Knowledge, proof and the Knower

Walter Dean & Hidenori Kurokawa

wdean@gc.cuny.edu & hkurokawa@gc.cuny.edu

May 16, 2009

Abstract

The Knower Paradox demonstrates that any theory T which 1) extends Robinson arithmetic Q, 2) includes a predicate $K(x)$ intended to formalize "the formula with godel number x is known by agent i," and 3) contains certain elementary epistemic principles involving $K(x)$ is inconsistent. The purpose of this paper is to show how this paradox may be redeveloped within a system of quantified explicit modal logic in the tradition of Artemov [4] and Fitting [10], [11] which we argue allows for a more faithful formulation of some of the epistemic principles on which it is based. Along the way, we isolate a principle – the Uniform Barcan Formula [UBF] – which we show is required to derive an explicit counterpart of the axiom $\mathbf{U}$ (i.e. $K(\ulcorner K(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$) which was used in the original formulation of the Paradox. We argue that since there are independent epistemic reasons to be suspicious of UBF, the paradox may be resolved by abandoning this principle (and thereby $\mathbf{U}$ as well).

1 Introduction

A simplified formulation of Montague and Kaplan's [19] Paradox of the Knower is as follows: if a predicate $K(x)$, intended to express that the sentence with godel-number x is known by an agent i, is adjoined to an arithmetic theory Z sufficiently strong to formalize reasoning about its own syntax (e.g. Robinson arithmetic [Q]) and it is assumed that knowledge implies truth – i.e.

$\mathbf{T}$ $\quad K(\ulcorner \varphi \urcorner) \to \varphi$

and that this fact is known – i.e.

$\mathbf{U}$ $\quad K(\ulcorner K(\ulcorner \varphi \urcorner) \to \varphi \urcorner)$

– then the resulting theory T is inconsistent. The derivation of a contradiction passes through the fact that the existence of self-referential statements about the predicate $K(x)$ such as

(1)   i) $\delta \leftrightarrow K(\ulcorner \neg \delta \urcorner)$
       ii) $\delta \leftrightarrow \neg K(\ulcorner \delta \urcorner)$

will be provable in T. Such statements respectively receive the interpretations δ *is true iff i knows δ to be false* and δ *is true iff i does not know δ*. As such, the Knower is conventionally taken to reflect a conflict between certain intuitively evident principles about knowledge and the possibility of accommodating epistemic self-reference within a single theory for reasoning about knowledge.

Montague and Kaplan [M&K] originally presented the Knower as a zero-day or 'instantaneous' version of the Hangman Paradox wherein a judge issues a decree of the form *i knows that this decree is false*. If we let δ stand for the proposition expressed by this statement, then it is evident that δ is true iff i knows $\neg\delta$ – i.e. if δ satisfies (1i). The situation envisioned by the original formulation of the Knower is thus one in which the existence of self-referential statements about knowledge comes about through linguistic stipulation. Others (e.g. Tymoczko [22], Burge [7] and Koons [14]) have argued that since epistemic self-reference can arise in a variety of other mechanisms, it would be arbitrary to exclude statements like (1i,ii) from the scope of an epistemic theory. M&K accommodate this observation by proposing that knowledge should be treated as a predicate of sentences and that this predicate be adjoined to an arithmetic background theory Z to which the Gödel-Carnap Fixed Point Theorem applies. But as we will see below, the same arguments can readily be formulated in the traditional framework of epistemic modal logic provided that appropriate fixed point formulas are either assumed as premises or added systematically in, e.g., the style of Smorynski diagonalization operators [20].[1]

M&K's original strategy in setting out the paradox was to seek a minimal set of epistemic principles which can be shown to lead to a contradiction in the presence of epistemic self-reference. In addition to **T** and **U**, they also assumed a closure principle **I** which states (roughly) that if $\varphi \to \psi$ is derivable in T, and φ is known, then so is ψ. The epistemic motivation of **T**, **U** ought to be clear: **T** expresses the factivity of knowledge (i.e. that knowledge of φ entails its truth) and **U** expresses that this feature of knowledge is known. **T** is almost universally accepted as a conceptual truth about knowledge. But in this case, i can presumably come to know that **U** is true simply by reflecting on the meaning of "knows that."

