The Limits of Knowledge

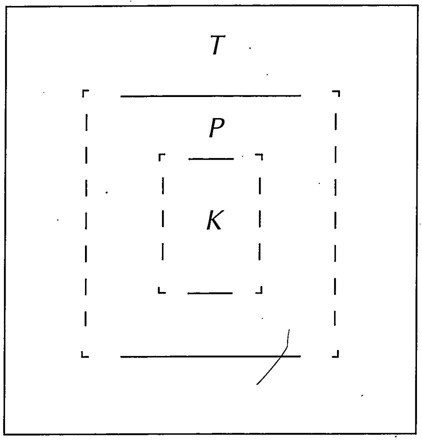
Graham Priest

1. Introduction

Are there limits to knowledge? Well, there are certainly many things that we do not, as a matter of fact, know. We do not know (at the moment) whether Iraq will continue its downward spiral into anarchy. We will know in due course. We do not know how to make the Theory of Relativity and the Theory of Quantum Mechanics consistent with each other. Maybe we will in due course. More interesting is the question of whether there are things that it is not *possible* to know—in any of the many senses of 'possible'. Perhaps there are things that are so difficult, remote, or recondite, that they transcend anything we could find out. If this is the case, there are even limits to what it is possible to know. In what follows, we will look at matters concerning the limits of knowledge more closely.

2. Setting Up the Issue

Let us start by getting the geography of the issue straight. Some notation: I will use Kx as the predicate 'is known', $\Diamond$ and $\Box$ as the usual modal operators of possibility and necessity, and $<.>$ as a name-forming device. Now, let T be the set of truths. The question is how what we know relates to this. There are two relevant subsets. The first is what is known, $K = \{x : Kx\}$. The second is what it is possible to know, $P = \{x : x \in T \wedge \Diamond Kx\}$. (Note that $K<\alpha>$ entails that α is true; but $\Diamond K<\alpha>$ does not—only that α is possibly true.) Since what is known is possibly known (in any sense of possibility), the general relationship between the three sets is as follows:



K is certainly non-empty. Melbourne, for example, is known to be in Australia. $P - K$ is also non-empty. As I have already observed, there are things about the future that we do not know, but will; so *that* knowledge is certainly possible. Similarly, the ancient Greeks did not know that there was a planet beyond Uranus; but it is possible to know this: we do.

The status of $T - P$ is less clear. Is it possible to know anything that is true? Arguably not. We do not know, to the minute, when the last dinosaur died. Nor would it now seem possible for us to find out. But, it might be replied, in principle it is knowable. Had we been around at the time, we could have clocked the event. Maybe, though, there are things that are not knowable even in this sense, some things that are so complex or evidentially remote, that we could not know them, even in principle. One group of people who deny this comprises the verificationists, who identify truth itself with what it is possible, in principle, to verify. Thus, they endorse the principle $\alpha \rightarrow \Diamond K\langle\alpha\rangle$, for the appropriate sense of possibility. There is also a well-known argument that this principle cannot be true in any sense of possibility; that there are some things that are true but cannot be known. The argument was first published by Fitch,[1] and is to the effect that if it is possible to know whatever is true, then everything true is not just *possibly* known, but is *actually* known. That is a *reductio ad absurdum* of the view. *A priori,* there is something highly suspect about the argument, however. Surely one cannot get from the mere fact that it is *possible* to know something to the fact that it is known?

Let us start our detailed investigations by looking at the argument more carefully. Informally, it goes as follows. Suppose that everything true is

1. Fitch (1954). Fitch himself attributes the argument to an anonymous source. It lay dormant for some time, but was published again by Hart and McGinn (1997), whose attention was drawn to it, again, anonymously.

knowable, and suppose for *reductio* that there is something, α, which is true but not known, $\alpha \wedge\neg K<\alpha>$. Then it must be possible to know this, $\Diamond K(\alpha \wedge\neg K<\alpha>)$. By a few straightforward inferences concerning knowledge, it follows that it is possible to both know α and not know it, $\Diamond(K<\alpha>\wedge\neg K<\alpha>)$, which it isn't.

3. The Fitch Argument

3.1 Stage 1: Knowledge

Let me spell out the argument in detail (in natural deduction form), so that we may look at the moves in it more carefully. For the purpose of discussing the argument, and in the cause of simplicity, I will write $K<\alpha>$ as $K\alpha$, effectively turning the predicate K into the more usual operator. (As long as we are not quantifying-in, there is no real difference.)

