Editor's Introduction

The Knowability Paradox

In his seminal paper "A Logical Analysis of Some Value Concepts" (1963; Reprinted, *this volume*), Frederic Fitch articulates an argument that threatens to collapse a number of modal epistemic distinctions. Most directly, it threatens to collapse the existence of fortuitous ignorance into the existence of necessary unknowability. For it shows that there is an unknown truth, only if there is a logically unknowable truth. Fitch called this 'Theorem 5', which usually is represented formally as follows:

> (*Theorem 5*) $\exists p(p \,\&\, \neg Kp) \vdash \exists p(p \,\&\, \neg\Diamond Kp)$,
>
> where p holds a place for sentence letters; $\Diamond$ is normal possibility, read 'it is possible that'; and K is the epistemic operator, 'it is known (by someone [like us] by some means or other at some time) that'.

The theorem rests on tremendously modest modal epistemic principles, which we will turn to shortly. The converse of Theorem 5 is modest as well. So Theorem 5 does the interesting work in erasing the logical difference between there being truths forever unknown and there being truths logically unknowable.

The contrapositive of Theorem 5 is better known as the knowability paradox:

> (*Knowability Paradox*) $\forall p(p \to \Diamond Kp) \vdash \forall p(p \to Kp)$.

If each truth is knowable in principle, then it follows logically that each truth is at some time known. That's the result. It is thought to be paradoxical for a number of related reasons. First, it refutes all too easily interesting brands of anti-realism which are committed to the knowability principle, $\forall p(p \rightarrow \Diamond Kp)$. It refutes them since the knowability principle entails the obviously false omniscience principle, $\forall p(p \rightarrow Kp)$. The knowability principle has been claimed for a number of historic non-realisms, among them Michael Dummett's semantic anti-realism, Hilary Putnam's internal realism, the logical positivisms of the Berlin and Vienna Circles, Peirce's pragmatism, Kant's transcedental idealism, and Berkeley's metaphysical idealism. How strange that the knowability principle and every brand of non-realism that avows it, are only as plausible as the exceedingly implausible, and obviously false, omniscience principle. An extension of Fitch's result, found in Williamson (1992: 68), shows that a traditional strengthening of the knowability principle forecloses on the very distinction between what is possible and what is actual. Roughly, if truth is possible knowledge then possibility is actuality.[1] The paradoxicality is that sophisticated forms of anti-realism could be so easily refuted.

A second reason to regard the proof as paradoxical is that it threatens to erase the logical distinction between the knowability principle and the omniscience principle. More specifically, the proof logically collapses the relatively moderate and plausible claim that each truth *can* be known into the apparently stronger and unbelievable claim that each truth is *in fact*

[1]More carefully Williamson shows this: if necessarily something is true if and only if it is knowable, then necessarily p is possible if and only if p.

known. The claims seem to carry distinct logical commitments, but they do not if Fitch's result is valid.

Fitch's result presupposes the following principles.

Knowing a conjunction requires knowing each of the conjuncts:

(A) $K(p\ \&\ q) \vdash Kp \,\&\, Kq$

Knowing entails truth:

(B) $Kp \vdash p$

Theorems are necessarily true:

(C) If $\vdash p$, then $\Box p$

And, *a necessarily false proposition is impossible*:

(D) $\Box\neg p \vdash \neg\Diamond p$

The proof may be characterized this way:

$$\begin{array}{c l}
\hline K(p\ \&\ \neg Kp) & (1) \\
\hline Kp\ \&\ K\neg Kp & (\mathrm{A}) \\
\hline Kp\ \&\ \neg Kp & (\mathrm{B})\ \&\ \text{trivial logic} \\
\hline \neg K(p\ \&\ \neg Kp) & (1) \\
\hline \Box\neg K(p\ \&\ \neg Kp) & (\mathrm{C}) \\
\hline \neg\Diamond K(p\ \&\ \neg Kp) & (\mathrm{D})
\end{array}$$

At the top of the tree we suppose for reductio that the Fitch-conjunction, $p \,\&\, \neg Kp$, is known. By (A), it follows that each conjunct is known. The third line demonstrates an application of factivity, (B), to the right conjunct of the second line. In the face of the ensuing contradiction, we discharge and deny our only assumption. By necessitation, (C), and then by (D), we conclude with the impossibility of our initial assumption—giving, $\neg\Diamond K(p \,\&\, \neg Kp)$.

