Knowability as Learning

The aim of this paper is to revisit Fitch's Paradox of Knowability in order to challenge an assumption implicit in the literature, namely, that the key formal sentences in the proof adequately represent the intended epistemic meaning of their informal counterparts (call this the "standard interpretation"). The assumption in question drives much of the technical work done in recent years on Fitch's paradox, and is central to the philosophical debate concerning realist and anti-realist views of truth and knowledge (a debate which brought Fitch's formal proof to the forefront of philosophical logic not long ago). The central challenge I pose to the standard reading of Fitch's paradox concerns claims of the form "p is knowable". I argue that questions about the knowability of a proposition are analogous to questions about the provability of a theorem or the computability of a function, questions which should not be answered in the style of the standard interpretation. In taking the analogy to provability and computability seriously, I call attention to different formal semantics for knowability claims that are, however, partially independent of semantics for simple knowledge claims (i.e. claims of the form "p is known"). The semantics in question borrow heavily from *learning theory.* The paper aims to illustrate the central difference between the standard modal logic semantics and the learning theoretic semantics in terms of the difference of the truth-conditions they require for knowability claims. I argue that unlike the standard modal semantics, the learning theoretic semantics better accommodates the intended informal epistemic meaning of such claims, insofar as they are taken to be *epistemic* claims concerning knowledge instead of *metaphysical* claims about possibility.

1 Introduction

Consider the question of whether a given proposition p is knowable or not. Informally, *knowability* is often understood as a possibility of sorts. For example, one can say of an inconclusive criminal investigation that the identity of the perpetrator is unknowable given the evidence; one can say of a purported historical fact that it is unknowable whether it really happened or not given the lack of evidence and distant time of occurrence; one can say that some events will forever remain unknowable to us, merely in virtue of the fact that we are a certain type of creatures with certain limitations, both physically, cognitively and spatio-temporally (i.e. it might be claimed to be unknowable to us what it is like to have telepathic abilities, or perfect memory, or perfect mathematical reasoning, etc.). In all these examples, certain facts are said to be such that it is *not possible* to know them due to whatever reasons.[1] The examples also illustrate the variety of the reasons in question, some appealing to a purported relationship between the agent and her evidence, some appealing to the agent's overall epistemic environment, some appealing to the agent's constitution and inherent limitations. What is knowable and what is not can thus be understood as a function of a number of variables.

A formal exploration of the question of knowability, and how best to model it,

[1] Notice that I am understanding knowability as pertaining to true propositions (facts). This is to be distinguished from knowability understood counterfactually, i.e. as the question of whether certain non-facts could be known other things being different (could we know who the perpetrator is if we had an oracle that doled out the relevant missing evidence? Could we know whether the purported historical fact really happened or not if we had a time machine? etc.).

can thus begin by looking at what might plausibly be the most specific level, and generate more general levels by mere quantification. So if we let $P = \{\varphi | \varphi \in Prop\}$ (where $Prop$ is just the standard recursive set of propositions from classic propositional logic), $A = \{a, b, c, ...\}$ denote a set of agents, $W = \{w_0, w_1, ...\}$ denote a set of epistemic environments, and $E = \{H_i | H_i \subseteq \wp(Prop)\}$ (thus modeling an agent's evidence as a set of propositions indexed as usual), one could initially consider, as the most specific level, a predicate asserting that a specific proposition p is knowable by a specific agent a on a specific epistemic environment w_n given the agent's specific evidence H_nm perhaps something that looks like $Knowable(p, a, w_n, H_n)$. By quantifying on the parameters we can ask more general questions, such as "what is knowable to the agent a in general?" or "are there epistemic environments and evidence sets such that any agent could in principle know that p?", by considering truth conditions for the predicates $\forall\varphi\exists w_i\exists H_i Knowable(\varphi, a, w_i, H_i)$ and $\exists w_i\exists H_i Knowable(p, a, w_i, H_i)$ respectively.