Although the justification of **T** and **U** appear to be closely linked in this way, the consensus about the *epistemic* significance of the Knower seems to be that principle **U** ought to be rejected – c.f., e.g., Anderson [1], Cross [8], Égré [9].[2] The argument typically adduced against **U** turns on the fact that the following instance of this principle

(2) $K(\ulcorner K(\ulcorner\bot\urcorner) \to \bot\urcorner) \equiv K(\ulcorner\neg K(\ulcorner\bot\urcorner)\urcorner)$

expresses that i knows that his reasoning is consistent. But as long as T is recursively axiomatizable and sufficiently strong to show that $K(x)$ shares certain features of traditional

---

[1]This latter observation glosses over the significance which Montague [18] assigned to the Knower –i.e. that since self-reference is a feature of even very weak arithmetic systems, the inconsistency of these systems with **T** and **U** shows that we must reject Quine's view that sentential operators like "knows that" and "necessarily" should be treated as predicates of sentences. But we now know that in treating these operators as modalities, we do not necessarily escape self-reference. For instance, a modal analogue to the arithmetic Fixed Point Theorem obtains for the provability logic GL in the form of the De Jongh-Sambin Fixed Point Theorem. For a discussion of the Knower in the context of modal provability logic, see [9].

arithmetic proof predicates, it follows by an argument which mirrors the proof of Gödel's Second Incompleteness Theorem that a statement like (2) cannot be derived in T as long as T is and consistent. It has been argued on this basis (e.g. by Anderson [1]) that the outer occurrence of K in **U** cannot be taken to express knowledge which is on the same 'level' as the inner occurrence.

We do not so much object to the view that **U** is problematic, as much as we find the argument just sketched for this conclusion lacking in epistemic motivation. For if self-reference really is an integral part of our everyday conception of knowledge, then it seems that a satisfactory resolution of the Knower should not be grounded in the specifically arithmetic features of the system we adopt in order to ensure the existence of statements like (1i,ii).[3] Our primary goal in this paper will thus be to adduce an independently motivated argument for rejecting **U**.

In order to do this, we will reconstruct the Knower in a system of explicit modal logic known as the Quantified Logic of Proofs (QLP) due to Fitting [10], [11]. When reformulated in this setting, it becomes clear that an additional principle beyond those assumed by M&K is required in order to derive a contradiction and that this principle is also needed in order to underwrite **U**. In demonstrating this, it will be useful to take as our starting point a streamlined version of the Knower based on a latter presentation of Montague [18]. In this case, the axioms **U** and **I** are dropped in favor of a rule – which we will refer to as the *internalization principle* – which subsumes them both:

Int $T \vdash \varphi$ $\therefore$ $T \vdash K(\ulcorner \varphi \urcorner)$

Montague's result can now be stated as follows:

Proposition 1.1. Let $T \supseteq \mathsf{Q}$, and let be $K(x)$ a predicate of $\mathcal{L}_T$ which satisfies **T** and **Int**. Then T is inconsistent.

Proof: Apply the Fixed Point Theorem to find a sentence δ such that

FP $T \vdash \delta \leftrightarrow \neg K(\ulcorner \delta \urcorner)$

We may now reason in T as follows:

1)	$T \vdash \neg K(\ulcorner \delta \urcorner) \rightarrow \delta$	**FP**
2)	$T \vdash K(\ulcorner \delta \urcorner) \rightarrow \neg \delta$	**FP**
3)	$T \vdash K(\ulcorner \delta \urcorner) \rightarrow \delta$	**T**
4)	$T \vdash \neg K(\ulcorner \delta \urcorner)$	2), 3)
5)	$T \vdash \delta$	1), 4)
6)	$T \vdash K(\ulcorner \delta \urcorner)$	**Int** 5)
7)	$T \vdash \bot$	4), 6)

---

[2] The caveat "epistemic" reflects the fact that soon after M&K's original derivation of the Knower, it was realized that the properties they assumed of $K(x)$ also held for a variety of other "truth-like" predicates such as *necessity* – c.f., e.g., Montague [18], McGee [16] and Friedman & Sheard [12]). In this broader context, arguments in favor of dropping **T** have also been advanced – cf., e.g., [17].

