Notre Dame Journal of Formal Logic
Volume 49, Number 2, 2008

Fitch's Argument and Typing Knowledge

Alexander Paseau

Abstract Fitch's argument purports to show that if all truths are knowable then all truths are known. The argument exploits the fact that the knowledge predicate or operator is untyped and may thus apply to sentences containing itself. This article outlines a response to Fitch's argument based on the idea that knowledge is typed. The first part of the article outlines the philosophical motivation for the view, comparing it to the motivation behind typing truth. The second, formal part presents a logic in which knowledge is typed and demonstrates that it allows nonlogical truths to be knowable yet unknown.

1 Fitch's Argument

Fitch's simple argument shows that if $p \to \Diamond Kp$ for all p then $p \to Kp$ for all p. Substituting $q \wedge \neg Kq$ into schema $p \to \Diamond Kp$ yields $(q \wedge \neg Kq) \to \Diamond K(q \wedge \neg Kq)$; hence $(q \wedge \neg Kq) \to \Diamond(Kq \wedge K\neg Kq)$ by the distributivity of the knowledge operator over conjunction; hence $(q \wedge \neg Kq) \to \Diamond(Kq \wedge \neg Kq)$ by the factivity of the knowledge operator; and since $Kq \wedge \neg Kq$ is a contradiction, it follows that $\neg(q \wedge \neg Kq)$ or, in other words, $q \to Kq$. If we read the operator 'K' as 'It is known at some time by some subject that', knowability in this sense implies that all truths are known at some time by some subject.

The argument constitutes an apparent reductio of any view committed to the knowability thesis for nonomniscient subjects. It is not specific to knowledge since it applies to any operator with the logical properties appealed to in the proof. Indeed, Fitch's original presentation of the argument (Fitch [7]) was not specifically targeted at knowledge. Various ways of avoiding the reductio have been touted in the literature (see the outline below). I share in the general consensus that at this stage—and much is currently being written on these issues, so the jury is still out—none of the responses has proved sufficiently convincing to have undermined Fitch's argument.

Received July 5, 2006; accepted April 23, 2007; printed March 12, 2008
2000 Mathematics Subject Classification: Primary, 03B42
Keywords: Fitch, knowledge, type theory, knowability, epistemic paradox, epistemic logic
 10.1215/00294527-2008-005

On the contrary, given the diffuse range of responses and the strength of the case for the defense, prosecuted notably by Williamson, there is more than a feeling in the air that Fitch's argument will turn out to be sound.

The literature to date, however, has ignored a natural and independently motivated response. The substitution of sentences of the form $p \wedge \neg Kp$ into the schema $p \to \Diamond Kp$ exploits the fact that the knowledge operator K is untyped. Given that p and $\neg Kp$ and hence $p \wedge \neg Kp$ are of the same type, it follows that the very same operator K which appears in one conjunct can operate on the conjunction itself. This suggests blocking the argument by revising the knowability thesis to each type within a modal typed epistemic logic. Naïvely, one would have thought that typing the knowledge operator in this way would suffice to avoid the conclusion that all truths are known. The discussion below shows this rigorously, by demonstrating that a logic incorporating the knowability thesis does not imply any new sentence of the form $p \to Kp$ where p does not contain a knowledge operator. Before that, we present a philosophical motivation for typing knowledge.

2 Independent Motivation for Typing

Any interesting proposal for blocking Fitch's alleged proof must be supported on independent grounds. We outline the case that the justification for typing knowledge is roughly on a par with that of typing truth.

The conclusion drawn by many who have thought about the Liar and related semantic paradoxes is that the only way to overcome them is to type the notion of truth. As is familiar from the work of Tarski and others, on the typed view there is not one truth predicate but a linear hierarchy of them: 'is true$_0$', 'is true$_1$', ...[1] I shall take it as given that this is a live and important, if not an inevitable, response to the semantic paradoxes and assume broad familiarity with its details.

Suppose now that the same kinds of reasons apply to the case of knowledge; indeed suppose more generally that knowledge should be typed if truth is. It follows that the typing approach has a well-grounded motivation independent of the desire to block Fitch's argument. Following an introduction to this motivation in this section and the next, Section 4 considers the parallels between truth and knowledge.

The semantic paradoxes have analogues in terms of knowledge (for short, epistemic analogues). The Liar's epistemic analogue, for instance, is

(1) This sentence is unknown.

Suppose (1) is known. It is thus true; hence it is unknown. This contradicts the assumption and establishes that (1) cannot be known. From the fact that (1) cannot be known, however, it follows that it is unknown. Hence (1) is true. Since assumptionless proof is a means of gaining knowledge, we therefore know (1). Contradiction.

Likewise there are pairs of sentences that produce contradictions in tandem. For example,

(2) Sentence (3) is known.
(3) Sentence (2) is unknown.

The reasoning here is similar to that above. Further instances of epistemic paradox abound, but there is no need to multiply examples as they are mostly straightforward analogues of familiar semantic ones.

Just as typing the truth predicate blocks Liar-type paradoxes, typing the knowledge predicate blocks epistemic paradoxes. If we assume that there is not one untyped knowledge predicate but rather a typed linear hierarchy of them—'is known$_0$', 'is known$_1$', ..., (see Section 5 for the formal details)—we may rewrite (1) as follows:

(1) This sentence is unknown$_i$.

The first step in the paradox-generating reasoning is now blocked, because the typing rules do not allow us to suppose that (1) is known$_i$, but only that it may be known$_n$, where $n > i$. And there is no contradiction in a sentence's being unknown$_i$ but known$_n$ if $n > i$. Similarly for the other epistemic paradoxes. As is familiar from the case of truth typing, typing knowledge more generally blocks the epistemic paradoxes.

In Sections 5 and 6 we present an epistemic logic that types knowledge and thereby blocks Fitch's argument. All principles of the usual modal and epistemic logic may be taken as principles of this logic, the only difference being the typing of the knowledge operator. Section 3 briefly situates the typing response, and Section 4 examines whether typing knowledge is as well motivated as typing truth by discussing some objections to it.

We officially take knowledge as a predicate applying to sentence names. However, since the distinction between an operator- and a predicate-based treatment of knowledge is not relevant to the informal Sections 3 and 4, the discussion there is cast in terms of either depending on convenience. The formal discussion in Sections 5 and 6 assumes an operator treatment purely for ease. The final Section 7 touches on the importance of this difference and briefly explains why the predicate approach is preferable.

3 Situating the Typing Response

Much has been and continues to be written about Fitch's argument, so it will be useful to briefly situate the typing response within the matrix of existing responses. The aim of this section is not to offer a survey or history or assessment of these responses, but more modestly to give a rough idea of how the typing response stands vis-à-vis some of the leading candidates.[2]

First of all, the typing response rejects the conclusion of Fitch's argument. It is thus distinct from a form of verificationist antirealism that accepts the conclusion but finds it unthreatening, indeed trivial, on its preferred semantics. On this view, to assert p is to lay claim to possession of a verification procedure for p, which—according to this response—implies that p is known.

Second, the typing response is not revisionary with respect to nonepistemic logic. For example, Fitch's argument is not intuitionistically valid, so one might try to block it by adopting intuitionistic rather than classical logic. The typing proposal, however, maintains classical logic.

Third, some authors (e.g., Tennant [19], ch. 8) have responded to the argument by restricting the knowability thesis to certain types of sentences (e.g., consistently knowable ones or "basic" ones). The typing proposal, in contrast, does not restrict knowability: if it succeeds, it vindicates the claim that any truth whatsoever (of any given type) is knowable (at some type or other). A typing response is also distinct from situation-theoretic responses. A situation-theoretic response, generically characterized, allows knowledge in one situation to be about the truth of a proposition in another situation. For instance, Edgington [5] proposes that for any p, the fact *actually p* (i.e., the fact that p holds in the actual world) is knowable in some possible world; in particular, the fact *actually p and no one actually knows p* can be known in some nonactual possible world.[3] More generally, Edgington restricts knowability to propositions relativized to situations (e.g., the actual world), where situations need not be as complete as worlds. The typing response is unlike these situation-theoretic responses in that it does not put a situation-specific gloss on the knowable propositions. It does not reinterpret knowable propositions as being (implicitly or explicitly) about the actual world or some specified situation; on the contrary, it maintains that situation-unrelativized propositions—for example, '$p \wedge \neg K_i p$' for some index i—can be known.[4] For powerful critiques of some responses along these lines, consult Williamson's work (e.g., [20], ch. 12; [21]; [22]).

Finally, the typing response is intended to be neutral between realism and antirealism. Indeed, as I see it, Fitch's argument is a worry not only for many antirealists but also for some realists. You might accept knowability because you see it as a conceptual truth, perhaps one dictated from your theory of meaning, as Dummettians maintain. Or you might accept knowability because you are an optimistic realist. From this perspective, knowability is a metaphysical fact about reality rather than a conceptual fact about the nature of truth. This is a kind of generalization of Gödelian optimism about mathematics, according to which every mathematical truth is provable.[5] What follows is not intended to take sides on the realism-antirealism debate—at least not once classical and modal logic are accepted, which some antirealists might reject for reasons linked to knowability.[6] Indeed I take it to be a strength of the typing response that it cuts across (some) realist and antirealist lines.

4 Objections to Typing: Some Parallels and a Disanalogy

This section runs through a representative (but not exhaustive)[7] sample of objections to typing knowledge and, where appropriate, briefly assesses whether the same, or a similar, problem afflicts the typing approach to truth. (The fourth to sixth objections, which are specific to knowledge, are dealt with directly.) If the truth and knowledge cases are exactly parallel, typing knowledge is no worse a response to the epistemic paradoxes than typing truth is to the semantic ones. If so, that is motivation enough for the approach since, as is widely appreciated, the typing response to the semantic paradoxes is a leading contender (and arguably the only game in town once strengthened or revenge versions of the paradoxes are taken into account). As we will see, the two cases are mostly parallel, but there is one significant difference between them. A brief indication of the standard defense to the objections from the typed perspective for both truth and knowledge is also given. I assume familiarity with the case of truth throughout.