It seems altogether clear that questions of knowability are questions about a possibility of sorts. The simplest model in the literature takes the standard Kripke models for possibility and necessity and applies them to a simple epistemic logic. If you let K be an operator denoting "it is known (by someone at some time that)...", then this model parses a knowability claim of the form "p is knowable" as $\Diamond Kp$. Given the standard semantics for the modal and epistemic operators, this model has as a consequence: $\models (\varphi \rightarrow \Diamond K\varphi) \rightarrow (\varphi \rightarrow K\varphi)$ for any φ. This result is known as *Fitch's Paradox*, and is informally taken to mean that if every truth is knowable, then every truth is known. The aim of this paper is to use Fitch's result as motivation for thinking that the model itself is in some respect defective and that it fails to account for the kind of possibility involved in knowability claims. An alternative model is sketched that remedies these defects, and some epistemological considerations are brought to bear to justify the proposed model.

2 Fitch's Paradox

2.1 Semantics and Proof

Let L be the language of classical propositional modal logic augmented by an operator K. Let $w \in W$ corresponds to the set of "points" or "worlds" and let $R \subseteq W \times W$ be a relation between worlds, one for the operators $\Box, \Diamond$ denoted R_M and one for the epistemic operator K, denoted R_E. A *frame* will be a triple $F =< W, R_M, R_E >$, and a model will be a quadruple $M =< W, R_M, R_E, \Vdash>$where $\Vdash$ is an assignment of formulas of the language to worlds in W defined in the standard recursive manner:

- $w \Vdash \neg\varphi$ iff $w \nVdash \varphi$
- $w \Vdash \varphi \to \psi$ iff $w \Vdash \neg\varphi$ or $w \Vdash \psi$
- $w \Vdash \varphi \wedge \psi$ iff $w \Vdash \varphi$ and $w \Vdash \psi$ (similar condition for $\vee$)
- $w \Vdash \Box\varphi$ iff for all w_i such that $R_M(w, w_i)$, $w_i \Vdash \varphi$
- $w \Vdash \Diamond\varphi$ iff for some w_i such that $R_M(w, w_i)$, $w_i \Vdash \varphi$
- $w \Vdash K\varphi$ iff for all w_i such that $R_E(w, w_i)$, $w_i \Vdash \varphi$

No special restrictions are placed for R_M, but we shall assume that R_E is reflexive the guarantee the factivity of the K operator (i.e. to ensure that $K\varphi \models \varphi$ as knowledge is said to entail the truth of that which is known).

Lemma (Distribution). $\models K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$

Proof: Let M be an arbitrary model. Let $w \Vdash K(\varphi \wedge \psi)$ for some arbitrary $w \in W$. Then $w_i \Vdash \varphi \wedge \psi$ for all $R_E(w, w_i)$. Then $w_i \Vdash \varphi$ and $w_i \Vdash \psi$. But then since $R_E(w, w_i)$, then $w \Vdash K\varphi$ and $w \Vdash K\psi$. If so, then $w \Vdash K\varphi \wedge K\psi$.

Lemma (*Moore's Sentence*)**.** $\models \Box\neg K(\varphi \wedge \neg K\varphi)$

Proof: Let *M, w* be arbitrary model and world. Assume towards a contradiction that $w \Vdash K(\varphi \wedge \neg K\varphi)$. Then $w \Vdash K\varphi$ and $w \Vdash K\neg K\varphi$(by Distribution Lemma). Since R_E is reflexive, then $w \Vdash \neg K\varphi$. Then $w \Vdash K\varphi \wedge \neg K\varphi$, which is a contradiction. Discharging our assumption, $w \nVdash K(\varphi \wedge \neg K\varphi)$ so then $w \Vdash \neg K(\varphi \wedge \neg K\varphi)$. Since w is arbitrary, $M \models \Box\neg K(\varphi \wedge \neg K\varphi)$. Since M was an arbitrary model, $\models \Box\neg K(\varphi \wedge \neg K\varphi)$.