The part of the proof concerning knowledge goes as follows. Call it Π_1.

$$\dfrac{\dfrac{K(\alpha\wedge\neg K\alpha)\quad \dfrac{[\alpha\wedge\neg K\alpha]}{\alpha}}{K\alpha}\qquad \dfrac{\dfrac{K(\alpha\wedge\neg K\alpha)}{\alpha\wedge\neg K\alpha}}{\neg K\alpha}}{K\alpha\wedge\neg K\alpha}$$

Π_1 uses four inferences:

$$\dfrac{\beta\quad\gamma}{\beta\wedge\gamma}\qquad \dfrac{\beta\wedge\gamma}{\gamma}\qquad \dfrac{K\beta}{\beta}\qquad \dfrac{K\beta\quad \begin{matrix}[\beta]\\ \vdots\\ \gamma\end{matrix}}{K\gamma}$$

In the fourth of these, the column from β to γ represents an argument with premise β (and only β), and conclusion γ. The square brackets represent the fact that the inference discharges β, so that the final argument no longer depends on it.

The first two inferences require no comment; nor does the third: what is known is true. The fourth says that knowledge is closed under entailment. In general terms, this is not true. It might not be known that β entails γ. (This might be because the argument depends on inferences that are not known to hold. Thus, a medieval monk could not infer from 'God exists' to 'God exists or the formalism of quantum mechanics deploys Hilbert spaces'. Or it may be because the argument is just so long that no one has yet put it together; thus, the Peano postulates—we may suppose—entail

Fermat's Last Theorem, but no one realised this till recently.) And if it is not known that β entails γ there is no reason one should know γ, even if one knows β. But the particular application of the rule at issue here is not of this kind. The entailment involved is a very simple and known one. Indeed, the application of the rule could be replaced by the much simpler:

$$\frac{K(\beta \wedge \gamma)}{K\gamma}$$

which is hard to contest.[2]

There is therefore little scope for faulting this part of the argument.

3.2 Stage 2: Possibility

The second part of the argument embeds Π_1 in an argument concerning possibility. This is as follows, where the right-hand column represents Π_1. Call this part Π_2.

$$\dfrac{\dfrac{\alpha\wedge\neg K\alpha}{\Diamond K(\alpha\wedge\neg K\alpha)} \qquad \begin{array}{c}[K(\alpha\wedge\neg K\alpha)]\\ \nabla \\ K\alpha\wedge\neg K\alpha\end{array}}{\Diamond(K\alpha\wedge\neg K\alpha)}$$

Π_2 applies two new rules, which are as follows:

$$\frac{\beta}{\Diamond K\beta} \qquad \frac{\Diamond\beta \quad \begin{array}{c}[\beta]\\ \vdots \\ \gamma\end{array}}{\Diamond\gamma}$$

The first captures the idea that everything true is possibly known, which is what we are assuming. Let us call this the *verificationist inference*. The second says that possibility is closed under entailment. Of course, there are many notions of possibility (e.g., logical, physical, epistemic) but in all of its senses it would seem to be closed under entailment—provided, as we have already, in effect, observed, that the possibility is in-principle possibility. (Suppose that β is known, that β entails γ, and that knowing the entailment is the only way to come to know that γ. Suppose, also, that

2. Though it can be. Connexivist logicians (including some medievals) held that $\beta\wedge\gamma$ does not entail γ—for example, if β is $\neg\gamma$, this simply cancels out the γ. Such a logician could know $\beta\wedge\gamma$, but not believe, and *a fortiori* know, γ. To avoid this kind of problem we can just restrict the class of knowers in question to those who have the normal beliefs about the validity of inferences concerning conjunction—which includes us.

the inference is so complex that it is physically impossible, in practice, to grasp it. Then it is physically possible, in practice, to know β, but not γ. It is possible to know γ only in principle.)

There is little in this stage of the inference that one can balk at, then.