[3] It bears noting in this regard that the argument summarized in the previous paragraph requires that Second Incompleteness Theorem (or equivalently Löb's Theorem) is provable in T. This is straightforwardly true in the case of PA. But the situation is much more subtle for Q – cf. [5]. This highlights the extent to which Anderson's argument relies on arithmetic considerations in addition to epistemic ones.

In order to motivate the use of QLP in reconstructing this derivation, it will be useful to begin by considering the epistemic rationale for adopting **Int**. Such a principle presumably reflects the fact that if the statement φ is derivable in T, then i can come to know φ by constructing a proof of this statement himself by reasoning in T. This principle thus embodies the view that if we are to view T as a medium for reasoning about i's knowledge, then derivability in T must be understood as a means by which i can not only extend his knowledge, but also that by proving a statement φ in T, he thereby gains justification for φ. As such, we take it to be implicit in the formulation of the Knower that knowledge requires proof.

A related issue which this observation helps to expose is that since the class of statements which are derivable in T will inevitably outstrip those which i will have explicitly derived at any given time, the epistemic notion described implicitly by the hypotheses of Proposition 1.1 is not precisely that of *knowledge*, but closer to a conception of *knowability* grounded in deductive provability.[4] But since to say that φ is provable just is to say that there exists a proof of φ. This suggests that a faithful interpretation of the meaning of $K(x)$ in Proposition 1.1 ought to resemble

(3) $K(\ulcorner\varphi\urcorner) \rightleftharpoons$ there exists a proof of φ in T

One obvious means of attempting to formalize such an interpretation is to think of $K(x)$ not as a primitive predicate, but rather as defined using a conventional arithmetic proof predicate $Proof_Z(x, y)$ for Z. This would in turn suggest defining $K(x)$ as follows:

(4) $K(\ulcorner\varphi\urcorner) =_{df} \exists x Proof_Z(x, \ulcorner\varphi\urcorner)$

The problem with adopting such a definition is that its consistency with **T** and **Int** will depend quite sensitively on the details of how $Proof_Z(x, y)$ is defined and also on the strength of Z. In particular, if provably in Z, $Proof_Z(x, y)$ satisfies the Gödel-Löb-Bernays conditions on proof predicates, it follows by a well-known argument that the provability principle analogous to **T** – i.e. the so-called *reflection axiom* $\exists x Proof_Z(x, \ulcorner\varphi\urcorner) \rightarrow \varphi$ – cannot be consistently adjoined to Z.[5]

If, on the other hand, we think of epistemic self-reference as a phenomenon which arises independently of our decision to represent i's reasoning using an arithmetic theory, then nothing about the formulation of the Knower requires that the notion of provability be analyzed arithmetically. This suggests the possibility of formalizing (3) using a system for reasoning about informal proofs and provability. QLP is such a system. This theory descends from Artemov's LP (the Logic of Proofs) which in turn descends from Gödel's [13] use of the modal logic S4 to provide a provability interpretation for intuitionistic proposal logic. QLP inherits from LP statements of the form $t : \varphi$ (which are known as *explicit modalities*) subject to the intended interpretation

---

[4]This observation in turn raises the question whether the paradox can be resolved by simply abandoning **Int** or similar closure principles like **I**. Such a strategy has been advocated by Maitzen [15]. However, Cross [8] has shown that a paradox analogous to the Knower can be formulated on the basis of a knowledge predicate which fails to satisfy substantial closure conditions.

[5]Cf. Boolos [6], p. 63.

(5) $t : \varphi \rightleftharpoons t$ is a proof of φ

Herein t is a potentially complex expression known as a *proof term* intended to denote an informal proof. In addition to explicit modalities, QLP also possess first-order quantifiers intended to range over informal proofs. In this system, a statement of the form $(\exists x)x : \varphi$ thus receives the intended interpretation "there exists a proof of φ," which thereby provides us with a means of formalizing the right hand side of (3).