Theorem (Fitch's Paradox)**.** $\models (\varphi \rightarrow \Diamond K\varphi) \rightarrow (\varphi \rightarrow K\varphi)$

Proof: Let *M, w, φ* be arbitrary model, world and sentence. Assume that $w \Vdash (\varphi \rightarrow \Diamond K\varphi)$, and assume towards a contradiction that $w \nVdash (\varphi \rightarrow K\varphi)$. Then $w \Vdash \neg(\varphi \rightarrow K\varphi)$ so that $w \Vdash (\varphi \wedge \neg K\varphi)$. By our assumption, since φ was arbitrary, then it can be instantiated by $(\varphi \wedge \neg K\varphi)$, so that we get $w \Vdash ((\varphi \wedge \neg K\varphi) \rightarrow \Diamond K(\varphi \wedge \neg K\varphi))$. Since $w \Vdash (\varphi \wedge \neg K\varphi)$, by *modus ponens* then $w \Vdash \Diamond K(\varphi \wedge \neg K\varphi)$. Then for some w' such that $R_M(w, w')$, $w' \Vdash K(\varphi \wedge \neg K\varphi)$. By the above Lemma (Moore's sentence), $w' \Vdash \neg K(\varphi \wedge \neg K\varphi)$. Contradiction. Discharging our second assumption, then $w \Vdash (\varphi \rightarrow K\varphi)$. Since M, w, φ are all arbitrary, then $\models (\varphi \rightarrow \Diamond K\varphi) \rightarrow (\varphi \rightarrow K\varphi)$.

2.2 Paradoxicality

Informally speaking, Fitch's result demonstrates, in the given model, the inconsistency of the claims that every truth is knowable, on one hand, and that not every truth is known, on the second hand. Since it would be rather pressing to maintain that every truth is indeed known, the standard response to Fitch's result is to take it as a *reductio* of the claim that all truths are knowable (see Hart (1976), Williamson (1987), and especially Williamson (2000) where the result is taken as demonstrating a "structural" limit to knowledge).

There are a number of philosophical theses that are committed in one form or another to the claim that all truths are knowable, so Fitch's result is often discussed in debates concerning the appropriateness of such so-called anti-realist theses. Some authors find as cause of puzzlement that sophisticated versions of anti-realist positions can be refuted, or at least seriously impaired, by Fitch's result (see Salerno (2009)). One might also be puzzled by the result to the extent that it shows how contingent ignorance entails necessary ignorance. Indeed, a key ingredient required for Fitch's result is the Lemma I have dubbed "Moore's sentence". This Lemma exploits the fact that if ignorance exists, i.e. there is some truth that is not known by anyone at anytime, then knowledge *of that such fact* is impossible, to the extent that it would require knowledge of that which is not known.[2] One may reasonably be suspicious of the claim that if ignorance is expressed by $\varphi \wedge \neg K\varphi$ for some true φ, then knowledge of one's own ignorance, i.e. $K(\varphi \wedge \neg K\varphi)$ cannot be had. Another reason to be somewhat perplexed by the result is that it falls right out of the system as a logical consequence of its semantics, without any external assumptions playing any role. As such, some have found it odd that a substantial claim about the limits of knowledge could be so easily established by only a very few semantic conditions on knowledge and possibility operators.

The literature on Fitch's result is multifarious, ranging from treating it as a paradox requiring a revision or reformulation of certain anti-realist theses,[3] treating

[2] This is similar to G. E. Moore's famous discussion about sentences of the form "p is true, but I don't believe it".

[3] E.g. Tennant (1997), Edginton (1985), or Dummet (2009) for an account that aims to demonstrate that Fitch's result is not threatening to certain intuitionist theses.

it as a paradox concerned with the sheer collapse of necessity and possibility in epistemic contexts (due to a fallacious aspect of the model),[4] treating it as a robust result concerning the limits of knowledge,[5] and finally as a challenge to revise the appropriateness of the model in which the result is framed. We shall assume this last attitude. [6]

3 Knowability, Computability and Learning

3.1 Ways to come to know

As we have stated a number of times, knowability is a possibility of sorts. If the standard, simple model outlined above is to be found wanting, then a case can be made that it fails to represent the relevant type of possibility involved in knowability sentences. The parsing of knowability sentences that the above model employs treats the possibility in question as bound to the relation R_M, which we often call a "metaphysical" accessibility relation between possible states of affairs. The relation R_E, which is often called the "epistemic" accessibility relation, only comes to bear on the question of whether some given proposition *is known* in a given world. To the extent that these relations are structurally and philosophically distinct, knowability is cashed out as a metaphysical possibility concerning epistemic claims.[7]

[4] E.g. Kvanvig (2006)

[5] E.g. Williamson (2000)

[6] For some examples of work done in this same general direction, see van Benthem (2004), van Benthem (2009), Restall (2009).