It is worth emphasizing that my aim is not to answer the objections to typing, nor more generally to advocate either typing. I am no partisan, and it is certainly beyond the scope of this paper to adjudicate between untyped and typed truth theories. My goal is rather to explore the parallels between the cases of truth and knowledge, and by doing so to weigh the typing response's pros and cons so as to come to an initial assessment of its potential.

First Objection There is little or nothing in ordinary language to suggest typing the knowledge operator. At best, then, typing knowledge is poorly motivated; at worst, it cannot be a correct analysis of our concept of knowledge, which is univocal.

Parallel The parallel with typing truth is very close. There is equally little or no evidence in ordinary language for typing truth.[8]

General line of response The paradoxes (semantic or epistemic) effectively force the typing on us. The resulting account is not an analysis of our conceptual scheme; it is rather a theoretically serviceable replacement for it. Its virtue lies in the fact that it honors most of the linguistic phenomena while satisfying crucial theoretical desiderata. In the case of truth, for instance, it issues in a consistent theory that preserves classical semantics and logic, standard truth principles (e.g., disquotation) and allows for self-reference.

Second Objection It seems that there is an underlying concept of knowledge that knowledge_0, $\text{knowledge}_1, \ldots,$ are all instances of. However, not only does the theory not posit such a concept, it deems it illicit. What makes knowledge_0, $\text{knowledge}_1, \ldots,$ all *knowledge* predicates (or operators)?

Parallel Exactly the same objection can be raised about the predicates is true_0, is $\text{true}_1, \ldots$. What makes all these predicates truth predicates? Is there not an overarching, untyped, unrestricted concept of truth, truth simpliciter, of which these are all instances?

General line of response Type theorists bite the bullet and concede that there is no consistent overarching untyped concept (of truth or knowledge). On this view, what the semantic paradoxes show is precisely that there cannot be a single all-encompassing truth predicate or operator. Likewise, the epistemic paradoxes show that there cannot be a single all-encompassing knowledge predicate or operator. The various truth and knowledge predicates or operators only have their inferential behavior in common (see Section 5 for some inferential rules and axioms).

Untyped approaches to the semantic paradoxes, such as those inspired by Kripke ([11], see especially p. 710), derive the predicates 'true_i' as instances of the unrestricted truth predicate conjoined with the predicate 'is a sentence of language_i'. They apparently respect the intuition, illicit on the typed approach, that 'truth_i' is relational, that is, that it is 'truth-in-language-$\mathcal{L}_i$', with '$\mathcal{L}_i$' a variable place. That would certainly be an advantage of untyped over typed theories if they could have it. However, defenders of typing maintain that strengthened or revenge versions of the Liar (e.g., 'this sentence is not true in any language $\mathcal{L}_i$') show that purportedly untyped theories ultimately cannot avoid typing. Kripke himself seemed to recognize this[9] and considered his untyped theory to offer models plausible only "as models of natural language at a stage *before* we reflect on the generation process associated with the concept of truth, the stage which continues in the daily life of nonphilosophical speakers" ([11], p. 714, n. 34, my emphasis) rather than a broader theory encompassing reflective semantic discourse as well. It should be emphasized that these strengthenings of the Liar are not a minor defect of untyped accounts; they are the root problem at issue. Moreover, any solution to the semantic paradoxes must either severely restrict classical truth theory[10]—for example, standard disquotational principles and inferences—or reject classical logic (or both). Arguably, then, despite the revival of interest in untyped theories in the past thirty-odd years, typing remains unavoidable if one is to get around all forms of semantical paradox. So if successful the parity argument that epistemic typing is just as well-motivated as truth typing really does constitute a strong case for epistemic typing, and it constitutes a correspondingly strong case for the response to Fitch's argument based on it.

Third Objection (to motivating blocking Fitch's argument by typing knowledge) If we treat the subscripts as variables, Fitch's argument rearises. Consider the operator K_ω defined as the infinite disjunction $\vee_{1 \leq i < \omega} K_i$, that is, $K_1 \vee K_2 \vee K_3 \vee \cdots$. Substituting $(q \wedge \neg K_\omega q)$ into the schema $p \rightarrow \Diamond K_\omega p$ yields a contradiction, as before. This objection has been raised by Williamson ([20], p. 281) among others. More generally, define K_+ as the disjunction over all ordinals α of K_α. Assuming $p \rightarrow \Diamond K_+ p$ and allowing K_+ to operate on sentences containing itself, $p \rightarrow K_+ p$ follows as before.

Parallel Exactly the same objection can be raised about truth. If we treat the subscripts as variables, a strengthened or explicit version of the Liar reappears: 'this sentence is not true$_i$ for any finite i' or more generally 'this sentence is not true$_\alpha$ for any ordinal α'.

General line of response The hierarchy may consistently be described from two standpoints: either completely in some external metalanguage or partially from within the hierarchy itself. A theorist operating in a metalanguage distinct from the hierarchy of typed languages may, of course, treat the indices of the truth or knowledge predicates as variables. But speakers of those languages cannot; only partial and incomplete descriptions of the hierarchy may be given from the speakers' standpoint. The base language $\mathcal{L}_0$ is characterized as having no semantic vocabulary; the next language $\mathcal{L}_1$ contains $\mathcal{L}_0$ plus the predicate 'is true$_0$', which applies to all truths$_0$, and so on. In brief, the hierarchy of languages is elliptically described as the hierarchy generated by iterating this process indefinitely. The idea of this indefinite iteration may be intuitively conveyed, but on pain of inconsistency it cannot be precisely defined from within the perspective of the hierarchy. Speakers of these typed languages are therefore not vulnerable to the objection.

On the typing approach, then, what the paradoxes show is that there is no acceptable language which gives its own semantic theory, or to put it another way, that the metalanguage of any sufficiently expressive object language must be stronger than it. There is no ultimate global language that contains the whole hierarchy of predicates 'is true$_\alpha$' or 'is known$_\alpha$', including predicates applicable to that language itself. In particular, the predicates 'not known$_i$ for any finite i' or 'not known$_\alpha$ for any ordinal α' are not available to speakers of one of the languages in the hierarchy. Untyped theories of truth must, of course, also accept this conclusion or else give up some highly intuitive principles about truth (or classical logic).

Fourth Objection We should not proliferate typings. Even if in light of the semantic paradoxes we accept that the best reconstruction of our conceptual scheme is to type truth, there is no need for a further epistemic typing.

Fifth Objection The knowledge typing follows from the truth typing because the concept of truth features in the analysis of knowledge. Hence a further knowledge typing is otiose.

Responses to the Fourth and Fifth Objections The fourth objection is mistaken and the diametrically opposed so-called fifth objection is really no objection at all. In response to the fourth objection, the epistemic paradoxes show that knowledge has to be typed along with truth. Typing truth alone does not solve the paradox generated by sentences such as 'this sentence is unknown'. In response to the fifth objection, one might well argue that it follows from the analytic connection between knowledge and truth—that knowledge of p entails the truth of p—that knowledge has to be typed if truth is. But this so-called objection is really an argument *for* typing knowledge rather than against it. Whether the typing of knowledge is primitive or derivative on the truth typing does not affect the conclusion that knowledge is also typed.

Sixth Objection Doesn't the contradiction[11] associated with sentences such as (1)—'This sentence is unknown'—depend on us actually carrying out the reasoning given in Section 2 (for example, to get to the subconclusion that (1) is known)? And are such epistemic paradoxes not therefore more akin to pragmatic paradoxes like 'I am not speaking' or 'I am not here' or 'p but I don't believe that p' than they are to semantic paradoxes such as the Liar?[12]

Response Epistemic paradoxes such as (1) are very different from pragmatic paradoxes. For one thing, our interest here is primarily in logically omniscient subjects, as standard in epistemic logic. In particular, the axiom 'if $\vdash \varphi$ then $\vdash K\varphi$' is assumed (more precisely, in Section 5 we assume this axiom's typed versions). Now this axiom is all that is required to infer from the inconsistency of (1)'s falsehood that (1) is known. *Our*—my and your, my reader's—reasoning about, discussion of or contemplation of (1) does not in any way affect or impact on whether a logically omniscient subject knows (1). The discussion in Section 2 revealed that (1) is known (and also not known!) by a logically omniscient subject at all times; it did not create that knowledge.[13]

The disanalogy between epistemic and pragmatic paradoxes remains even if we interpret the knowledge operator as concerned with limited subjects such as you and me rather than logically omniscient ones. On this interpretation of the operator, the paradox generated by (1) does, of course, depend on facts about actual subjects' reasoning (and the 'we' in our exposition of the paradox in Section 2 can be taken literally). But that does not make it a pragmatic paradox, since it is due not to assertion (nor to the thought it expresses being entertained) but rather to the fact that knowledge can be attained by reasoning. Epistemic concepts naturally depend on subjects' mental states and reasoning, but that does not make the paradoxes based on them pragmatic.

To appreciate this difference in diagnosis, observe that the key symptoms of pragmatic paradox are missing in the epistemic case. In the pragmatic case, the paradoxical sentence is no longer contradictory if it is not uttered (e.g., consider 'I am not speaking'), but (1) remains contradictory whether or not anyone utters it. In the pragmatic case, the paradoxical sentence is typically no longer contradictory if a different subject utters a sentence expressing the same proposition (e.g., 'you are not speaking'), but (1) remains contradictory whoever utters a sentence expressing that same proposition. In the pragmatic case, the paradoxical sentence is typically no longer contradictory if the tense is changed (e.g., 'I was not here'), but (1) is tenseless, in the sense that it is understood as 'This sentence never was, is not, and never will be known'. Finally, epistemic paradoxes pass other tests that pragmatic paradoxes fail. For instance, as Hintikka has noticed ([10], p. 51), pragmatic paradoxes are no longer contradictory if prefixed by a word like 'suppose' (e.g., 'Suppose that p but I don't believe that p'), but (1) remains contradictory even with this prefix. Epistemic and pragmatic paradoxes are therefore birds of a very different feather.

Of course, epistemic paradoxes are not exactly analogous to semantic ones, as is to be expected since knowledge is a different notion from truth. For instance, there is nothing akin to the T-schema for knowledge (one half of the T-analogue for a knowledge operator, $p \to Kp$, is precisely the absurd consequence to which the knowability thesis is reduced by Fitch's argument). But the parallels between them run deep, and, as we have seen, typed theories of truth and typed theories of knowledge share many common features.