3.3 Stage 3: Contraposition

The third part of the argument embeds Π_2 in an argument deploying negation. This is as follows, where the left-hand column represents Π_2. Call this Π_3.

$$\dfrac{\begin{array}{c}[\alpha\wedge\neg K\alpha]\\ \nabla\\ \Diamond(K\alpha\wedge\neg K\alpha)\end{array}\qquad \neg\Diamond(K\alpha\wedge\neg K\alpha)}{\neg(\alpha\wedge\neg K\alpha)}$$

Π_3 employs one premise and one further rule of inference. The premise is $\neg\Diamond(\beta\wedge\neg\beta)$, or equivalently, given the usual connections between $\Box$ and $\Diamond$:

$$\Box\neg(\beta\wedge\neg\beta)$$

The inference is contraposition:

$$\dfrac{\begin{array}{c}[\beta]\\ \vdots\\ \gamma\end{array}\qquad \neg\gamma}{\neg\beta}$$

The only plausible way to contest these steps is to suppose that contradictions may be true. The rationale for contraposition is that if β delivers something that is not true, γ, it must be false. This rationale collapses if γ can be true despite the truth of $\neg\gamma$. Unsurprisingly, then, the inference fails in many paraconsistent logics (including the one whose semantics I will describe below). Suppose, for example, that the logic contains the Law of Excluded Middle (LEM), $\beta\vee\neg\beta$. Then we have $\gamma\vdash\beta\vee\neg\beta$. Contraposing, $\neg(\beta\vee\neg\beta)\vdash\neg\gamma$, that is (assuming De Morgan Laws), $\beta\wedge\neg\beta\vdash\neg\gamma$—which fails, since γ was arbitrary. This stage of the argument may therefore be broken by appealing to dialetheism.

It might be thought that dialetheism would invalidate the new premise of the argument as well: if contradictions may be true, one might expect $\neg(\beta\wedge\neg\beta)$, and so its necessitation, to fail. Surprising as it might be to those meeting paraconsistency for the first time, it does not. There are many paraconsistent logics where the law holds (including the one whose semantics I will describe below). Of course, any contradiction, $\beta\wedge\neg\beta$, will then generate a secondary contradiction, $(\beta\wedge\neg\beta)\wedge\neg(\beta\wedge\neg\beta)$, but there is nothing in a paraconsistent logic to rule this out.

Actually, the simplest way of avoiding $\neg(\beta\wedge\neg\beta)$ (and so its necessitation) is to appeal, not to truth-value gluts, but to truth-value gaps. If β is neither true nor false, so (given the natural semantics for the connectives) is $\neg(\beta\wedge\neg\beta)$. Appealing to truth-value gaps also invalidates contraposition unless the logic is paraconsistent. If the logic is not paraconsistent, we have $\beta\wedge\neg\beta\vdash\gamma$, and so $\neg\gamma\vdash\neg(\beta\wedge\neg\beta)$, i.e., $\neg\gamma\vdash\beta\vee\neg\beta$, which is not the case if we do not have the LEM.

It might therefore be thought that appealing to truth-value gaps is a way of avoiding the argument without an appeal to gluts. Unfortunately (for the friends of consistency) it is not. As Π_2 shows, $\alpha\wedge\neg K\alpha$ already leads to $\Diamond(K\alpha\wedge\neg K\alpha)$, and thus to the possibility of true contradictions. Moreover, if the logic is not paraconsistent, we have, for an arbitrary β, $K\alpha\wedge\neg K\alpha\vdash\beta$. By the closure of possibility under entailment, we have $\Diamond(K\alpha\wedge\neg K\alpha)\vdash\Diamond\beta$. Given that $\Diamond(K\alpha\wedge\neg K\alpha)$, then, everything is possible—not an enticing conclusion. One way or another, then, true contradictions are required to break this step of the argument.

3.4 Stage 4: Double Negation

There is one final part of the argument. This embeds Π_3 in the argument which actually takes us from α to $K\alpha$. This goes as follows, where the right-hand column represents Π_3.

$$\dfrac{\dfrac{\dfrac{\alpha \quad [\neg K\alpha]}{\alpha\wedge\neg K\alpha} \qquad \begin{array}{c}\neg\Diamond(K\alpha\wedge\neg K\alpha)\\ \nabla \\ \neg(\alpha\wedge\neg K\alpha)\end{array}}{\neg\neg K\alpha}}{K\alpha}$$