Our goal in the rest of this paper will be demonstrate that when the Knower is reconstructed in QLP, we discover that an additional epistemic principle is required in order to legitimate the application of **Int** at line 6) in the derivation of Proposition 1.1. Since this principle underwrites the transposition of a proof quantifier across an explicit modality, it is known as the Uniform Barcan Formula [UBF]. We will suggest in Section 3 below that there are a variety of reasons – both technical and conceptual – to regard UBF with suspicion. In addition, we will also see that it is precisely such a principle which is required in order to internalize a proof of the QLP analogue of **T** – i.e.

T$_q$ $(\exists x)x : \varphi \rightarrow \varphi$

to derive the QLP analogue of **U** – i.e.

U$_q$ $(\exists y)y : ((\exists x)x : \varphi \rightarrow \varphi)$

As such, we will argue that the source of the inconsistency highlighted by the Knower Paradox is not located in **U** itself, but rather in the principle UBF on which it depends. Such a resolution provides not only the desired epistemic rationale for rejecting **U**, but it also points towards a system wherein epistemic self-reference, factivity and a limited form of internalization may be consistently accommodated.

2   QLP and the explicit reconstruction of the Knower

The language of QLP is similar to that of propositional modal logic. However, rather than a single modal operator $\Box$, QLP possesses an infinite family of so-called *explicit modalities* consisting of a proof term t followed by the operator symbol ":" (which is intended to express "is a proof of"). The class of QLP proof terms is specified by the grammar

$$t := x_i \mid a_i(x) \mid !t \mid t_1 \; \cdot \; t_2 \mid t_1 + t_2 \; \mid (t\forall x)$$

$x_1, x_2, \ldots$ are known as *proof variables*, $a_1(x), a_2(x), \ldots$ as *primitive proof terms*. !, $\cdot$, + and $(\cdot\forall\cdot)$ denote *proof operations* respectively called *proof checker* (unary), *application* (binary), *sum* (binary) and *uniform verifier* (binary). The class of formulas and sentences of QLP is defined as follows:

Definition 2.1. If $P_0, P_1, \ldots$ are propositional letters, then the class of formulas of LP is specified by the grammar

$$\varphi := \bot \mid P_i \mid \; \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 \rightarrow \varphi_2 \mid \neg\varphi_1 \mid t : \varphi \mid (\forall x)\varphi(x) \mid (\exists x)\varphi(x).$$

The free variables of proof terms and formulas are respectively defined in the same manner as those of terms and formulas in first-order logic with the exception that in uniform verifier terms of the form $(t\forall x)$, x is considered bound. As usual, a sentence is taken to be a formula with no free variables.

We now present a Hilbert-style proof system for QLP, presenting first the axioms and then rules.

Definition 2.2. The axioms of LP are as follows:

LP1	all tautologies of classical propositional logic
LP2	$t : (\varphi \to \psi) \to (s : \varphi \to t \cdot s : \psi)$
LP3	$t : \varphi \to \varphi$
LP4	$t : \varphi \to !t : t : \varphi$
LP5	$t : \varphi \to t + s : \varphi$ and $s : \varphi \to t + s : \varphi$
QLP1	$(\forall x)\varphi(x) \to \varphi(t)$, for any proof term t that is free for x in $\varphi(x)$
QLP2	$\varphi(t) \to (\exists x)\varphi(x)$, for any proof term t that is free for x in $\varphi(x)$
QLP3	$(\forall x)(\psi \to \varphi(x)) \to (\psi \to (\forall x)\varphi(x))$, where $x \notin \mathrm{FV}(\psi)$
QLP4	$(\forall x)(\varphi(x) \to \psi) \to ((\exists x)\varphi(x) \to \psi)$, where $x \notin \mathrm{FV}(\psi)$
UBF	$(\forall x)t(x) : \varphi(x) \to (t\forall x) : (\forall x)\varphi(x)$

Axioms LP2, LP3 and LP4 will be respectively recognized as explicit versions of the S4 axioms K, T and 4 wherein $\Box$s have been replaced with explicit modalities. These axioms can be taken to display functional dependencies between occurrences of S4-$\Box$s, read in accordance with Gödel's [13] informal provability interpretation (cf. Artemov [4] for details). UBF corresponds to the Uniform Barcan Formula mentioned above. It expresses that if for all x, the proof term $t(x)$ serves as a proof of $\varphi(x)$, then the uniform verifier term $(t\forall x)$ serves as a proof of $(\forall x)\varphi(x)$.