Seventh Objection It is difficult to understand the difference between is 'known$_m$' and 'is known$_n$' for $m \neq n$. What exactly is the difference? What does it correspond to? It seems to be a formal distinction without a substantive difference. (See Williamson ([20], p. 281) for a version of this objection.)

Putative parallel The same objection can be raised about the predicates 'is true$_m$' and is 'true$_n$'. It equally appears to be a formal distinction without a substantive difference.

A disanalogy It is at this juncture that the most significant disanalogy between knowledge and truth typing appears if the former is constructed so as to block Fitch's reasoning. Consider sentences drawn from theories $\mathrm{T}_0 \subset \mathrm{T}_1 \subset \mathrm{T}_2 \subset \cdots \subset \mathrm{T_n} \subset \cdots$, where a successor theory T_{i+1} extends its predecessor T_i by containing a theory of truth$_i$ for T_i. It is natural to accept the following principle governing the truth hierarchy:[14]

(Truth Minimality) Suppose p is a sentence of type i. Then for any $n > i$,

$$\mathrm{True}_n(p) \to \mathrm{True}_{i+1}(p).$$

In other words, a sentence is true at some type only if it is true at the lowest type at which it is eligible for truth. This respects the informal intuition that, as we might put it, there is no more to truth at level i than eligibility for truth at level i and truth simpliciter. This is an intuition which, of course, cannot be formally expressed in the typed theory, since the point of that theory, as indicated earlier, is precisely that there is no room for the notion of truth simpliciter. It is one, however, which any speaker of an apparently untyped language, such as English, would like to see any typed replacement language(s) preserve. A theory of truth types that respects minimality stays closer to our pretheoretic understanding of truth than one that rejects it.

The epistemic version of minimality is

(Knowledge Minimality) Suppose p is a sentence of type i. Then for any $n > i$,

$$\mathrm{Known}_n(p) \to \mathrm{Known}_{i+1}(p).$$

Knowledge minimality is compatible with a typed epistemic logic (excluding knowability) of the kind we shall investigate in Sections 5 and 6. But it is evidently incompatible with knowability (Section 5 contains the straightforward explanation). Thus anyone who blocks Fitch's argument by typing knowledge must face the question of whether there are independent grounds for rejecting minimality.

This seems to be the point at which the parallel between the knowledge and truth hierarchy is at its weakest. In both cases, resolving the paradoxes leads to a counterintuitive type hierarchy. But in the case of knowledge, the first-pass counterintuitiveness seems to run somewhat deeper than in the case of truth. In both cases, our grip on the difference between a predicate of level n and its analogous predicate of level m seems to be formal rather than substantive, meaning that it *seems* to depend merely on the formal character of the sentence in question. (Putting it in propositional rather than sentential terms, it seems to depend on the nature of the proposition's constituents.) This is not something that speakers raised on untyped languages can grasp intuitively—although the hope or expectation of typing's proponents is, as Quine put it, that "a time will come when truth locutions without implicit subscripts, or like safeguards, will really sound as nonsensical as the antinomies show them to be" ([15], p. 9). At present, however, it is only slightly better than being told that there are, say, two different types of truth predicate, one applicable to sentences containing words with the letter 'e' in them, the other to all other sentences or, in propositional terms, that, say, propositions containing binary concepts/properties may be true$_1$ but that those containing ternary concepts/properties may only be true$_2$.[15] Still, on this view we know that the reason for typing, namely, the paradoxes, must lead to this consequence. But within the constraints of adopting a typed solution to the paradoxes, respecting minimality seems, on the surface at least, preferable. The only difference between the predicates 'is true$_n$' and 'is true$_m$' respectively licensed by the paradoxes is that they apply to different types of sentences. Should this not also extend to knowledge?

Can the disanalogy be defended? The difference in verdict on minimality (if knowability is preserved) does indeed constitute an important disanalogy between knowledge and truth typing. What remains to be seen is whether the disanalogy can be justified. I shall not settle the matter definitively here; I shall merely try to indicate what I take to be the typing response's best shot at justifying the disanalogy.

We should first of all be clear that it really is illicit to understand truth$_i$ as truth simpliciter for a sentence of language $\mathcal{L}_i$, or as we might informally express it,

$$\text{Truth}_i = \text{Truth} + \text{Language } i.$$

This equation may be a useful psychological crutch for speakers of untyped languages who find themselves forced, on theoretical grounds, to operate with typed languages. But it is clearly a false picture of what is going on, as it employs what on the typed approach is an incoherent notion of truth simpliciter, and it also illicitly treats the truth index as a variable. For exactly the same reason it is illicit to understand knowledge$_i$ as

$$\text{Knowledge}_i = \text{Knowledge} + \text{Language } i.$$

This observation cuts off an important argument for Knowledge Minimality. Arguably, the main reason that we expect minimality to hold is because we subscribe at some level to these informal equations. But since the equations are inconsistent with the typed perspective, this avenue to justifying Knowledge Minimality is closed off. This potential justification for minimality is simply inconsistent with the typed perspective.

We should be clear about the import of this point, which is not about plausibility itself but about what underpins it. I do not deny that Truth Minimality is what speakers brought up on untyped languages expect to see in a typed theory. Nor do I deny that this extends to Knowledge Minimality, which is also prima facie plausible to those same speakers. The point is simply that from the typed perspective the obvious justification for minimality is not available: one cannot argue that a principle is a natural component of some theory if doing so invokes assumptions inconsistent with the latter! Still, that does not amount to a positive argument for rejecting Knowledge Minimality; it is merely an argument against the main idea behind its appeal.

A second point is that knowability seems to be a generally held pretheoretical conviction prior to reflection on Fitch's argument. It is plausible that any truth whatsoever could in some (perhaps very remote) circumstances be known—this is after all why Fitch's argument has been called a paradox by several authors.[16] In a typed theory, however, minimality and knowability are incompatible; hence we have to jettison (at least) one of these two pretheoretic commitments. Rejecting knowability is thus also counterintuitive.

This argument against minimality is controversial. Some will complain that knowability is only appealing to a verificationist or to a certain kind of idealist or to some other sort of philosopher currently regarded as unsavory. Others will complain that there is an intuitive and nowadays generally endorsed view that there are truths that are simply unknowable—"forever beyond our ken" as they say. Both these complaints miss the point, however. As regards the first, I take knowability to be a tenet of optimistic realism, as explained earlier. Also, some philosophical antirealists (e.g., Dummett) who have espoused knowability cannot be dismissed as crude verificationists.[17] Moreover, knowability's plausibility is not only confined to those who are philosophically *parti pris*. It has some appeal to nonphilosophers as well, especially when it is explained to them that the modality in question transcends physical possibility.[18] The second complaint is compatible with the fact that knowability is also plausible. Simply because two principles are inconsistent does not mean that they cannot both be plausible. (Materialism and dualism in the philosophy of mind provide a classic example: they are both, to some degree—the degree depending on the individual or society—intuitive views, yet they are inconsistent.) The question is which of the two should be dropped. This is a question that will be settled on broadly theoretical grounds, with some weight given to initial plausibility. The two complaints are therefore unwarranted.

Nevertheless, as I see it, though knowability does have intuitive pull,[19] Knowledge Minimality is also highly intuitive. Perhaps that is only because it is so hard to disengage oneself from the grip of the equation '$\text{Knowledge}_i = \text{Knowledge} + \text{Language } i$'. That may well be; even so, the typing response would be on much stronger ground if it came with a positive justification for the rejection of minimality. At the very least, the disanalogy with the truth case remains, and something has to be said by way of explaining the difference between them. Moreover, unless we understand what it would be for evidence to lead to knowledge_n of p but not to knowledge_{i+1} (where $n > i + 1$ and p is of type i), the idea that evidence for any statement could always be presented so that it may become known seems to support minimality as well. So the fact that knowability has some intuitive plausibility, perhaps no less than minimality, is indeed something to be reckoned with in this debate. But for the reasons just given, it does not much boost the typing response's dialectical position.

It seems that the best way for the defender of the typing response to proceed is to argue that knowledge typing corresponds to the manner in which the proposition has come to be known. In other words, the typing is one of epistemic access rather than (just) content.[20] In a Fitch case, for example, that some proposition p of type 0 is unknown$_1$ but known$_2$ might be explained on the grounds that it has been derived by reasoning involving propositions of type 1 (as well as of type 0), in this instance by deriving $K_2 p$ from $K_2(p \wedge \neg K_1 p)$.

Now why should knowledge typing depend on knowledge route? Two observations are in order. First, addressing this question would move us into firmly philosophical territory. We should distinguish what we may call a logic of knowledge from a philosophical theory of knowledge, just as we may distinguish a logic of truth from a philosophical theory of truth. In the case of truth, this distinction (though not necessarily these labels) is familiar: for example, the typing account is a logic of truth, whereas, say, the correspondence theory or the coherence theory or the minimalist theory are philosophical theories of truth. Now to ask why knowledge typing should depend on epistemic route is to ask for a philosophical theory of knowledge. The answer to that question will form an important part of the overall typing response, but it will be supplementary to a logic of knowledge. Just as I have not touched on whether the typing approach to truth should be coupled with a correspondence or coherence or minimalist or some other philosophical theory of truth, in the same way I shall not investigate which philosophical theory should underpin the typing approach to knowledge. This is not to deny that there are important consonances and discords between logics of knowledge and philosophical theories of knowledge—indeed there are, as there are in the case of truth—nor that the feasibility of the approach might ultimately depend on it. But it would be to open a new, entirely philosophical, chapter, distinct from the present logical one though of course related to it.

Having said that, we can preempt the discussion a little, since there appears to be a very natural line of thought, once typing is in place, for knowledge of p's type being constitutive on the epistemic route to p. Compare two cases of knowledge of the same empirical fact, say that my computer's keyboard is black: (i) perception of the keyboard, and (ii) inferential knowledge, obtained by modus ponens from the known propositions that q and that if q then my keyboard is black, where q has knowledge type $i > 0$. Part of what makes the perceptual belief that my keyboard is black knowledge is that it was arrived at in a certain way, namely, visual perception. Part of what makes the inferential belief that my keyboard is black knowledge is that it was arrived at in a certain way, namely, deduction via modus ponens from some premises, one of which is a fact of knowledge type i. My knowledge of p is constituted in the perception case by K_0-facts; my knowledge of p is constituted in the inference case partly by K_i-facts. It is thus natural to see the knowledge acquired in the first case as being of type 1 and the knowledge acquired in the second case as being of type $i + 1$. More generally, it seems to be constitutive of any belief's status as knowledge that it was acquired in some particular way, involving knowledge facts of some particular type(s).