[7] The relations would be structurally distinct to the extent that the epistemic accessibility relation is restricted in at least one aspect in which the metaphysical relation of accesibility is not, namely, its reflexivity. Analogously, the metaphysical possibilities (i.e. those accesible from a given world) are often intended to be stronger than epistemic possibilities. For example, given that the Evening Star is identical to the Morning Star, the claim "Evening Star = Morning Star" is true across all worlds where Venus exists. However, it was an epistemic possibility that the Morning Star $\neq$the Evening Star, even though it was not a metaphysical possibility. Examples of this sort are meant to illustrate the difference between epistemic possibilities and metaphysical possibilities (cf. DeRose (1991))

So a worry is that the model should be about an epistemic possibility rather than a metaphysical possibility. A second worry is that perhaps the model places too much of the burden of the semantics on there being certan types of *states* somehow connected (accesible) in the relevant manner. What if knowability is more about the nature of the connection between these states, and not so much about the states themselves?

To illustrate these worries, suppose you are taking your first propositional logic class in college, and its the end of semester and you have become proficient at proving things in the standard propositional calculus. So far you have been asked to prove things that you have been told before hand can indeed be proved, so that so far all you have had to do is engineer a sequence of valid steps from premises to conclusions. Now your instructor writes a formula on the board, and asks the class whether that formula is a theorem of propositional logic, i.e. whether it is provable in propositional logic alone, with no premises. Suppose that the formula happens to be indeed provable in classical propositional logic alone. Here is what a bad answer would be: to claim that the formula is indeed provable since there is some scenario, very similar to the actual scenario, where the formula has indeed been proved *somehow.* A logic instructor will quickly correct the student providing one such answer by pointing out that she was not asking whether it was *conceivable* that the formula could be proved, somehow, but rather she was asking about the *existence* of a proof. Similarly, imagine that you are now in a class on computability theory, and you are now familiar with the standard Turing models. You know that a great deal of functions are computable, e.g. addition, division, etc. The instructor defines a function f you have never seen before on the board, and asks the class if that function is computable. Suppose it is indeed computable. Here is again what a bad answer would look like: f is computable to the extent that there is a scenario, not unlike the present scenario, where f is the function being computed by *some* Turing machine M_i. The instructor will quickly point out that that is not what is *meant* by the question, i.e. whether you could *conceive* of a scenario where the function is being computed by some Turing machine.What her question was *meant* to be asking was, rather, whether some Turing machine does indeed exist that can be shown to *be computing* the function f in question.

What both examples are drawing attention to is that for questions about provability or computability, it is the existence of some procedure or transition between items what is at stake. In the case of provability we are concerned with the existence of a chain of inferences linking axioms or theorems to further theorems. In the case of computability we are concerned with the existence of *programs* or recursive procedures that can be demonstrated to generate all and only the elements of a given set. Knowability is better understood as standing in analogous grounds to provability and computability. In asking whether p is knowable for an agent a (given some environment, etc.), the question of central epistemological interest is whether the agent could *come to know* that p, whether p is within the epistemological ken of the agent or not. This emphasis leads one to consider models that implement some dynamics between an agent's states.[8]

Given the analogy to questions of computability, a natural avenue to explore is the application of concepts and methods from computability theory to models of epistemic logic. The application in mind borrows heavily from *learning theory* (sometimes also known as computational This application is not novel by any means (see Kelly (1996) for a rich discussion and application to a number of epistemological questions; I borrow heavily from the presentation therein). I aim to emphasize the contrast with the standard model, and demonstrate the

[8] van Benthem (2004,0) are primary examples of applications of *dynamic epistemic logic* in light of Fitch's result.

flexibility of this *learning theoretic* model in terms of the flexibility it provides in exploring issues of knowability in the fashion we saw in the introduction to this paper. The model's complexity, compared to the simple standard Kripke model, is the price one pays for a richer and deeper apparatus to explore the intricacies of the concept of knowability and its cognates.