Now suppose the knowability principle, $\forall p(p \to \Diamond Kp)$, and take the following instance: $(p \,\&\, \neg Kp) \to \Diamond K(p \,\&\, \neg Kp)$. This together with the above theorem, $\neg\Diamond K(p \,\&\, \neg Kp)$, entails $\neg(p \,\&\, \neg Kp)$, which may be generalized to $\forall p \neg(p \,\&\, \neg Kp)$. The classical equivalent is the omniscience principle, $\forall p(p \to Kp)$. At a glance:

$$\dfrac{\dfrac{\dfrac{\overline{\neg\Diamond K(p \,\&\, \neg Kp)} \qquad \dfrac{\forall p(p \to \Diamond Kp)}{(p \,\&\, \neg Kp) \to \Diamond K(p \,\&\, \neg Kp)}}{\neg(p \,\&\, \neg Kp)}}{\forall p \neg(p \,\&\, \neg Kp)}}{\forall p(p \to Kp)}$$

In sum, if all truths are knowable, then all truths are known:

$$\forall p(p \to \Diamond Kp) \;\; \vdash \;\; \forall p(p \to Kp).$$

The Generalized Paradox

Fitch generalized the knowability result, showing that any operator O that is both factive and closed under conjunction-elimination, generates the following aporia:

$$\begin{array}{c} \underbrace{\forall p(p \to \Diamond Op) \qquad \exists p(p \,\&\, \neg Op)} \\ \vdots \\ \bot \end{array}$$

To prove this Fitch begins with Theorem 1, which holds of any factive operator O that is closed under conjunction-elimination:

(Theorem 1) $\quad \vdash \quad \neg\Diamond O(p \,\&\, \neg Op)$.

Theorem 1 generates the above aporia.[2] Others have noted that it is not just factive, conjunction-distributive operators that validate Theorem 1 and generate the aporia. Belief, for instance, is closed under conjunction-elimantion but is not factive. Yet arguably a belief-instance of Theorem 1 is provable, giving

$\neg\Diamond B(p \,\&\, \neg Bp)$.[3]

In this way the corresponding aporia is generated for the belief operator:

$$\begin{array}{c} \underbrace{\forall p(p \to \Diamond Bp) \qquad \exists p(p \,\&\, \neg Bp)} \\ \vdots \\ \perp \end{array}$$

That is, the plausible notion that any truth could be believed is inconsistent with the truism that some truths are not ever believed. Such proofs about belief avoid unrestricted factivity principles in favor of restricted principles about the transparency of beliefs about one's own beliefs. To take another example, a knowledge-version of the result may be derived without the conjunction-distributivity principle (Williamson: 1993) .

Most generally, then, a Fitch-aporia, or Fitch paradox, is generated for any operator O just when

1. The conjunction $p \,\&\, \neg Op$ is un-O-able: $\forall p \neg\Diamond O(p \,\&\, \neg Op)$;
2. The O-ability principle, $\forall p(p \to \Diamond Op)$, is plausible; and

[2]To see how, substitute $p \,\&\, \neg Op$ for p in $\forall p(p \to \Diamond Op)$. By Theorem 1, it follows that $\neg(p \,\&\, \neg Op)$. This in turn may be may be generalized, giving $\forall p\neg(p \,\&\, \neg Op)$, or equivalently $\neg\exists p(p \,\&\, \neg Op)$.

[3]See, for instance, Wright (****) and Linsky (1986; and *this volume*).

3. Clearly, some truths are un-O-ed: $\exists p(p \mathbin{\&} \neg Op)$.

Operators that seem to generate Fitch-aporias include

It is written truthfully on the board that

Somebody brought it about that

God brought it about that

The laws of nature made it that case that

It is believed that

It is thought that

So, for instance, the paradox of omnipotence may be seen, logically, as a special case of Fitch's paradox. It says, roughly, that God can do anything that is in fact done, but only if God does in fact do everything. Another example: any truth can in principle be thought, but only if every truth is (at some time) thought. This latter result, like the result about belief, requires (in lieu of factivity) a principle that avows some minimal transparency of one's thoughts about one's thoughts.