The idea, in sum, is that when a subject knows that p, her knowledge of p is constituted by various facts, some of which are facts about her epistemic route to p. And from a typed perspective, it is natural to assume that one cannot know anything at a lower level than the facts which constitute that knowledge.

That whether a subject knows p should depend on the manner in which her belief that p was acquired is hardly a radical thought. Quite the contrary, it is familiar and philosophically orthodox. To give just one example, a standard form of reliabilism in epistemology takes knowledge to be true belief *acquired* by a reliable method. Contrast the case of truth, where there is no more to the truth of the sentence expressing p than the fact that p. Since there is nothing else for the truth of p to depend on other than p itself—unlike knowledge that p—it is only appropriate that *True(p)* may always be of type $i + 1$ (where $\text{type}(p) = i$). This seems to be the relevant difference between truth and knowledge case, which explains why minimality applies to the former but not the latter.

Conclusion The remarks in this section constitute a first assessment of the typing response. Our aim has been not to settle that case, but to open and explore it. If the case against minimality ultimately turns out to be a strong one—if what I described as the typing response's best shot succeeds—then Fitch's argument is blocked by a motivation not significantly weaker than the motivation for typing truth. Clearly, more needs to be said to fill out the response, in particular, its philosophical dimension. For instance, the sense in which a knowledge fact is constituted by other facts, for example, belief-acquisition facts, requires further explication.[21] And, of course, our discussion has been driven by considerations apparently specific to knowledge, so it may not happily generalize to instances of Fitch's general argument based on other epistemic operators (e.g., true conceivability). At least, that remains to be investigated.[22] Moving on to the formal discussion, I demonstrate that typing knowledge allows nonlogical truths to be unknown yet knowable.

5 A Modal Typed Epistemic Logic

5.1 Language The language is that of propositional logic supplemented with a propositional operator $\Box$ and propositional operators $\mathrm{K_i}$ for $1 \leq \mathrm{i} < \omega$. A formula φ in which the highest index of the $\mathrm{K_i}$-operators appearing in it is m is said to have type m, written $\tau(\varphi) = \mathrm{m}$; concatenations of propositional letters, propositional connectives and $\Box$, that is, sentences φ such that $\tau(\varphi) = 0$, are said to be nonepistemic or K-free. The usual formation rules apply for the propositional connectives and the modal operator $\Box$. The formula $\mathrm{K_j}\varphi$ is well-formed if and only if $\mathrm{j} > \tau(\varphi)$.

5.2 Deductive system The first component of the deductive system is any rule-based system for propositional logic that makes it sound and complete with respect to the usual semantics. The second component consists of the usual logical rules for the modal logic **S5**, to wit the K-axioms $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$; the necessitation rule if $\vdash \varphi$ then $\vdash \Box\varphi$; and the **S5** axioms $\Box\varphi \to \varphi, \Box\varphi \to \Box\Box\varphi$ and $\Diamond\varphi \to \Box\Diamond\varphi$. The third component is a typed version of the epistemic logic KT. It contains the epistemic version of necessitation, if $\vdash \varphi$ then $\vdash \mathrm{K_n}\varphi$ for every $\mathrm{n} > \tau(\varphi)$; the closure axioms $\mathrm{K_n}(\varphi \to \psi) \to (\mathrm{K_n}\varphi \to \mathrm{K_n}\psi)$ for every $\mathrm{n} > \max\{\tau(\varphi), \tau(\psi)\}$; and the factivity axioms $\mathrm{K_n}\varphi \to \varphi$ for every $\mathrm{n} > \tau(\varphi)$. The fourth component is the strongest possible version of the knowability thesis: all instances of the schema $\varphi \to \Diamond\mathrm{K_n}\varphi$ hold for $\mathrm{n} > \tau(\varphi)$. (In Section 7 we explain that this is much stronger than what a proponent of knowability need be committed to: in fact he is only committed to the claim that $\varphi \to \Diamond\mathrm{K_n}\varphi$ holds for some $\mathrm{n} > \tau(\varphi)$, and similarly for necessitation.) These four components make up the logic $\mathrm{L^K}$.

5.3 Fitch's argument Fitch's argument is blocked straightforwardly by the typing. Consider $A \wedge \neg K_{\tau(A)+j}A$ where $j \geq 1$. From $K_{\tau(A)+j+m}(A \wedge \neg K_{\tau(A)+j}A)$ where $m \geq 1$, it follows that $K_{\tau(A)+j+m}(A) \wedge K_{\tau(A)+j+m}(\neg K_{\tau(A)+j}A)$; hence $K_{\tau(A)+j+m}(A) \wedge \neg K_{\tau(A)+j}(A)$, but this is no contradiction.

We now show that no Fitch-like argument can be run in L^K. In fact we prove a more general conservativeness result: if A is nonepistemic and $L^K \vdash A \rightarrow K_n A$ then either A or $\neg A$ is a theorem of **S5**. In other words, even the strongest possible version of knowability incorporated in L^K fails to imply that any nonepistemic nontheorem is known, never mind that all truths are known.[23]

5.4 The model We describe a simple possible worlds model[24] for this logic in which any instance of $\varphi \rightarrow K_n\varphi$ where φ is nonepistemic and neither φ nor $\neg\varphi$ is an **S5** theorem is false at some world. It follows that no such instance of $\varphi \rightarrow K_n\varphi$ is valid under a semantics of this kind. Since the logic is sound for the semantics, no such instance of $\varphi \rightarrow K_n\varphi$ is provable in L^K.

Here is the idea behind the model, which is constructed from a model for **S5**. The new domain is made up of ordered pairs of worlds and stages. The truth-value assignments of nonepistemic sentences at all world-stages are the same as those at the corresponding world in the old model. At the initial stage $\langle\,\rangle$, no statements that are not **S5**-theorems are known at any world. For any given world, any stage s where s is a finite sequence of positive integers, and any positive m, there is a next stage in which all the true sentences at that world (of type $< m$) are known$_m$. This is achieved by defining an accessibility relation corresponding to K_m such that any world at stage $s\hat{\ }m$ only accesses itself.

More formally, consider a model $\langle W, R, Val\rangle$ for the nonepistemic (type 0) language whose frame validates **S5**, R being the accessibility relation for the worlds in W. This model could, for instance, be the canonical model for **S5**. Using $\langle W, R, Val\rangle$, we now construct a new model $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, Val^*\rangle$, where R^* is the accessibility relation for $\Box$ and S^*_n is the accessibility relation for K_n ($n \geq 1$). The new model $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, Val^*\rangle$ validates L^K, as we show.

The new set of worlds W^* consists of ordered pairs of members of W and finite sequences of members of $\mathbb{N}^+$ (positive integers). $\langle\,\rangle$ is the null sequence, and $s\hat{\ }n$ is the concatenation of the sequence s with the positive integer n. Intuitively, the world $\langle w, m\rangle$ is a world at which all truths are known$_m$, the world $\langle w, \langle m, n\rangle\rangle$ is a world at which all truths are known$_n$ and anything true at both $\langle w, m\rangle$ and $\langle w, \langle m, n\rangle\rangle$ is known$_m$, and so on. We say that m is a member of s if m is an element of the sequence s.

The new modal accessibility relation R^* is defined by

$$\langle w, s\rangle R^* \langle x, t\rangle \quad \text{iff} \quad wRx.$$

Note that R^* is reflexive, since R is, and more generally that $\langle w, s\rangle R^* \langle w, t\rangle$ for any sequences s and t.[25] The epistemic accessibility relations S^*_n, one for each $n \in \mathbb{N}^+$, are specified by induction on the length of the sequence s (where s is a finite sequence of positive integers, that is, $s \in (\mathbb{N}^+)^{<\aleph_0}$) as follows:

$$\begin{array}{lll} \langle w, \langle\,\rangle\rangle S^*_n \langle x, t\rangle & & \text{for every } w, x, t \\ \langle w, \langle s\hat{\ }m\rangle\rangle S^*_n \langle x, t\rangle & \text{iff} & \langle x, t\rangle = \langle w, \langle s\hat{\ }m\rangle\rangle \\ & & \textbf{or } m \neq n \text{ and } \langle w, s\rangle S^*_n \langle x, t\rangle. \end{array}$$

Note that S^*_n is reflexive. We say that $\langle w, s\rangle S^*_n$-accesses $\langle x, t\rangle$ if $\langle w, s\rangle S^*_n\langle x, t\rangle$; ditto for R^*.

The new valuation Val* is defined straightforwardly:

$$\text{for atomic A, Val}^*(A, \langle w, s\rangle) = \text{Val}(A, w).$$

The clauses for the sentential connectives are the usual ones. The clauses for A and K_nA are likewise as expected:

$$\text{Val}^*(\Box A, \langle w, s\rangle) = 1 \text{ iff Val}^*(A, \langle x, t\rangle) = 1 \text{ for all x, t such that } \langle w, s\rangle R^*\langle x, t\rangle;$$

$$\text{Val}^*(K_nA, \langle w, s\rangle) = 1 \text{ iff Val}^*(A, \langle x, t\rangle) = 1 \text{ for all x, t such that } \langle w, s\rangle S^*_n\langle x, t\rangle.$$

The proof that $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, \text{Val}^*\rangle$ is a model of L^K is now straightforward, as is the proof that any instance of $\varphi \to K_n\varphi$ where A is nonepistemic and neither A nor ¬A is an **S5** theorem is false at some world.[26]

Lemma 5.1 *If* A *is* K_m*-free, then* $\text{Val}^*(A, \langle w, s\hat{}m\rangle) = \text{Val}^*(A, \langle w, s\rangle)$.