For simplicity we shall imagine we are concerned with a single agent a. The agent has at her disposal a finite segment of a denumerable *sequence* of data, encoded in some canonical language (we shall assume the evidence is propositional). *Time* will be modeled as a function of the growth of the data. The agent's *environment* corresponds to her finite segment of data, any and all background knowledge she has, and some set of *hypothes*es. We shall imagine the agent is exclusively concerned with the task of determining whether some hypothesis she is pondering is true or not. Formally put:

- $\omega_i =< \varphi_o, \psi_1, ... >$ denotes an infinitely denumerable ω-sequence of *datums*, whose index corresponds to the *time* the datum is presented to the agent, and where $\varphi, \psi, ...$ are formulas of a propositional language.
- ω^ωdenotes the set of all ω-sequences.
- $E =< K, H >$ is the agent's environment, where K is some set of propositions that constitute the agent's background knowledge, and $h \in H$ is an hypothesis the agent is concerned about. The set K serves to rule out some members of ω^ω so as to reduce the agent's possible data sequences to a proper subset of all possible sequences.

We shall idealize the situation to be such that the truth of any $h \in H$ is not under determined by any data sequence. In other words, if h is true, then it neatly partitions the set of data sequences ω^ω into those that make h true and those that make h false. Conversely, any particular data sequence is such that it entails h or it entails $\neg h$. The agent can, upon presentation of a new *datum*, engage in one of the following options: she can conjecture h, she can conjecture $\neg h$, or she can suspend judgment (perhaps until a new *datum* is presented). We shall call a *method* any well-defined *program* that encodes the agent's reactions to the data as it pours in. Methods can be such that they can either converge in the limit or loop indefinitely. Formally:

- $[\omega_h] \cup [\omega_{\neg h}] = \omega^\omega$(any hypothesis h partitions the set of data sequences)
- $M = \{m|m : \omega^\omega \rightarrow \{h, \neg h, undef\}\}$ (M is the set of methods, i.e. the set of functions from data sequences to a set that either outputs h, $\neg h$ or neither -we use *undef* to make clear the possibility of the method being a *partial* function to the set $\{h, \neg h\}$)
- $Conv(m, \omega_i) \Leftrightarrow_{df} (\exists \varphi_i)[(\varphi_i \in \omega) \wedge \forall \varphi_{j \geqq i} m(\varphi_j) = m(\varphi_i)]$ (m converges on sequence ω at entry i iff the method outputs the same output for any subsequent entry on ω)
- $Loop(m, \omega) \Leftrightarrow_{df} \neg Conv(m, \omega)$

A method will be said to be *logically reliable in the limit* if, for all data sequences, it converges on the correct hypothesis for that data sequence.:

- $Rel(m, h) \Leftrightarrow_{df} (\forall \omega_i \in \omega^\omega)(Conv(m, \omega_i) \wedge (m(\omega_i) = h \leftrightarrow \omega_i \in [\omega_h]))$ (these methods are called *verifiers,* since they converge on h when h is true)[9]
- $Rel(m, \neg h) \Leftrightarrow_{df} (\forall \omega_i \in \omega^\omega)(Conv(m, \omega_i) \wedge (m(\omega_i) = \neg h \leftrightarrow \omega_i \in [\omega_{\neg h}]))$ (these methods are called *refutators* since they converge on $\neg h$ when h is false)

[9] These correspond to *limiting computable r.e. sets.*

Notice that a method being logically reliable with respect to h does not mean that the method in question needs to converge on $\neg h$ if h is false. It only needs to *avoid* converging on $\neg h$. When a method m is logically reliable with respect to both h and $\neg h$, then it is a *reliable decider* of whether h *or not* h*:*

- $Dec(m, (h \vee \neg h)) \Leftrightarrow_{df} Rel(m, h) \wedge Rel(m, \neg h)$

Notice that a method can be logically reliable while still producing some false outputs until it manages to converge. But it is necessary for its being logically reliable that it only produces a finite amount of such outputs. Also notice that a method can converge without there being any signal or indication *for the agent* telling her that the method has indeed converged.