This stage of the argument uses contraposition again, discharging $\neg K\alpha$. (And in this application, there is also another assumption in the subproof. As is to be expected, this does nothing to restore validity in a paraconsistent logic. It just makes matters worse.) It uses one further rule, double negation:

$$\frac{\neg\neg\beta}{\beta}$$

Double negation fails in intuitionist logic, which is intimately connected with verificationism. Hence, breaking the argument by denying this step is a very plausible move. If we do, we can get from α only to $\neg\neg K\alpha$, which is not so bad. Well, perhaps. But not really. Π_3 already gives us a proof of $\neg(\alpha\wedge\neg K\alpha)$. And since the argument is formally valid, K can mean anything as long as it satisfies the pertinent rules. So interpret K as 'Frege knew that', and take α to be 'Fermat's Last Theorem is true'. Since Fermat's Last Theorem is true and Frege did not know this, we have $\alpha\wedge\neg K\alpha$, and we are back with contradiction.[3]

4. A Simple Model

So far, we have seen that appealing to dialetheism breaks the Fitch argument against verificationism. Moreover, it is the *only* way that we have found to do so.[4] We can do more than this, however. It can be shown that once contraposition (and only contraposition) is removed from the principles employed, the inference from α to $K\alpha$ is not forthcoming. I demonstrate this with a semantics for a simple paraconsistent modal/epistemic logic.[5]

Interpretations are of the form $<W, R, S, v>$. W is a set of worlds. R is the modal binary accessibility relation. S is the epistemic binary accessibility relation, which is reflexive. v maps every world and propositional parameter to $\{1\}$, $\{0\}$ or $\{1,0\}$ (true, false, both). I write the value of α at w as $v_w(\alpha)$. We also single out one world, ∞, as special. For every propositional parameter, p, $v_\infty(p)=\{1,0\}$; for all $w \in W$, $wR\infty$, and ∞ accesses itself, and only itself, under both R and S. Truth conditions are as follows. For all worlds, w:

$1 \in v_w(\alpha\wedge\beta)$ iff $1 \in v_w(\alpha)$ and $1 \in v_w(\beta)$
$0 \in v_w(\alpha\wedge\beta)$ iff $0 \in v_w(\alpha)$ or $0 \in v_w(\beta)$
$1 \in v_w(\neg\alpha)$ iff $0 \in v_w(\alpha)$
$0 \in v_w(\neg\alpha)$ iff $1 \in v_w(\alpha)$
$1 \in v_w(\Diamond\alpha)$ iff for some w'such that wRw', $1 \in v_{w'}(\alpha)$
$0 \in v_w(\Diamond\alpha)$ iff for all w'such that wRw', $0 \in v_w(\alpha)$
$1 \in v_w(K\alpha)$ iff for all w'such that wSw', $1 \in v_{w'}(\alpha)$
$0 \in v_w(K\alpha)$ iff for some w'such that wSw', $0 \in v_{w'}(\alpha)$

Validity is defined in terms of truth-preservation at all worlds.

3. For a further discussion of the argument in the context of intuitionist logic, see Percival (1990) and the references cited therein.

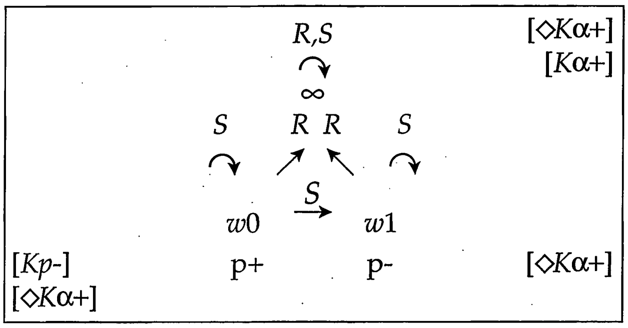
4. Human ingenuity being what it is, there are, of course, others. A number of these are discussed (and rejected) in Williamson (2000), ch. 12. The chapter also contains references to other discussions of the argument in the literature.

5. This is an extension of the propositional paraconsistent logic *LP*. (See Priest [1987], ch. 5.)

Leaving aside the verificationist inference for the moment, it is not difficult to check that the semantics verifies all the inferences involved in the Fitch argument (including the premise $\neg\Diamond(\beta\wedge\neg\beta)$) except contraposition.