A *primitive term specification* is a mapping $\mathcal{F}$ which assigns to each primitive proof term a a (possibly empty) set of formulas $\mathcal{F}(a)$ such that if $\varphi \in \mathcal{F}(a)$, then $\mathrm{FV}(a) = \mathrm{FV}(\varphi)$. $\mathcal{F}$ is *axiomatically injective* if for all (and only) instances φ of the QLP axioms, there exists a unique primitive proof term a such that $\varphi \in \mathcal{F}(a)$. We will henceforth assume that we are working with a fixed axiomatically injective constant specification $\mathcal{F}$.

Definition 2.3. If Γ is a finite set of formulas, then the derivability relation $\Gamma \vdash \varphi$ for QLP is now given by specifying the following rules:

RQLP1	If $\varphi \in \Gamma$, then $\Gamma \vdash \varphi$.
RQLP2	If $\Gamma \vdash \varphi \to \psi$ and $\Gamma \vdash \varphi$, then $\Gamma \vdash \psi$.
RQLP3	If $\vdash \varphi(x)$, then $\vdash (\forall x)\varphi(x)$.
RQLP4	If φ is an instance of a QLP axiom, then $\Gamma \vdash \varphi$.
RQLP5	If φ is an instance of one of LP1-LP5 and $\varphi(x_1, \ldots, x_n) \in \mathcal{F}(a(x_1, \ldots, x_n))$, then $\vdash a(x_1, \ldots, x_n) : \varphi(x_1, \ldots, x_n)$.

RQLP5 is known as the *axiom internalization rule* and is meant to reflect the idea that if $\varphi(x_1, \ldots, x_n)$ is an instance of an axiom, then $a(x_1, \ldots, x_n)$ serves as a name for a justification of this axiom provided that $\varphi(x_1, \ldots, x_n) \in \mathcal{F}(a(x_1, \ldots, x_n))$.

RQLP5 should be contrasted with the internalization rule **Int** which we have suggested above implicitly represents the fact that there is a proof of any statement derivable in T. While this does not hold in virtue of any basic rule in QLP, much of the interest of this and related systems of explicit modal logic derives from the fact that they are able to internalize their own proofs in the manner reflected in the following theorem.

Theorem 2.1. Constructive Internalization Theorem (Artemov [4], Fitting [10])
Suppose that $x_1 : \varphi_1, \ldots, x_n : \varphi_n \vdash \psi$. Then there exists a proof term $t(x_1, \ldots, x_n)$ such that

$$x_1 : \varphi_1, \ldots, x_n : \varphi_n \vdash t(x_1, \ldots, x_n) : \psi$$

Something must finally be said about the status of self-referential statements such as **FP** with respect to QLP. Since this system has no mechanism for reasoning about its own syntax, there is no reason to expect that anything like the Fixed Point Theorem holds for QLP. In fact QLP derivability bears a close relationship to $\mathsf{S4}$ derivability in a sense which is made precise by the following:

Theorem 2.2. Embedding Theorem (Fitting [10]) Define the mapping $(\cdot)^{\exists}$ between the language of $\mathsf{S4}$ and QLP as follows: $(P_i)^{\exists} = P_i$ for P_i a propositional letter; $(\cdot)^{\exists}$ commutes with propositional connectives; $(\Box A)^{\exists} = (\exists x)x : A^{\exists}$. Then

$$\mathsf{S4} \vdash A \text{ if and only if } \mathsf{QLP} \vdash A^{\exists}.$$