With these elements we can construct a rather natural condition for the knowability of some hypothesis h given a *fixed* epistemic environment, set of data streams and extensions of them, namely, the *existence* of a logical reliable method, the implementation of which would lead the agent (in the limit) to conjecture the truth of h ever after.

- $Knowable(h) \Leftrightarrow_{df} (\exists m \in M) Rel(m, h)$

As in the case of reliable decidability, a distinction needs to be made between h being knowable by itself, and *whether* h *or not* h is knowable:

- $Knowable(h \vee \neg h) \Leftrightarrow_{df} (\exists m \in M) Rel(m, h) \wedge (\exists m \in M) Rel(m, \neg h)$

Notice that it being knowable whether h need not entail that one and the same method is logically reliable as to whether h. In other words, it being knowable whether h need not require there being a decider of h.[10] These definitions can be

[10] Often, if there exists a verifier and a refutator for h, a decider can be built by composing or dovetailing both the verifier and refutator, but this is not in general always the case.

generalized to consider the knowability of a hypothesis not just on a fixed environment and set of data streams, but on *any* environment and *any* data stream (whenever $K = \emptyset$ then *every* data sequence must be considered in assessing a method's logical reliability). Since different notions of convergence are amenable to specification (i.e. convergence with n conjecture flips, convergence by a fixed time, convergence to a gradual interval, etc), corresponding notions of knowability can be easily defined (knowability with n conjecture flips, knowability by some fixed time, etc.).[11]

As a rather simple example, let us consider the following example: Audrey is pondering whether all ravens are black or not. Unbeknownst to her, all ravens re indeed black. She has not collected any data yet, and nothing she knows already bears on the issue (thus $K = \emptyset$). She settles on employing the following method:

If the latest data observed consists of a black raven, conjecture "all ravens are black" and continue obtaining data;

else, conjecture "not all ravens are black" forever after.

Assuming all her data consists of nothing but observations of ravens that are black or not, we can determine whether via this method Audrey can come to know that all ravens are black. Since all ravens are indeed black, then whatever data sequence Audrey comes across will be such that she will correctly converge, upon the first observation, that all ravens are black (naturally Audrey does not know that she has converged when it does happen). *If* not all ravens were black, then Audrey would, in the limit, find an data entry of a non-black raven, in which case her method will correctly have her conjecture "not all ravens are black" forever after. In either scenario, her method would converge on the right

[11] See (Kelly, 1996, Ch. 7, 8, 9)

answer no matter what is the case (perhaps after a long time of producing an incorrect conjecture).

3.2 Epistemological Virtues and Worries

To use these concepts in providing semantics for an epistemic logic, we can interpret the operator $\Diamond$ in $\Diamond Kp$ as an existential quantifier not over possible worlds (i.e. static descriptions that contain an agent's knowledge) but rather over *logically reliable methods*, which can be understood as *procedures* to help the agent navigate through a infinitely denumerable sequence of states (defined by her finite amount of evidence at any given time, her background knowledge and her current reaction to her evidence). The provided definitions might make on worry that the model would translate into an account of full-fledged knowledge along the following lines: S knows that p at t iff S has converged on p by t through a logically reliable method. This might worry an epistemologist concerned with issues of justification or warranted assertability. Consider Audrey's case above. Upon her very first observation, she will converge on the correct conjecture. Does that mean that she thereby *knows* that all ravens are black? One observation seems too meager an evidence set to ascribe Audrey with any knowledge yet. Surely, she needs more instances before being *warranted* or *justified* in believing that all ravens are black. Similarly, perhaps because of how beliefs work (if knowledge is indeed something like justified true belief), Audrey quite likely doesn't yet *believe* that all ravens are black after her very first observation -she might merely be *saying* it because her method requires her to do so.