For the verificationist inference: it is not difficult to establish that, for all β (and so for everything of the form $K\alpha$) $v_\infty(\beta) = \{1,0\}$. (The proof is by an induction on the formation of sentences.) Hence for all $w \in W$, $1 \in v_w(\Diamond K\alpha)$. The inference $\alpha \vdash \Diamond K\alpha$ is therefore (vacuously) valid.

To finish the job, we just take an interpretation where there are worlds, $w0$ and $w1$, such that $1 \in v_{w0}(p)$, $w0Sw1$, but $1 \notin v_{w1}(p)$. Then $1 \notin v_{w0}(Kp)$. We can depict the simplest interpretation of this kind as follows (+ indicates that a formula holds; - indicates that it fails; square brackets indicate things that hold at worlds, other than what is part of the specification):



5. Enter the Knower

We see that subscribing to dialetheism makes it possible to hold that every truth is possibly known, without this collapsing into the absurdity that every truth is known. But how plausible is dialetheism in this context? Very.

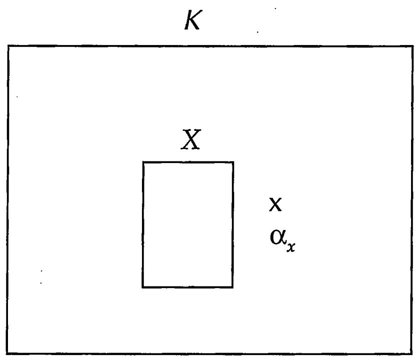
By applying techniques of self-reference, we can construct a sentence, κ, that says of itself that is not known. That is, κ is of the form $\neg K{<}\kappa{>}$. (I now revert to writing K as a predicate. Self-referential constructions require this.) Suppose that $K{<}\kappa{>}$; then κ is true, so $\neg K{<}\kappa{>}$. Hence, $\neg K{<}\kappa{>}$. That is, κ, but we have just demonstrated this, so it known to be true, $K{<}\kappa{>}$. (This is the Knower paradox.[6]) We have demonstrated $K{<}\kappa{>}\wedge\neg K{<}\kappa{>}$. This is therefore necessarily true (in whatever sense of necessity one cares for); *a fortiori*, $\Diamond(K{<}\kappa{>}\wedge\neg K{<}\kappa{>})$. We see, then, that quite independently of the Fitch argument there are sentences of the form required to invalidate the contraposition in Π_3.

6. The connection between the Fitch argument and the Knower was first made by Beall (2000).

We can bring this to bear explicitly on the question of the limits of knowledge as follows. Let $X \subseteq K$. Provided that X has a name, and given appropriate techniques of self-reference, we can form a sentence that says of itself that it is not in X; that is, a sentence, α_x, of the form $\langle \alpha_x \rangle \notin X$. We can show that $\langle \alpha_x \rangle \notin X$ but that $\langle \alpha_x \rangle \in K$ as follows:

$$\begin{aligned} \langle \alpha_x \rangle \in X &\Rightarrow \langle \alpha_x \rangle \in K \\ &\Rightarrow \alpha_x \\ &\Rightarrow \langle \alpha_x \rangle \notin X \end{aligned}$$

Hence, $\langle \alpha_x \rangle \notin X$. But this is α_x, and we have just established this, so it is known to be true; that is, $\langle \alpha_x \rangle \in K$. The situation may be depicted thus:



When X is the empty set, α_x can be located anywhere in $K - X$ (= K). As X gets bigger and bigger, there is less and less space in which α_x can be consistently located; until, at the limit, when X coincides with K there is nowhere consistent for α_x to go. $\langle \alpha_k \rangle \in K \wedge \langle \alpha_k \rangle \notin K$. (This is the Knower paradox. κ is just α_k.) The limit of what is known is dialetheic. That is, there are certain truths that are both within the known and without it.