Proof The result is entirely expected, given the informal idea that the world $\langle w, s\hat{}m\rangle$ extends $\langle w, s\rangle$ by forcing all the true sentences at $\langle w, s\hat{}m\rangle$ of type $< m$ to become known$_m$ there. Note that the lemma implies that if A is nonepistemic then $\text{Val}^*(A, \langle w, s\hat{}m\rangle) = \text{Val}^*(A, \langle w, s\rangle)$ for any m, and hence that $\text{Val}^*(A, \langle w, s\rangle) = \text{Val}^*(A, \langle w, \langle\,\rangle\rangle)$ for nonepistemic A. We prove the lemma by induction on the complexity of A. The basis case follows from the fact that $\text{Val}^*(A, \langle w, s\rangle) = \text{Val}(A, w)$ for all atomic A. The induction step for truth-functors is routine, as is the step corresponding to □, which does not even require the inductive hypothesis since by definition $\langle w, s\hat{}m\rangle$ and $\langle w, s\rangle R^*$-access the same worlds. The remaining case is the one in which $A = K_nB$ for some B. Observe first that since A is K_m-free, $m \neq n$. Now if $\text{Val}^*(K_nB, \langle w, s\hat{}m\rangle) = 0$, then B is false at some W^*-world S^*_n-accessed by $\langle w, s\hat{}m\rangle$. If this is $\langle w, s\hat{}m\rangle$ itself, then $\text{Val}^*(B, \langle w, s\hat{}m\rangle) = 0$, and hence by the induction hypothesis, which applies since B too is K_m-free, $\text{Val}^*(B, \langle w, s\rangle) = 0$. Therefore, by the reflexivity of S^*_n, $\text{Val}^*(K_nB, \langle w, s\rangle) = 0$. Alternatively, if the W^*-world that $\langle w, s\hat{}m\rangle S^*_n$-accesses at which B is false is not $\langle w, s\hat{}m\rangle$ itself then by the definition of S^*_n, $\langle w, s\rangle$ also S^*_n-accesses that world. Hence once more $\text{Val}^*(K_nB, \langle w, s\rangle) = 0$. Putting this together, if $\text{Val}^*(K_nB, \langle w, s\hat{}m\rangle) = 0$, then $\text{Val}^*(K_nB, \langle w, s\rangle) = 0$. Suppose now that $\text{Val}^*(K_nB, \langle w, s\hat{}m\rangle) = 1$. As $m \neq n$, any world S^*_n-accessed by $\langle w, s\rangle$ is also S^*_n-accessed by $\langle w, s\hat{}m\rangle$. Thus $\text{Val}^*(K_nB, \langle w, s\rangle) = 1$. □

Lemma 5.2 *The model satisfies the strongest possible version of knowability; that is, for all* $n > \tau(A)$*, and* $\langle w, s\rangle \in W^*$, $\text{Val}^*(A \to \Diamond K_nA, \langle w, s\rangle) = 1$.

Proof Suppose $\text{Val}^*(A, \langle w, s\rangle) = 1$ and $n > \tau(A)$ so that A is K_n-free. By Lemma 5.1, $\text{Val}^*(A, \langle w, s\hat{}n\rangle) = 1$. Since $\langle w, s\hat{}n\rangle$ only S^*_n-accesses itself, it follows that $\text{Val}^*(K_nA, \langle w, s\hat{}n\rangle) = 1$, and hence that $\text{Val}^*(\Diamond K_nA, \langle w, s\rangle) = 1$ since $\langle w, s\rangle R^*$-accesses $\langle w, s\hat{}n\rangle$. □

Lemma 5.3 *Suppose that* $A \to K_nA$ *is valid (true at all worlds) in the model* $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, \text{Val}^*\rangle$*. Then either* A *or* ¬A *is valid.*

Proof If there is no world at which A is true, we are done since ¬A is then valid. So suppose A is true at some world $\langle w, s\rangle$ and hence K_nA is also true there since $A \to K_nA$ is valid. Now K_nA is well-formed (Val* only takes well-formed sentences of the language), so we know that A is K_n-free. If we can show that both A and K_nA are true at the world ⟨w, t⟩ where the sequence s directly extends the sequence t (i.e., s = t^m for some m), we will be done, since by repeated applications of the same reasoning it will follow that K_nA is true at ⟨w, ⟨ ⟩⟩, and hence that A is true at all worlds in the model. (Recall that ⟨w, ⟨ ⟩⟩S_n^*-accesses all worlds; if s = ⟨ ⟩ in the first place we are immediately done.) So suppose that A and K_nA are true at ⟨w, s⟩ = ⟨w, t^m⟩. If m ≠ n, then the worlds that ⟨w, t⟩S_n^*-accesses are also S_n^*-accessed by ⟨w, t^m⟩, and so since Val*(K_nA, ⟨w, t^m⟩) = 1 it follows that Val*(K_nA, ⟨w, t⟩) = 1. By the reflexivity of S_n^*, Val*(A, ⟨w, t⟩) = 1 as well. Hence both A and K_nA are true at ⟨w, t⟩. If alternatively m = n, then by Lemma 5.1, which applies since A is K_n-free, Val*(A, ⟨w, t⟩) = Val*(A, ⟨w, t^m⟩) = 1. Now since A → K_nA is true at all worlds, K_nA must also be true at ⟨w, t⟩, since A is. So whatever m might be, A and K_nA are true at ⟨w, t⟩ if they are both true at ⟨w, t^m⟩. □

Our conservativeness result now follows straightforwardly. Suppose that A is part of the nonepistemic fragment of the language (that is, τ(A) = 0 and so A is K_n-free for all n). If A → K_nA were a theorem of L^K, it would be valid in all of the logic's models, and hence by Lemma 5.3 one of A or ¬A would be valid in each of L^K's models. In particular, either A or ¬A would be valid in the *-extension of the canonical model for **S5**, which by inspection and Lemma 5.2 is a model of L^K. Hence either A or ¬A is valid in the canonical model for **S5** (recall that A is nonepistemic and use Lemma 5.1), and so either A or ¬A is a theorem of **S5**.[27]

The model thus establishes the conservativeness result for **S5**. Rerunning the argument for appropriately similar modal logics (e.g., **S4**) establishes it for them as well.

6 Variations

Though the model establishes the required result, it is far from "natural"; that is, it is far from capturing the intuitive picture behind knowability. For example, if we take the canonical model for **S5** as our starting model ⟨W, R, Val⟩, one artificial feature of the new model ⟨W*, R*, $\{S_n^*\}_{n\in\mathbb{N}^+}$, Val*⟩ is that at any world in the model knowledge$_m$ of many, many sentences stands or falls together. In particular, for any two literals A and B, if Val*(A, ⟨w, s⟩) = Val*(B, ⟨w, s⟩) then Val*(K_nA, ⟨w, s⟩) = Val*(K_nB, ⟨w, s⟩), as is easily proved by induction on the length of s. (Literals are atomic propositions of the language and negations thereof.) Briefly, if s = t^m, where m ≠ n, then by Lemma 5.1, Val*(K_nA, ⟨w, s⟩) = Val*(K_nA, ⟨w, t⟩), and also by Lemma 5.1 Val*(A, ⟨w, s⟩) = Val*(A, ⟨w, t⟩); ditto for K_nB and B, and now use the inductive hypothesis. Alternatively, if s = t^m, where m = n, then Val*(K_nA, ⟨w, s⟩) = Val*(A, ⟨w, s⟩) = Val*(B, ⟨w, s⟩) = Val*(K_nB, ⟨w, s⟩). Finally, if s = ⟨ ⟩, the result follows immediately. Yet if we think of the worlds in the model as stages in the buildup of knowledge, it is unrealistic to suppose that all true literals become known$_m$ if one of them does.[28]

A more natural model would therefore allow that only some true literals may become known$_m$ at a world ⟨w, ⟨s^m⟩⟩. One suggestion for implementing this idea is, given the original model ⟨W, R, Val⟩ which we can take to be the canonical model for **S5**, to define a new set of worlds W** as the set of ordered pairs of members of W and finite sequences of members of $2^{\aleph_0} \times \mathbb{N}^+$, that is, members of $(2^{\aleph_0} \times \mathbb{N}^+)^{<\aleph_0}$. ⟨ ⟩ is the null sequence, and s^⟨α, n⟩ is the concatenation of the sequence s with the ordered pair $\langle \alpha, n \rangle$, where $\alpha \in 2^{\aleph_0}$ and $n \in \mathbb{N}^+$. The ordinal index α here picks out the αth element of some fixed well-ordering of the power set of the set of literals of the language, which is of size $2^{\aleph_0}$. Intuitively, the world $\langle w, \langle \alpha, m \rangle \rangle$ is the world at which the members of some subset indexed by α of the literals of the language true at $\langle w, \langle\,\rangle \rangle$ are all forced to be known$_m$ there. For example, the subset of literals indexed by α true at $\langle w, \langle\,\rangle \rangle$ is stipulated to be known$_m$ at $\langle w, \langle \alpha, m \rangle \rangle$ and the subset of literals indexed by β true at $\langle w, \langle\,\rangle \rangle$ is stipulated to be known$_n$ at $\langle w, \langle \langle \alpha, m \rangle, \langle \beta, n \rangle \rangle \rangle$ (where $m \neq n$), and so on; α might be, say, $\{p_0, p_2, p_4, p_6, \ldots\}$ and β might be $\{p_1, p_2, \neg p_3\}$. Actually, this is only true of cases in which the subset indexed by α is satisfiable (i.e., does not contain an atomic proposition and its negation); if this subset is unsatisfiable (e.g., α is $\{p_0, \neg p_0\}$) then all sentences true at $\langle w, \langle s\hat{}\langle \alpha, m \rangle \rangle$ are known$_m$.