The Volume, Contributions and Literature

We here turn to some traditional and developing treatments of the paradox. The earliest version of the knowability proof appears in a 1945 referee report for the Journal of Symbolic Logic (Printed here as Essay 1). Its author, Alonzo Church, anonymously conveyed the proof to Fitch. The proof had the effect of undermining a certain definition of 'value' that Fitch was articulating—a definition that is trivialized if there are unknowable truths.

So the proof originates in a context that is very different from the one in which we discuss the proof today. We think of the knowability paradox today either as an all-too-quick refutation of anti-realism or as a logical collapse of apparently distinct philosophical commitments. More on the more recent debate in a moment. Church offers a number of potentially promising ways to block the proof. He is most sympathetic to a rejection of closure principles for knowledge and belief, and a fortiori the principle that knowledge is closed under conjunction-elimination. This is principle (A) in our earlier presentation of the proof. So Church ultimately takes the knowability proof to be invalid—dare I say, paradoxical. However, Church's proposal does not help Fitch, since Fitch is deeply committed to necessary logical connections between the relevant propositional attitudes. Church considers that one may alternatively appeal to Russell's theory of logical types, which would have the effect of blocking special instances of the conjunctive distributivity principle—principle (A). The appeal to types foreshadows Linsky (*this volume*) and Hart (*this volume*). However, Church notes that the type-theoretic approach, like the rejection of closure principles, is antithetical to the goals of Fitch's manuscript.

Fitch had a very different kind of reply in mind. He responds to the referee report with a letter to the editor, in which he restricts the relevant class of true propositions to ones that it is "empirically possible" to know. Fitch's definition of value is thus resuscitated. His restriction strategy foreshadows Neil Tennant (1997), where we find an analogous restriction to the class of true propositions that it is logically possible for somebody to know. We will discuss Tennant's restriction in a moment. It should be noted here that Church was unimpressed with Fitch's restriction strategy, and in a subsequent referee report (also in Essay 1) attempted a version of the knowability proof that respects Fitch's restriction. The debate between Fitch and Church is tracked in Salerno (Essay 3). Church's argument against Fitch's restriction strategy, I argue, is critically flawed.

Section I of the volume is dedicated to this, the early history of the Church-Fitch paradox of knowability. Essay 1 is the pair of referee reports from 1945. They record one side of a dialog between Fitch and the referee regarding the paper submitted by Fitch to JSL. Essay 2 is Fitch's seminal 1963 paper, shaped in no small part by those reports. Fitch's paper has been the logical fuel or foil for the literature on the knowability paradox. Essay 3 is my understanding of the first two essays. It offers an account of why Fitch included the the knowability result in the 1963 paper.

Section II, is about Michael Dummett's semantic anti-realism. The first wave of reactions to Fitch's 1963 paper, including Hart and McGinn (1976), Hart (1979), Mackie (1980), and Routley (1981), had a common theme. They all aimed to use Fitch's proof to discredit various forms of verificationism, the view that all meaningful statements (and so all truths) are knowable.[4] The knowability principle is commonly taken throughout the literature as a particularly clear expression of Hilary Putnam's internal realism (1981) and Michael Dummett's anti-realism (1959b; 1973, and else-

[4]An exception is Walton (1976), whose aim was to draw lessons in the philosophy of religion. For related discussion see Plantinga (1982); Humberstone (1985); MacIntosh (1991); Kvanvig (1995, and 2006); Rea (2000); Wright (2000); Cogburn (2004); Bigelow (2005); and Brogaard and Salerno (2005).

where). The Fitch paper then threatens these forms of anti-realism. Since Williamson (1982) and Rasmussen and Ravinkilde (1982), however, we find various proposals to vindicate at least Dummettian anti-realism. Fitch's reasoning is classically, but not intuitionistically, valid. Specifically, the move from $\neg(p \mathbin{\&} \neg Kp)$ to $p \rightarrow Kp$ (i.e., the final step in our version of the proof) is intuitionistically unacceptable, since it harbors an application of double-negation elimination—i.e., $\neg\neg p \ \vdash \ p$. Leading developments in Dummettian anti-realism favor intuitionistic revisions to classical logic.[5] As such, Dummettian anti-realism is said to evade the unwelcome classical consequences of Fitch. The proposal is further developed in Williamson (1988b; 1990; and 1992).

