, expressive power may increase. E.g., take any strict linear order satisfying the first-order theory of $(N, <)$ which extends beyond N, by adding copies of the integers Z. Its only first-order definable subsets of objects are the finite and the co-finite ones. Now consider the first-order learnability assertion that

'some true announcement makes the following true: the current object *n* is the greatest point, while every object different from zero has a predecessor'.

This can only be true for those objects *n* which lie at some finite distance from the zero of the model. These form an initial copy of *N*, which is not definable in first-order logic. Balder ten Cate has pointed out one might also do this argument in a temporal language, closer to the epistemic modal original.

10 References

Baltag, A., L. Moss and S. Solecki. 1998. The Logic of Public Announcements, Common Knowledge and Private Suspicions. *Proceedings TARK 1998*, 43–56. Los Altos: Morgan Kaufmann Publishers. Updated version, department of cognitive science, Indiana University, Bloomington, and department of computing, Oxford University, 2003.

Benthem, J. van. 2002. One is a Lonely Number, the logic of communicative update. Invited lecture, Colloquium Logicum, Muenster 2002. Report 2003-07, Institute for Logic, language and Information, University of Amsterdam.

Blackburn, P, de Rijke, M. and Y. Venema. 2001. *Modal Logic*. Cambridge: Cambridge University Press.

Brogaard, B. and J. Salerno. 2002. Fitch's Paradox of Knowability. Stanford Electronic Encyclopedia of Philosophy, http://plato.stanford.edu/entries/fitch-paradox/.

Tennant, N. 2002. Victor Vanquished. *Analysis* 62: 135–142.

Wansing, H. 2002. Diamonds are a Philosopher's Best Friends. The Knowability Paradox and Modal Epistemic Relevance Logic. *Journal of Philosophical Logic* 31: 591-612.

University of Amsterdam & Stanford University
Plantage Muidergracht 24, NL-1018 TV Amsterdam
johan@science.uva.nl