It is easy to see that $\mathsf{S4}$ is incompatible with the existence of a modal version of a statement like **FP** in the strong sense that $\mathsf{S4} \vdash \neg\Box(A \leftrightarrow \neg\Box A)$ for all modal formulas A.[6] One straightforward means of seeing this is to take $\Box(A \leftrightarrow \neg\Box A)$ as a reductio assumption, and proceeding as follows:

Proposition 2.1. For all A, $\mathsf{S4} \vdash \neg\Box(A \leftrightarrow \neg\Box A)$

Proof: Reasoning in $\mathsf{S4}$:

0)	$\Box(A \leftrightarrow \neg\Box A) \vdash A \leftrightarrow \neg\Box A$	T
1)	$\Box(A \leftrightarrow \neg\Box A) \vdash \neg\Box A \rightarrow A$	
2)	$\Box(A \leftrightarrow \neg\Box A) \vdash \Box A \rightarrow \neg A$	
3)	$\Box(A \leftrightarrow \neg\Box A) \vdash \Box A \rightarrow A$	T
4)	$\Box(A \leftrightarrow \neg\Box A) \vdash \neg\Box A$	2), 3)
5)	$\Box(A \leftrightarrow \neg\Box A) \vdash A$	1), 4)
6)	$\Box(A \leftrightarrow \neg\Box A) \vdash \Box A$	Necessitation 5)
7)	$\Box(A \leftrightarrow \neg\Box A) \vdash \bot$	4), 6)
8)	$\vdash \neg\Box(A \leftrightarrow \neg\Box A)$	0) - 7)

---

[6]Although the $\mathsf{4}$ axiom (i.e. transitivity) is required to justify the form of the Necessitation rule employed in Proposition (2.1) (i.e. if all formulas in Γ are of the form $\Box B$, then $\Gamma \vdash A \therefore \Gamma \vdash \Box A$) this feature of the derivation is eliminable as $\neg\Box(A \leftrightarrow \neg\Box A)$ is already a theorem of the modal logic T which contains only the normality and reflection axioms. We have elected to employ $\mathsf{S4}$ rather than T in our exposition to simplify the exposition of results about QLP such as Theorem 2.2. We have also extended the results presented here to cover versions of the Knower Paradox based on a belief operator (similar to that employed by Thomason [21]) in which an axiom similar to $\mathsf{4}$ is assumed.

It is notable that the right hand side of lines 1)-7) of this derivation have exactly the same form as the derivation given in Proposition 1.1. The foregoing derivation thus can be taken as a modal reconstruction of the Knower Paradox. Formulated in this context, it demonstrates the inconsistency of adding an instance of a modal analog of **FP** of the form $\Box(A \leftrightarrow \neg\Box A)$ together with the reflection axiom and necessitation rule of S4.[7]

It is now straightforward to transform the previous S4 derivation into a QLP derivation:

0)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash \varphi \leftrightarrow \neg(\exists x)x : \varphi$	LP3
1)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash \neg(\exists x)x : \varphi \rightarrow \varphi$	
2)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash (\exists x)x : \varphi \rightarrow \neg\varphi$	
3)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash (\exists x)x : \varphi \rightarrow \varphi$	derivable in QLP
4)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash \neg(\exists x)x : \varphi$	2), 3)
5)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash \varphi$	1), 4)
6)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash t(y) : \varphi$	for some $t(y)$ via Theorem 2.1
6')	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash (\exists x)x : \varphi$	QLP3
7)	$y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi) \vdash \bot$	4), 6')
8)	$\vdash \neg y : (\varphi \leftrightarrow \neg(\exists x)x : \varphi)$	0) - 7)
9)	$\vdash (\forall y)\neg y : [\varphi \leftrightarrow \neg(\exists x)x : \varphi]$	RQLP4
10)	$\vdash \neg(\exists y)y : [\varphi \leftrightarrow \neg(\exists x)x : \varphi]$	