An objection to the intuitionistic strategy is found in the view that the intuitionistic consequences of Fitch's reasoning are as bad, or almost as bad, as the classical consequences. The objection is developed in Percival (1990).[6] The main intuitionistic consequence is $p \rightarrow \neg\neg Kp$, which says that no truths are forever unknown. Some equivalent formulas include $\neg(p \mathbin{\&} \neg Kp)$, which denies that there are unknown truths, and $\neg Kp \rightarrow \neg p$, which says that anything forever unknown is false, and $\neg(\neg Kp \mathbin{\&} \neg K\neg p)$, which denies that there are any forever undecided statements. The potentially irksome consequence, which is a focus of Wright (1993a: 426–427) and Williamson (1994a), can be put this way. The intuitionistic anti-realist lacks the resources to express the apparent truism that there may be truths that never in fact will be known, formally $\exists p(p \mathbin{\&} \neg Kp)$. That is because the inconsistency derivable

[5] For alternative formulations of the anti-realist argument against classical logic, see Tennant (2000), Salerno (2000), and Wright (2001).

[6] Important further discussion and a reply appears in DeVidi and Solomon (2001).

from the joint acceptance of the knowability principle and $\exists p(p \mathbin{\&} \neg Kp)$ is intuitionistically acceptable.

In Essay 4 Dummett embraces the intuitionistic consequences without regret. The paper defends $p \to \neg\neg Kp$ as the best expression of semantic anti-realism.[7] In a letter to the editor of this volume, Dummett explains that the intuitionistic anti-realist

> as I conceive of him or her, does not think it irksome that the notion "never in fact" cannot be expressed by the use of the intuitionistic logical constants. Rather he or she thinks that the only meaning that can be given to "never" is that expressible by the intuitionistic logical constants. So there is no worry and no frustration. (Letter: September 27, 2005)

For insightful discussion of the intuitionistic use of 'never', see Williamson (1994a).

Incidently, Dummett does not endorse the position articulated in his (2001), which proposes a restriction of the knowability principle to "basic" or atomic sentences. Dummett tells me that he wrote that paper to dispell the myth that Fitch's paradox is an objection to *any* form of anti-realism.

In Essay 5 Stig Rasmussen further investigates and defends Dummett's newly favored knowability principle, $p \to \neg\neg Kp$. The centerpiece of the discussion is the "mapping objection," which points out that the Gödel's 1933 mapping of intuitionistic logic into S4 fails to preserve the original

[7] Cf., DeVidi and Solomon (2001), which offers a defense of this very position on behalf of the Dummettian anti-realist. Dummett embraces the truth of $p \to \neg\neg Kp$ in much earlier work, including (1977: 339 [2000: 236]).

formulation of the knowability principle, and that this fact counts against the original formulation as an expression of intuitionistic anti-realism.

In Essay 6 José Bermúdez argues that the Dummett (2001) position is well-motivated. The position restricts the knowability principle to atomic statements, and defines intuitionistic truth inductively from there. Bermúdez offers an instructive account of Dummett's development in (1990) and (1996). There Dummett attempts to clarify the notion of indefinite extensibility of such concepts as set, natural number, and real number, and argues that only intuitionistic logic can illuminate a proper understanding of the notion. It is argued that if this is correct, then the Dummett (2001) theory of truth is well-motivated, and so, we have a principled solution to the knowability paradox.

Section III is dedicated to paraconsistency and paracompleteness. The paraconsistent approach to the paradox is first suggested in Richard Routley (1981). While considering the liar ("This very statement is not true"), the knower ("This very statement is not known") and Fitch's proposition, $\Diamond K(p \mathbin{\&} \neg Kp)$, Routley entertains, but does not endorse, a uniform treatment:

> What the hardened paraconsistentist says is that [the liar] and $\Diamond K(p \mathbin{\&} \neg Kp)$, though inconsistent, are nonetheless coherent, that this is how things are: some (but not too many) inconsistencies hold true. (1981: 112, n.26)

Routley does not endorse the approach. His actual position is that Fitch's result is valid and that it indicates a *necessary* limitation of human knowledge. Fitch's result shows us that if there is in fact an unknown truth then there is a logically unknowable truth. On the assumption that our actual ignorance is a contingent matter, it is unclear whether the resulting unknowability is contingent or necessary. However, if *necessarily* we actually fail to know some truths, as Routley argues, then it follows by Fitch's main argument and the closure of necessity (over necessary implication) that, necessarily, some truths are unknowable. The passing insight about paraconsistency emerges in the context of Routley's more central discussion of the necessary limits of knowledge.