The only difference between the model $\langle W^*, R^*, \{S^*_n\}_{n \in \mathbb{N}^+}, Val^* \rangle$ and $\langle W^{**}, R^{**}, \{S^{**}_n\}_{n \in \mathbb{N}^+}, Val^{**} \rangle$ lies in the clauses for S^{**}_n:

$$\langle w, \langle\,\rangle \rangle S^{**}_n \langle x, t \rangle \qquad \text{for every } w, x, t,$$

$$\langle w, s\hat{}\langle \alpha, m \rangle \rangle S^{**}_n \langle x, t \rangle \quad \text{iff} \quad \langle x, t \rangle = \langle w, s\hat{}\langle \alpha, m \rangle \rangle$$
$$\textbf{or } m \neq n \text{ and } \langle w, s \rangle S^{**}_n \langle x, t \rangle$$
$$\textbf{or } n = m \text{ and } t = \langle\,\rangle \text{ and } x \cap w = \alpha.$$ [29]

If α is satisfiable, the third disjunct of the recursive clause enables all literals in α true at $\langle w, \langle\,\rangle \rangle$ (or w in the original model) to be known$_m$ at $\langle w, s\hat{}\langle \alpha, m \rangle \rangle$ and forces all literals not in α to be unknown$_m$ at $\langle w, s\hat{}\langle \alpha, m \rangle \rangle$. If α is unsatisfiable then all literals true at $\langle w, \langle\,\rangle \rangle$ (or w in the original model) are known$_m$ at $\langle w, s\hat{}\langle \alpha, m \rangle \rangle$. In fact, both claims generalize from literals to the nonepistemic fragment of the language, as is easily verified using an analogue of Lemma 5.1 (see next).

The proof of the conservativeness result is closely analogous to that in Section 5. The direct analogue of Lemma 5.1 in the **-model is, if A is K_m-free, then $Val^{**}(A, \langle w, s\hat{}\langle \alpha, m \rangle \rangle) = Val^{**}(A, \langle w, s \rangle)$ for all A, w, s, α, and m. The only (slight) difference in the proof concerns the case in which $A = K_n B$ for some B. Note as before that if A is K_m-free then $m \neq n$. If $Val^{**}(K_n B, \langle w, s\hat{}\langle \alpha, m \rangle) = 0$, then B is false at some W^{**}-world S^{**}_n-accessed by $\langle w, s\hat{}\langle \alpha, m \rangle$. If this is $\langle w, s\hat{}\langle \alpha, m \rangle$ itself, then $Val^{**}(B, \langle w, s\hat{}\langle \alpha, m \rangle) = 0$; hence by the induction hypothesis, since B too is K_m-free, $Val^{**}(B, \langle w, s \rangle) = 0$, and so by the reflexivity of S^{**}_n, $Val^{**}(K_n B, \langle w, s \rangle) = 0$. Alternatively, if the W^{**}-world that $\langle w, s\hat{}\langle \alpha, m \rangle S^{**}_n$-accesses at which B is false is not $\langle w, s\hat{}\langle \alpha, m \rangle$ itself then by the definition of S^{**}_n, either $\langle w, s \rangle$ also S^{**}_n-accesses that world or $n = m$. But $m \neq n$; hence the latter possibility is excluded, and so once more $Val^{**}(K_n B, \langle w, s \rangle) = 0$. In sum, if $Val^{**}(K_n B, \langle w, s\hat{}\langle \alpha, m \rangle) = 0$ then $Val^{**}(K_n B, \langle w, s \rangle) = 0$. Suppose now that $Val^{**}(K_n B, \langle w, s\hat{}\langle \alpha, m \rangle) = 1$. As $m \neq n$, any world S^{**}_n-accessed by $\langle w, s \rangle$ is also S^{**}_n-accessed by $\langle w, s\hat{}\langle \alpha, m \rangle \rangle$. It follows that $Val^{**}(K_n B, \langle w, s \rangle) = 1$. This establishes the analogue of Lemma 5.1 in the model $\langle W^{**}, R^{**}, \{S^{**}_n\}_{n \in \mathbb{N}^+}, Val^{**} \rangle$. For Lemma 5.2, suppose that A is true at $\langle w, s \rangle$ and let π be the set of all literals of the language.[30] Then for any $m > \tau(A)$, using the analogue of Lemma 5.1, A is true at $\langle w, s\hat{}\langle \pi, m \rangle \rangle$ and hence so is $K_m A$. The analogue of Lemma 5.3 follows as before using the analogue of Lemma 5.1.

One interesting feature of the model $\langle W^{**}, R^{**}, \{S^{**}_n\}_{n \in \mathbb{N}^+}, Val^{**} \rangle$ is that it is not monotonic with respect to knowledge of literals, since knowledge$_m$ of a literal at a world $\langle w, s \rangle$ need not be preserved at a world $\langle w, s\hat{}t \rangle$ (where s and t are now members of $(2^{\aleph_0} \times \mathbb{N}^+)^{<\aleph_0}$, not $(\mathbb{N}^+)^{<\aleph_0}$). For example, if, say, $\alpha = \{p_0\}$, $\beta = \{p_1\}$, and $\mathrm{Val}(p_0, w) = \mathrm{Val}(p_1, w) = 1$ then $\mathrm{Val}^{**}(K_1 p_0, \langle w, \langle \alpha, 1\rangle\rangle) = 1$, but $\mathrm{Val}^{**}(K_1 p_0, \langle w, \langle\langle \alpha, 1\rangle, \langle \beta, 1\rangle\rangle\rangle) = 0$. In this sense, the model allows for "forgetting" one's knowledge of literals and of nonepistemic sentences more generally (so long as they are not theorems of **S5**). (Of course, the model $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, \mathrm{Val}^*\rangle$ is not in general monotonic either. But using the original Lemma 5.1, it is easy to show that if $\mathrm{Val}^*(K_m\varphi, \langle w, s\rangle) = 1$ then $\mathrm{Val}^*(K_m\varphi, \langle w, s\hat{}t\rangle) = 1$ for any sequence t and nonepistemic φ.)

One might also wish to tweak the model $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, \mathrm{Val}^*\rangle$ so as to validate a thesis of ascending knowledge $K_{\tau(\varphi)+m}\varphi \rightarrow K_{\tau(\varphi)+m+j}\varphi$ ($j \geq 1, m \geq 1$). The thesis states that anything known at some knowledge type is known at all higher knowledge types. Note that a version of the KK-thesis, $K_{\tau(\varphi)+m}\varphi \rightarrow K_{\tau(\varphi)+m+j} K_{\tau(\varphi)+m}\varphi$ ($j \geq 1, m \geq 1$), implies the ascending knowledge thesis by the factivity of $K_{\tau(\varphi)+m}$. The obvious way of tweaking the model $\langle W^*, R^*, \{S^*_n\}_{n\in\mathbb{N}^+}, \mathrm{Val}^*\rangle$ would be to modify the definition of S^*_n-accessibility as follows:

$\langle w, \langle\,\rangle\rangle S^{***}_n \langle x, t\rangle$		for every w, x, t
$\langle w, s^\wedge m\rangle\rangle S^{***}_n \langle x, t\rangle$	iff	$\langle x, t\rangle = \langle w, s^\wedge m\rangle\rangle$
		or $n < m$ and $\langle w, s\rangle S^{***}_n \langle x, t\rangle$.

Thus true propositions are known$_n$ at $\langle w, s^\wedge m\rangle$ for any $n \geq m$.

An analogue of Lemma 5.1 may be proved in this model; namely, if A is K_i-free for all $i \geq m$, then $\mathrm{Val}^{***}(A, \langle w, s\hat{}m\rangle) = \mathrm{Val}^{***}(A, \langle w, s\rangle)$. We prove the case in which $A = K_n B$ for some B. Since A is K_i-free for all $i \geq m, n < m$. If $\mathrm{Val}^{***}(K_n B, \langle w, s\hat{}m\rangle) = 0$, then B is false at some W^{***}-world S^{***}_n-accessed by $\langle w, s^\wedge m\rangle$. If this is $\langle w, s\hat{}m\rangle$ itself, then $\mathrm{Val}^{***}(B, \langle w, s\rangle) = \mathrm{Val}^{***}(B, \langle w, s\hat{}m\rangle) = 0$ by the induction hypothesis, and by the reflexivity of S^{***}_n, $\mathrm{Val}^{***}(K_n B, \langle w, s\rangle) = 0$. If the W^{***}-world that $\langle w, s^\wedge m\rangle S^{***}_n$-accesses at which B is false is not $\langle w, s^\wedge m\rangle$ itself then by the definition of S^{***}_n, $\langle w, s\rangle$ also S^{***}_n-accesses it (since $n < m$). Hence once more $\mathrm{Val}^{***}(K_n B, \langle w, s\rangle) = 0$. Suppose, on the other hand, that $\mathrm{Val}^{***}(K_n B, \langle w, s\hat{}m\rangle) = 1$. Since $n < m$, any world S^{***}_n-accessed by $\langle w, s\rangle$ is also S^{***}_n-accessed by $\langle w, s\hat{}m\rangle$. It follows that $\mathrm{Val}^{***}(K_n B, \langle w, s\rangle) = 1$.

The analogue of Lemma 5.2 follows straightforwardly, but Lemma 5.3's analogue no longer holds. For instance, $K_m B \rightarrow K_{m+i} K_m B$ is valid for $i \geq 1$ yet neither $K_m B$ nor $\neg K_m B$ need be valid. This is a straightforward consequence of incorporating the ascending knowledge thesis into the logic. What is true, however, is the following: Suppose that A is nonepistemic and that $A \rightarrow K_n A$ is valid in the model $\langle W^{***}, R^{***}, \{S^{***}_n\}_{n\in\mathbb{N}^+}, \mathrm{Val}^{***}\rangle$; then either A or $\neg A$ is valid. For the proof, suppose that A is true at some world $\langle w, s\rangle$ and hence that $K_n A$ is also true there. Let $s = t^\wedge m$ (if $s = \langle\,\rangle$, we are done). Since $\tau(A) = 0 < m$, then by the analogue of Lemma 5.1 since A is K_i-free for every $i \geq m$, it follows that $\mathrm{Val}^{***}(A, \langle w, t\rangle) = \mathrm{Val}^{***}(A, \langle w, t^\wedge m\rangle) = 1$. Now since $A \rightarrow K_n A$ is true at all worlds, $K_n A$ must also be true at $\langle w, t\rangle$, since A is. Repeating the reasoning, it follows that $K_n A$ is true at $\langle w, \langle\,\rangle\rangle$ and hence that A is true at all worlds. The conservativeness result follows as before.