I do not aim to engage the traditional epistemologist on these points, as good as they are. I am willing to grant him that perhaps that is indeed the case, so that logically reliable convergence does not *ipso facto* entail knowledge. But

I do want to claim that in such scenario, the agent is *guaranteed* to have the possibility of forming a true belief *and* of being justified, whatever else it takes. After all, she will keep on gathering evidence upon convergence (remember that *she* doesn't *know* her method has indeed converged, because convergence does not entail any sort of marking or signal that it has happened). At some point, she is bound to have "enough" evidence by whatever standards one might want for evidential support. The evidence might be vast enough to *convince or cause* her to believe that all ravens are black. Whatever else needs to take place, the conditions for its obtaining can be guaranteed by employing a logically reliable method. If this is right, one can have a model of knowability separated (albeit not entirely independent) from a model of knowledge. Perhaps a lesson to derive from Fitch's proof is that a model of knowability is not so easily derived from a model of knowledge.

The most relevant consequence of this computational model, as far as Fitch's paradox goes, is that it falsifies the antecedent of Fitch's proof, at least when considered unrestricted. The reason is simply that it is false that every truth is knowable (generalizing over every possible data sequence, environment and background knowledge). For there are truths for which it can be proved that no logically reliable method exists (*modulo* the aforementioned generalizations).[12]

4 Coda

We have considered the possibility of furnishing an epistemic logic aimed at modeling issues of knowability with semantics borrowed from work on computational learning. The framework brings to bear a rich number of parameters

[12] For example, the real number whose binary expansion encodes the truth-set of FOL is non-computable in the limit, the notion of limiting computability corresponding (roughly) to that of knowability.

all informally taken to be a factor in what counts for or against some fact being knowable or not. Manipulation of a given parameter or set of parameters can give rise to different degrees or levels of knowability (and, more interestingly perhaps, unknowability). Factors which are amiss in the standard framework sometimes assumed to correctly model knowability sentences, and which has as a logical consequence an entailment deemed problematic both for philosophical and modeling reasons. In at least one respect, the learning theoretic framework *vindicates* the Fitch result, albeit for entirely different reasons (the proofs as to the existence of truths for which no logically reliable method exists do not involve anything like the Moore sentence Lemma, and are more akin to standard proofs using diagonalization techniques). But the result might not be so robust as to remain invariant under different manipulation of the parameters. This and a host of other questions are open for exploration on this framework (what class of sentences in a propositional language can be said to be knowable under a particular specification of the evidence? under a specification of the background knowledge? what about sentences in a first-order language? what about sentences expressing knowledge of one's ignorance? are they knowable, and under what conditions? etc.).

References

DeRose, K. 1991. "Epistemic Possibilities". *The Philosophical Review*, **100**(4), 581–605.

Dummet, M. 2009. "Fitch's Paradox of Knowability". *In:* Salerno, J. (ed), *New Essays on the Knowability Paradox*. Oxford University Press.

Edginton, D. 1985. "The Paradox Of Knowability". *Mind*, 557–568.

Hart, W.D. McGinn, C. 1976. "Knowledge and Necessity". *Journal of Philosophical Logic*, **5**, 205–208.

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

Kvanvig, J. 2006. *The Knowability Paradox.* Oxford University Press.

Restall, G. 2009. "Not Every Truth Can Be Known". *In:* Salerno, J. (ed), *New Essays on the Knowability Paradox.* Oxford University Press.

Salerno, J. 2009. "Knowability Noir: 1945-1963". *In: New Essays on the Knowability Paradox.* Oxford University Press.

Tennant, N. 1997. *The Taming of the True.* Oxford: Clarendon Press.

van Benthem, J. 2004. "What One May Come to Know". *Analysis*, **64**, 95–105.

van Benthem, J. 2009. "Actions That Make Us Know". *Pages 129–146 of:* Salerno, J. (ed), *New Essays On The Knowability Paradox.* Oxford University Press.

Williamson, T. 1987. "On the Paradox of Knowability". *Mind*, **96**, 256–261.

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