The paraconsistent approach is first defended in Beall (2000), where it is argued that the knower sentence provides independent evidence that knowledge is inconsistent. For the concept entails that Kp & $\neg Kp$, for some p. Further, it is argued that without a solution to the knower we should accept contradictions of this form and go paraconsistent. To this end Wansing (2002) defines a paraconsistent, constructive, relevant, modal, epistemic logic that evades Fitch.

The section of this volume on paraconsistency constitutes the most recent developments of the paraconsistent treatment of the problem. In Essay 7 Graham Priest develops the Routley/Beall proposal by countenancing the mere possibility of truth-value gluts and appealing to a paraconsistent logic with excluded middle. Beall, in Essay 8, compliments the development by exploring alternatives, some of which avoid the epistemic oddities of Priest's framework. Beall's centerpiece is a semantic framework that is paracomplete, but not paraconsistent, and avoids a commitment even to the mere possibility of truth-value gluts.

Section IV is an exploration of temporal and epistemic analogs of Fitch's reasoning. The strategy is to translate the modalities in the knowability principle into a favored temporal or epistemic logic, and to draw lessons from there about the plausibility of the knowability principle and the result in which it figures. Johan van Benthem (Essay 9) does this by placing the result in a dynamic epistemic setting—a setting in which the truth values of our epistemic attributions vary over time with the performance of various actions, such as announcements. The essay develops more thoroughly van Benthem (2004). John Burgess (Essay 10) translates the Fitch modalities into various Priorian temporal modalities. Each of these two approaches offers, not a rejection of Fitch's proof, but an investigation of the problematic nature of the corresponding knowability principle.

Bernard Linsky (Essay 11) proposes that we block Fitch's result by appealing to a theory of types in our account of epistemic and doxastic reasoning. Interestingly, this is one of the proposals that Alonzo Church (in Essay 1) considers when entertaining objections to the knowability proof. Linsky shows that the theory of types systematically treats a wide variety of contemporary paradoxes of knowledge and belief.

Section V is dedicated to Neil Tennant's Cartesian restriction strategy. Tennant's position is that intuitionistic logic alone will not free anti-realism from the grips of Fitch. His well-discussed proposal is to restrict the knowability principle to *Cartesian propositions*, that is, propositions that it is not provably inconsistent to know. Objections to the proposal include Hand and Kvanvig (1999), Williamson (2000b), and DeVidi and Kenyon (2003). For replies see Tennant (2000a; and 2000b). Further motivation for Tennant's proposal can be found in Jon Cogburn (2004) and Igor Douven (2005).

In Essay 12, Williamson continues the debate, specifically against Tennant (2001a), and renews his pessimism about the prospects for a successful defense of semantic anti-realism (Cf. Williamson 2000b: Chapter 12). Debate with Williamson continues in Tennant (Forthcoming). Essay 13 is Kvanvig's renewed discontent with Tennant's (and any other) restriction to the knowability principle. As he sees it, the real paradoxicality is not that Fitch's result threatens anti-realism, but that it threatens to collapse the very distinction between the existence of unknown truth and the existence of unknowable truth. The section is completed by Essay 14, which is Tennant's current position—a modification of the Cartesian restriction strategy.

Section VI is about modal and mathematical fictionalism. We learn in Brogaard (Essay 15) that modal fictionalism is threatened by Fitch's paradox. Otávio Bueno (Essay 16) evaluates the relevance of Fitch's paradox in the epistemology of mathematics. He argues that the mathematical fictionalist must contend with the unwelcome consequences of Fitch.