7 Knowability, But Not the Strongest Possible Kind

In Sections 5 and 6, we represented knowability in a typed logic in its strongest possible form, namely, as the schema $\varphi \rightarrow \Diamond K_n\varphi$ for *all* $n > \tau(\varphi)$. However, there is (at least) another way of cashing out knowability in a typed logic, namely, as the claim that $\varphi \to \Diamond K_n\varphi$ for *some* $n > \tau(\varphi)$.[31] Indeed the stronger version goes much further than the basic tenet of knowability that any proposition is knowable, since it adds to it a specification of the types at which knowledge is achieved.[32] This extra commitment clearly goes beyond knowability itself, though it remains an open question whether it goes beyond the motivation(s) behind it.

We used the strongest possible version of knowability earlier for two reasons. First, for simplicity of discussion, since it does not significantly affect the philosophical issues. Second, for ease of formal exposition in Sections 5 and 6, since a system that satisfies knowability simpliciter, namely, the schema $\varphi \to \Diamond K_n\varphi$ for some $n > \tau(\varphi)$, is not a logic proper. Adding this thesis to a typed modal epistemic logic does not give rise to a deductive system in the strict sense, since the set of theorems cannot be recursively generated without extra-logical information (in particular, the information of which type knowledge of φ is derivable at). In contrast, the maximally strong version of knowability, that $\varphi \to \Diamond K_n\varphi$ for all $n > \tau(\varphi)$, gives rise to a logic that can be recursively axiomatized. By proving the conservativeness result (and hence relative consistency) of the latter, we a fortiori proved the conservativeness (and hence relative consistency) of the former. But these two reasons should not blur the fact that the defender of knowability in a typed context is committed to no more than the schema $\varphi \to \Diamond K_n\varphi$ for some $n > \tau(\varphi)$.[33]

In fact, if cast within a predicate- (rather than operator-) based modal epistemic logic, the maximally strong version of knowability can be shown to lead to contradiction.[34] For any given K_n, let γ_n be a sentence such that $\gamma_n \leftrightarrow \neg\Diamond\ulcorner K_n\ulcorner\gamma_n\urcorner\urcorner$, with $\tau(\gamma_n) < n$. By standard diagonalization techniques, such a γ_n exists for any K_n. But together with the fact that $\gamma_n \to \Diamond\ulcorner K_n\ulcorner\gamma_n\urcorner\urcorner$, that is, the strongest version of knowability, it follows that $\neg\gamma_n$, and hence that γ_n is not known$_n$. Since this fact is provable, it is necessary that $\neg K_n\ulcorner\gamma_n\urcorner$, which by the equivalence of γ_n with $\neg\Diamond\ulcorner K_n\ulcorner\gamma_n\urcorner\urcorner$ yields γ_n itself. The proof of this contradiction, of course, crucially relies on the maximally strong version of knowability: without it all that can be assumed is that if γ_n then $\Diamond\ulcorner K_m\ulcorner\gamma_n\urcorner\urcorner$ for some $m \geq n > \tau(\gamma_n)$, and then the contradiction no longer follows.

Noting that the diagonalization procedure works because knowledge is treated as a predicate K(x) rather than a sentential operator Kp, some will doubtless respond that the trouble lies in taking knowledge to be a predicate rather than an operator. This has been a standard response to constructions of this kind following Montague [13], whose claim that there is no satisfactory, paradox-eschewing predicate-based modal logic is often taken at face value. Though it would take us too far afield to examine the merits of predicate versus operator systems of epistemic and alethic modality, it is worth giving two brief reasons why this reply is unpromising.

In the first place, a predicate-based epistemic logic is required to sustain the motivation for typing based on the epistemic paradoxes, as in this article's first half. There are operator-based systems in which the epistemic paradoxes are reproducible (Grim [9]) if, for example, propositional quantification is allowed, but such systems do not avoid the inconsistency generated by the maximally strong version of knowability precisely because of their paradox-generating resources. On the other hand, consistent partial fragments of predicate-based epistemic logics, investigated, for instance, in des Rivières and Levesque [16], remain regrettably unmotivated.[35] Worse, they too are consistent only because they proscribe the resources responsible for epistemic paradox.

Second, banishing a predicate conception (or an operator conception with propositional quantification) seems in any case an overreaction. Any epistemic logic that does not allow us to express such banal claims as 'something is known' or 'you know something I don't' or 'Bob and Bill both know something' or 'all I know about Albania I have learned from books', and so on, is expressively defective.

If the above is on the right lines, then the inconsistency should be taken at face value: the maximally strong version of knowability must be rejected. Perhaps that is only to be expected, as the type at which a proposition is known is not in general a logical fact but an extra-logical one depending on how it is derived. What the contradiction shows is precisely that for any type there is some sentence of lower type such that no knowledge-yielding derivations of that sentence at that very type exist.[36] In the absence of a reason for thinking that a proposition can always be known at its very next type, or at some specified higher type, the proponent of knowability need not be troubled by this fact and can accept it as revealing an unattainable upper bound on the strength of knowability. Indeed, he is accustomed to the idea that sentences of type τ are in certain circumstances only $\text{known}_{\tau+\text{m}}$ (where $\text{m} \geq 2$) but not $\text{known}_{\tau+1}$, say, because they are derived by deduction or by inference to the best explanation using sentences of type $\geq \tau + 1$. The earlier reasoning simply reveals that for a handful of sentences this will be true in *all* circumstances. Perhaps a further positive story could be told at this juncture, but even in its absence, I suggest, the knowability theorist should not be unduly troubled by the inconsistency of the maximally strong version of knowability. It goes far beyond what his view commits him to, namely, the schema $\varphi \rightarrow \Diamond\ulcorner \text{K}_\text{n} \ulcorner \varphi \urcorner\urcorner$ for some $\text{n} > \tau(\varphi)$.

Notes

1. The classic paper is Tarski [18]. One way of construing the typing view about truth when proposed about English itself (rather than some reconstruction of it) is that the predicate 'is true' is systematically ambiguous. An alternative proposal, that 'is true' is indexical, is explored in Burge [2]. This article only discusses finite types; an introduction to transfinite hierarchies of truth theories may be found in Feferman [6].

2. For recent surveys, see Brogaard and Salerno [1] and Kvanvig [12].

3. Others have proposed variants of this general idea. For example, Rückert [17] uses the distinction between subjunctive and indicative mood to mark the distinction between the claim that actually p and the claim that p holds at some other possible world.

4. The typing response should also not be confused with the subject-theoretic typing '$K_{(S)}p$' indicating that some subject S knows p. A subject typing is orthogonal to the epistemic typing here proposed. As mentioned in Section 1, the knowledge operator under discussion quantifies over subjects: it is thus equivalent to '$\exists SK_{(S)}p$' or, to make the time-dependence explicit, '$\exists S\exists tK_{(S,t)}p$'.

5. Provable in some absolute sense, that is—which, of course, transcends any given consistent formal system.

6. Almost all previous antirealists have accepted classical logic. Putnam—the middle-period Putnam who urged that “To claim a statement is true is to claim it could be justified” ([14], p. 56)—is perhaps the best-known contemporary antirealist. However, a Dummettian brand of antirealism (see, e.g., Dummett[4]), rejects classical logic in favor of intuitionistic logic. There are, of course, several types of antirealism; some of them have little in common other than their opposition to realism.

7. For instance, I omit the well-known objection that some sentences are paradoxical not because of their meaning but because of empirical facts.

8. In this respect, a typed set theory is arguably better motivated.

9. “…the present approach certainly does not claim to give a universal language, and I doubt that such a goal can be achieved.… The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us.” ([11], p. 714)

10. Friedman and Sheard [8].

11. If you object to calling this a contradiction, call it an absurdity or infelicity or what you like.

12. Or for the nonlinguistic variant, to pragmatic paradoxes arising from thoughts being entertained (rather than sentences being uttered).

13. Of course, logically omniscient subjects may be said to know logical consequences of facts they know in virtue of carrying out reasoning, and so in this etiolated sense the inconsistency of (1) does turn on some reasoning being carried out—but by the logically omniscient subject, not by you or me.

14. Some noncumulative type theories take ‘true_{i+1}’ to apply only to sentences of type i. Minimality is trivial for these.

15. I write ‘concepts/properties’ to preserve neutrality on the constituents of propositions.

16. The modal collapse of possibility into actuality is also a contributing factor.

17. For Dummett’s own restriction response to Fitch’s argument, see Dummett [3].

18. We might test this by asking nonphilosophers diagnostic questions of the kind: “Imagine that some alien creatures had unlimited physical powers and perfect minds: they can travel as fast as they like, observe things as small or large as they like—more generally they are unfettered by our constitution and our laws of nature; on top of that, they never forget anything, they reason perfectly.… Is it plausible or not that any given fact is one they can discover?”

19. But probably not enough to make Fitch’s argument, strictly speaking, a paradox, which is why I have resisted the label.

20. Williamson also raises this possibility: "Perhaps a claim could be known at level $i + 1$ but not at level i if the route to knowing it involved claims about knowledge, even though the target claim did not" ([20], p. 281). Williamson points out that Dummett—one of the targets of his discussion in this passage—could not easily take this route.

21. This dependence is not necessarily a form of reductionism about knowledge, be it a traditional analytic reduction (such as the justified true belief analysis) or a metaphysical reduction (which says that necessarily any instance of knowledge is an instance of something else).

22. Presumably failures of minimality can also be made to seem natural in the case of typed true conceivability predicates or operators TC_0, TC_1, TC_2, and so on. If we assume $p \wedge \neg TC_0 p$ then the Fitch reasoning leads to $\Diamond(TC_1 p \wedge \neg TC_0 p)$, but it seems mistaken to suppose that one can go from $TC_1 p$ to $TC_0 p$. Informally, the point is as follows. In the sort of scenario envisaged here, you $TC_1 p$ simply in virtue of TC_1-ing a conjunction one of whose conjuncts is p. Since the other conjunct is not TC_0-able, it is completely natural that in such a situation TC_1-ing p should not entail TC_0-ing p. You $TC_i p$ only in whatever sense you TC_i the conjunction.