Exactly the same is true of P. Let $X \subseteq P$. As before, we can construct a sentence, α_x, of the form $\langle \alpha_x \rangle \notin X$.

$$\begin{aligned} \langle \alpha_x \rangle \in X &\Rightarrow \langle \alpha_x \rangle \in P \\ &\Rightarrow \langle \alpha_x \rangle \in T \wedge \Diamond K \langle \alpha_x \rangle \\ &\Rightarrow \alpha_x \\ &\Rightarrow \langle \alpha_x \rangle \notin X \end{aligned}$$

Hence, $\langle \alpha_x \rangle \notin X$. But this is α_x, and we have just established this, so it is true and known to be so, $K\langle \alpha_x \rangle$. *A fortiori*, it is possible to know it,

$\Diamond K<\alpha_x>$. Thus, $<\alpha_x> \in T \wedge \Diamond K<\alpha_x>$. That is, $<\alpha_x> \in P$. Just as with K, when X is small, there is plenty of room for α_x to reside, consistently, outside it but inside P. As X gets bigger and bigger, there is less and less room, until when X is P, a contradiction arises: $<\alpha_p> \in P \wedge <\alpha_p> \notin P$. The boundary of possible knowledge is inconsistent too.

An Inclosure involving a set, Ω, a predicate, ψ, and a function, δ, is a structure satisfying the following conditions:

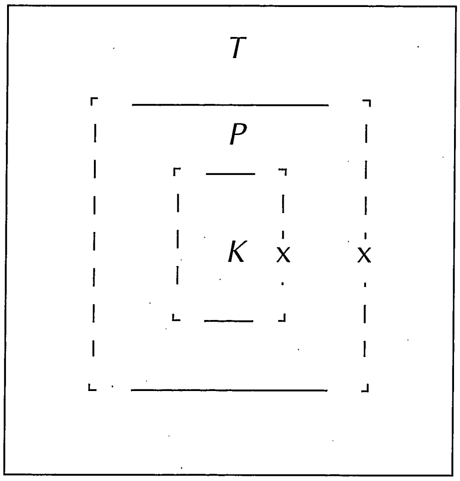
1. $\psi(\Omega)$
2. if $X \subseteq \Omega$ and $\psi(X)$
   (a) $\delta(X) \notin X$ (Transcendence)
   (b) $\delta(X) \in \Omega$ (Closure)

Whenever we have an Inclosure, a contradiction arises at the limit, when $X = \Omega$. For we then have $\delta(\Omega) \notin \Omega \wedge \delta(\Omega) \in \Omega$. All the standard paradoxes of self-reference are limit-paradoxes of this kind.[7]

The two contradictions we have just looked at are of this form. In the first, Ω is K; in the second, Ω is P. In both, $\psi(X)$ is 'X is definable (has a name)', and $\delta(X)$ is α_x. Hence, both are inclosure contradictions.

7. Conclusion

Let us recall our original diagram, and take stock:



K, we know, is non-empty, as is $P - K$. And this is so if $T - P$ is empty, and so the verificationist inference is correct, since the Fitch argument fails.

7. See Priest (1995), Part 3. For the Knower paradox, see 10.2. There, Ω is defined as $\{x : \varphi(x)\}$, where φ is the appropriate predicate.

We have also learned that the boundaries between K and $T - K$, and P and $T - P$ are dialetheic. That is, there is a true sentence, α_k, such that $<\alpha_k> \in K$ and $<\alpha_k> \notin K$, and a true sentence, α_p, such that $<\alpha_p> \in P$ and $<\alpha_p> \notin P$. (This is what the 'x's on the new version of the diagram indicate.) And since α_p is true, $<\alpha_p> \in T \wedge <\alpha_p> \notin P$, so $T - P$ is also non-empty. For all I have said here, this might be its only denizen. It is therefore still possible that it is empty as well, which it is if verificationism is correct. Whether this is so, or whether some other view entailing the emptiness of $T - P$ is correct, are matters for another occasion.

References

Beall, J.C. 2000. Fitch's proof, verificationism, and the knower paradox. *Australasian Journal of Philosophy.* 78: 241–47.

Fitch, F.B. 1963. A logical analysis of some value concepts. *Journal of Symbolic Logic.* 28: 135–42.

Hart, W.D. and C. McGinn. 1976. Knowledge and necessity. *Journal of Philosophical Logic.* 5: 205–8.

Percival, P. 1990. Fitch and intuitionistic knowability. *Analysis.* 50: 182–87.

Priest, G. 1987. *In Contradiction.* Dordrecht: Martinus Nijhoff. 2nd ed. Oxford: Oxford University Press.

———. 1995. *Beyond the Limits of Thought.* Cambridge: Cambridge University Press. 2nd ed., 2002. Oxford: Oxford University Press.

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