Section VII, Knowability Reconsidered, includes papers that reconsider the anti-realist thesis about the knowability of truth. There is a history of attempts to either refute or reformulate anti-realism in reaction of Fitch. I mentioned some refuters earlier. The reformulater is one who rejects, or offers an alternative to, the knowability principle as a characterization of anti-realism. They include Edgington (1985), Melia (1991), Wright (2000), Hand (2003), and Jenkins (2005), among many others. Michael Hand (Essay 17) further develops his 2003 proposal that Dummett's anti-realist is not committed to the knowability principle, owing to the fact that it carelessly blurs semantic conditions about verification procedures with pragmatic conditions about the performance of such procedures. Carrie Jenkins (Essay 18) agrees that the knowability principle fails as an expression of anti-realism. Her own statement of anti-realism (2005) is echoed here, but her primary concern is to take issue with Kvanvig (2006), in which it is argued that the real paradoxicality of Fitch's proof is the modal collapse that occurs in the reasoning from the knowability principle to the omniscience principle. W.D. Hart (Essay 19) takes Fitch's proof to be evidence for realism. He argues that the prospects are not good for a solution coming from the theory of types. Christoph Kelp and Duncan Pritchard (Essay 20) offer some hope for an anti-realism that endorses a justified believability principle in place of the knowability principle. They evaluate the thesis that, for all true propositions, it must be possible to justifiably believe them. An alternative weakening of the knowability principle is proposed by Greg Restall (Essay 21). His principle states that, for every truth p, there is a collection of truths, such that (i) each of them is knowable and (ii) their conjunction is equivalent to p. Restall proves that this formulation evades the paradox, and draws lessons about the operant notion of possibility.

I regret that the volume is incomplete. It includes no extensive discussion of Dorothy Edgington's important 1985 proposal, in which the knowability principle is reformatted as a thesis about the knowability of *actual* truth. Important criticisms are found in Wright (1987 [2nd ed., 1993: 428-432]), Williamson (1987a; 1987b; 2000a, Ch. 12.), and Percival (1991). Developments of Edgington's proposal are found in Rabinowicz and Segerberg (1994), Linström (1997), Rückert (2004), Fara (*manuscript*), and Murzi (*manuscript*). Related proposals that focus on the modal semantics of

Fitch's paradox, include Kvanvig (1995; and 2006), Brogaard and Salerno (2006) and Costa-Leite (2006). These latter three approaches diagnose various modal fallacies. Kvanvig appeals to issues about when we are licensed to substitute into modal contexts. Brogaard and Salerno appeal to Stanley and Szabo's (2000) theory of quantifier domain restriction, according to which there is hidden structure in quantified noun phrases. Costa-Leite appeals to the fusion of Kripke frames—the insight being that knowability is not to be understood compositionally out of one-dimensional possibility and knowledge operators.

Acknowledgements

I am indebted to the contributors for their hard work and their patience with the editorial process. Each of them has furthered my understanding of the knowability proofs and other matters modal epistemic. I am extremely grateful to the graduate students that have served as my research assistants: Amy Broadway (Missouri), Heidi Lockwood (Yale), Julien Murzi (Sheffield), Jonathan Nelson (Saint Louis) and Nick Zavidiuk (Saint Louis). They helped me enormously with various aspects of the project—among them, plundering archives, transcribing, compiling the bibliography, typesetting, and proof-reading for typographical and philosophical errors. Thanks also to my graduate students at Saint Louis University for stimulation, and to those grads at the Goethe University of Frankfurt who attended my seminar on the knowability paradox in May 2006. I thank the editor, Peter Momtchiloff, and assistant editor, Victoria Patton, for their assistance and dedication to the project and to two anonymous OUP readers for their exceedingly helpful comments and suggestions. Many people have helped to improve the volume in one way or another, including Aldo Antonelli, JC Beall, Scott Berman, Jim Bohman, Susan Brower-Toland, John Burgess, David Chalmers, Roy Cook, Judy Crane, Michael Della Rocca, Herbert Enderton, Saul Feferman, Bas van Fraassen, John Greco, Nick Griffin, Michael Hand, Monte Johnson, John Kearns, Jon Kvanvig, Bernard Linsky, Heidi Lockwood, Ruth Barcan Marcus, Robert Meyer, Gualtiero Piccinini, Graham Priest, Krister Segerberg, Wilfried Sieg, Roy Sorensen, Kent Staley, Jim Stone, Eleonore Stump, Neil Tennant, Achille Varzi and Ted Vitali. My wife, colleague, and favorite philosopher, Berit Brogaard, deserves special mention for being a constant source of feedback. My daughter, Rebecca, has been a endless source of sleepless nights but also inspiration, both of which were needed to complete this project. It is to her that I dedicate the volume.

J.S.