23. The proof generalizes simply to systems similar to L^K. Note that although $\varphi \rightarrow \Diamond K_i \varphi$ is a theorem of L^K for any $i > \tau(\varphi)$, the converse $\Diamond K_i \varphi \rightarrow \varphi$ will not generally be a theorem of L^K. That $\Diamond K_i \varphi \rightarrow \varphi$ is not in general a theorem of L^K is in fact a necessary condition for the failure of omniscience, given knowability. For if $\Diamond K_i \varphi \leftrightarrow \varphi$ (with $i > \tau(\varphi)$) were a theorem, it would follow that $\Diamond\Diamond K_i \varphi \leftrightarrow \Diamond\varphi$. In **S4** that reduces to $\Diamond K_i \varphi \leftrightarrow \Diamond\varphi$, which combines with $\Diamond K_i \varphi \leftrightarrow \varphi$ to give modal collapse, $\Diamond\varphi \leftrightarrow \varphi$, and therefore epistemic collapse, $K_i \varphi \leftrightarrow \varphi$, that is, omniscience.

24. The following avails itself of some key suggestions by Williamson on how to improve my original proof, which contained a flaw. Williamson has found other proofs of similar conservativeness results.

25. Hence the importance of the frame for the model $\langle W, R, Val\rangle$ being an **S5**-frame, to exclude models for **S5** based on frames whose accessibility relation is not an equivalence relation, for example, the model $W = \{w_1, w_2\}, R = \{\langle w_1, w_2\rangle, \langle w_2, w_1\rangle\}$ in which w_2 is just a duplicate of w_1 with respect to its valuation of literals under Val.

26. If Val* applied to ill-formed sentences, then in the canonical model for **S5** it would be true that for any nonepistemic A which is not a theorem of **S5**, $Val^*(K_m \neg K_n A, \langle w, m\rangle) = 1$ since $Val^*(\neg K_n A, \langle w, \langle\rangle\rangle) = 1$, irrespective of whether $m > n$. This illustrates the fact that if Val* were to apply to type-violating sentences, one would have to disregard the truth-values of the type-violating propositions as an artificial feature of that model, superfluous to the basic idea. We have avoided the need for this disinterpretation by allowing only sentences that respect the type restrictions to be evaluated by Val* in the first place.

27. Notice that the set of formulas valid in the model $M = \langle W^*, R^*, \{S_n^*\}_{n \in \mathbb{N}^+}, Val^*\rangle$ may not be closed under uniform substitution. For instance, if M is obtained from the canonical model for **S5** and φ is a nonepistemic non-**S5**-theorem, $M \models \Diamond\neg K_1 \varphi$, but for the substitution instance in which φ is a nonepistemic **S5**-theorem, $M \models \Box K_1 \varphi$. Given the validity of L^K in M, this does not affect our conservativeness result, as we have

proved that adding strong knowability does not generate any extra theorems of the form $A \to K_n A$ if A is nonepistemic.

28. Note that it's not true that any world $\langle w, s\rangle$ in which m is a member of the sequence s is K_m-omniscient; that is, it's not true that $Val^*(K_m A \leftrightarrow A, \langle w, s\rangle) = 1$ in such a world (where $m > \tau(A)$). For instance, suppose that in a model that extends the canonical model for **S5**, the literal q is true at $\langle w, \langle\,\rangle\rangle$. Then K_2q, $\neg K_1 q$, and q are true at $\langle w, \langle 2\rangle\rangle$, and K_2q, K_1q, and q are true at $\langle w, \langle 2, 1\rangle\rangle$. But since $\langle w, \langle 2, 1\rangle\rangle S^*_2 \langle w, \langle 2\rangle\rangle$, K_2K_1q is not true at $\langle w, \langle 2, 1\rangle\rangle$ though K_1q is true there and 2 is a member of $\langle 2, 1\rangle$.

29. We are assuming here that $\langle W, R, Val\rangle$ is a model in which literals of the language are true at w if and only if they are members of w. More generally, we could take the final disjunctive clause to be 'n = m and t = $\langle\,\rangle$ and $\{\varphi : \varphi$ is a true literal at x$\} \cap \{\varphi : \varphi$ is a true literal at $\langle w, \langle\,\rangle\rangle\} = \alpha$'.

30. The important fact about the set π is that it is unsatisfiable (any other such set would do). Its unsatisfiability implies that $\langle w, s\hat{}\langle \pi, m\rangle\rangle$ only S^{**}_m-accesses itself.

31. If an ascending knowledge thesis of the form $K_{\tau(\varphi)+m}\varphi \to K_{\tau(\varphi)+m+j}\varphi$ for $j \geq 1$, $m \geq 1$ is part of the system, infinitely many conditionals $p \to \Diamond K_n p$ hold if some conditional of the form $p \to \Diamond K_m p$ does.

32. Do not confuse this with rejection of minimality: that it's possible that $K_{\tau(\varphi)+1}\varphi$ is compatible with it being possible that $K_{\tau(\varphi)+j}\varphi$ but not $K_{\tau(\varphi)+1}\varphi$ for some $j \geq 2$.

33. Likewise, he is only committed to the inference rule if $\vdash \varphi$ then $\vdash K_n\varphi$ for some $n > \tau(\varphi)$.

34. This was brought to my attention by Volker Halbach. Note that modal as well as epistemic predicates should be typed in such a logic (and taken as predicates—as they are in this paragraph). Typing modality is motivated in the same way that typing knowledge is, for example, with sentences such as 'This sentence is not possible'. (If the sentence is impossible then it's true and thus possible; from this we may infer that it's necessarily false, that is, impossible.) We ignore the modal typing in this section, to keep the exposition simple and because both the objection leveled at knowability and my response to it go through, mutatis mutandis, in a logic in which modality is also typed.

35. For example, systems in which acceptable instances of the schema $K\ulcorner\varphi\urcorner \to \varphi$ are restricted to translations of operator-based modal sentences.

36. This is a mild generalization of the stated contradiction.

References

[1] Brogaard, B., and J. Salerno, "Fitch's paradox of knowability," *The Stanford Encyclopedia of Philosophy*, (2004). http://plato.stanford.edu/entries/fitch-paradox. 171

[2] Burge, T., "Semantical paradox," *Journal of Philosophy*, vol. 76 (1979), pp. 169–98. 171

[3] Dummett, M., "Victor's error," *Analysis*, vol. 61 (2001), pp. 1–2. 172

[4] Dummett, M., "The philosophical basis of intuitionistic logic," pp. 5–40 in *Logic Colloquium '73 (Bristol, 1973)*, edited by H. E. Rose and J. C. Shepherdson, vol. 80 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1975. Reprinted in *Truth and Other Enigmas*, Duckworth, London, 1978, pp. 215–47. Zbl 0325.02005. MR 0439563. 172

[5] Edgington, D., "The paradox of knowability," *Mind*, vol. 94 (1985), pp. 557–68. MR 878514. 156

[6] Feferman, S., "Reflecting on incompleteness," *The Journal of Symbolic Logic*, vol. 56 (1991), pp. 1–49. MR 1131728. 171

[7] Fitch, F. B., "A logical analysis of some value concepts," *The Journal of Symbolic Logic*, vol. 28 (1963), pp. 135–42. MR 0170792. 153

[8] Friedman, H., and M. Sheard, "An axiomatic approach to self-referential truth," *Annals of Pure and Applied Logic*, vol. 33 (1987), pp. 1–21. MR 870684. 172

[9] Grim, P., "Operators in the paradox of the knower," *Synthese*, vol. 94 (1993), pp. 409–28. MR 0803.03003. 170

[10] Hintikka, J., *Knowledge and Belief: An Introduction to the Logic of the Two Notions*, Cornell University Press, New York, 1962. 160

[11] Kripke, S., "Outline of a theory of truth," *Journal of Philosophy*, vol. 72 (1975), pp. 690–716. 157, 172

[12] Kvanvig, J., *The Knowability Paradox*, Oxford University Press, London, 2006. 171

[13] Montague, R., "Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability," *Acta Philosophica Fennica*, vol. 16 (1963), pp. 153–67. Reprinted in H. Thomason, editor, *Formal Philosophy: Selected Papers of Richard Montague (1974)*, Yale University Press, New Haven, pp. 286–302. MR 0163841. 170

[14] Putnam, H., *Reason, Truth and History*, Cambridge University Press, London, 1981. 172

[15] Quine, W., "The ways of paradox," *Scientific American*, vol. 26 (1962). First published as "Paradox"; reprinted in his *The Ways of Paradox and Other Essays* (2d ed., 1976), Harvard University Press, Cambridge, pp. 1–18. 161

[16] Riviéres, J. des, and H. Levesque, "The consistency of syntactical treatments of knowledge," pp. 115–30 in *Theoretical Aspects of Reasoning about Knowledge*, edited by J. Halpern, Morgan Kaufman, San Francisco, 1986. 170

[17] Rückert, H., "A solution to Fitch's paradox of knowability," pp. 351–80 in *Logic, Epistemology and the Unity of Science*, edited by S. Rahman and J. Symons, Springer Verlag, Dordrecht, 2004. 171

[18] Tarski, A., "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, vol. 1 (1935), pp. 261–405. Translated by J. H. Woodger as "The Concept of Truth in Formalized Languages" and reprinted in J. H. Woodger, editor, *Logic, Semantics, Metamathematics*, The Clarendon Press, 1956, pp. 152–278. Zbl 0013.28903. 171

[19] Tennant, N., *The Taming of the True*, The Clarendon Press, New York, 1997. Zbl 0929.03001. MR 1469976. 155

[20] Williamson, T., *Knowledge and Its Limits*, Oxford University Press, London, 2000. 156, 158, 160, 173

[21] Williamson, T., "Tennant on knowable truth," *Ratio*, vol. 13 (2000), pp. 99–114. 156

[22] Williamson, T., "Knowability and constructivism," *The Philosophical Quarterly*, vol. 38 (1988), pp. 422–32. MR 986929. 156

Acknowledgments

I am indebted to Tim Williamson for his work on Fitch's argument that prompted this article and for generous discussion and comments on previous drafts. I have also benefited from discussions with Andrew Bacon on how to motivate typing, with Volker Halbach on the issues dealt with in Section 7, and with Bruno Whittle, who made several constructive suggestions. Finally, thanks to some *NDJFL* referees for comments.

Wadham College
Oxford OX1 3PN
UNITED KINGDOM
alexander.paseau@philosophy.ox.ac.uk