With the exception of the quantifier steps 8)-10), this derivation also shares the structure of Proposition 1.1. However, as presented the argument is an enthymeme as neither steps 3) nor step 6) correspond to QLP axioms nor are they derivable from the preceding steps by QLP rules. The first of these gaps is filled in as follows:

i)	$\vdash x : \varphi \rightarrow \varphi$	LP3
ii)	$\vdash (\forall x)(x : \varphi \rightarrow \varphi)$	RQLP4
iii)	$\vdash (\forall x)(x : \varphi \rightarrow \varphi) \rightarrow ((\exists x)x : \varphi \rightarrow \varphi)$	QLP4
iv)	$\vdash (\exists x)x : \varphi \rightarrow \varphi$	2), 3)

In order to construct the term $t(y)$ used at step 6), i)-iv) must now be internalized in the manner suggested by Theorem 2.1. This may be accomplished as follows:

v)	$\vdash x : \varphi \rightarrow \varphi$	LP3
vi)	$\vdash r(x) : (x : \varphi \rightarrow \varphi)$	RQLP3
vii)	$\vdash (\forall x)r(x) : (x : \varphi \rightarrow \varphi)$	RQLP2
viii)	$\vdash (\forall x)r(x) : (x : \varphi \rightarrow \varphi) \rightarrow (r(x)\forall x) : (\forall x)(x : \varphi \rightarrow \varphi)$	UBF
ix)	$\vdash (r(x)\forall x) : (\forall x)(x : \varphi \rightarrow \varphi))$	3), 4)
x)	$\vdash q : (\forall x)(x : \varphi \rightarrow \varphi) \rightarrow ((\exists x)x : \varphi \rightarrow \varphi)$	RQLP3
xi)	$\vdash q \cdot (r(x)\forall x) : ((\exists x)x : \varphi \rightarrow \varphi)$	LP2, 5), 6)

---

[7]If after Gödel [13] we interpret the $\Box$ operator of S4 as expressing informal provability, $\Box(A \leftrightarrow \neg\Box A)$ can be interpreted as stating the informal provability of a fixed point statement. While Proposition 2.1 demonstrates that such statements are incompatible with S4, it is easy to see that no inconsistency arises if we assume the weaker statement $A \leftrightarrow \neg\Box A$ as in this case the application of the Necessitation rule at line 6) would not be licensed. This might initially seem to threaten the parallelism between Propositions 1.1 and 2.1. But note that in the context of Proposition 1.1 **FP** is not an additional principle assumed in order to derive a contradiction. Rather, it is an instance of the Diagonal Lemma which is provable in T itself. As such, the arithmetic analog of $\Box(A \leftrightarrow \neg\Box A)$ (i.e. $K(\ulcorner\delta \leftrightarrow \neg K(\ulcorner\delta\urcorner)\urcorner)$) will also be derivable in T via **Int**.

It is now routine to internalize steps 0) - 5) in the original derivation so as to construct the term $t(y)$.[8] What is of more immediate interest is that since $\mathbf{T}_q$ appears in the derivation of 10) above, its derivation must consequently be internalized to construct the term $t(y)$ at line 6). This leads to line xi), which immediately implies $\mathbf{U}_q$ via existential generalization (i.e. QLP2).

Another notable fact about the derivation v)-xi) is that it appears to depend on UBF. For although $\mathbf{T}_q$ can be derived via i)-iv), there appears to be no way of internalizing this proof without transposing the universal quantifier in vii) across the explicit modality. This is in fact the case:

Theorem 2.3. Let QLP^- denote the axiom system consisting of QLP without UBF and $\mathcal{L}_{\mathsf{QLP}^-}$ be the language of QLP without the uniform verifier symbol $(\cdot\forall\cdot)$. Then $\mathsf{QLP} \nvdash \varphi$, then

$$\mathsf{QLP}^- \nvdash (\exists y)y : ((\exists x)x : \varphi \rightarrow \varphi)$$

This result – which demonstrates the non-conservativity of QLP over QLP^- – can readily be established using the so-called *evidential* semantics for QLP presented by Fitting [10], [11] . It immediately follows that xi) in the foregoing derivations could not be derived without the aid of UBF. If we do not assume this principle, there is no way of constructing the proof term $t(y)$ required by 6). As such, we take Theorem 2.3 to demonstrate that when knowledge is understood in terms of proof in the manner we have suggested, UBF is required in order to reconstruct the Knower. What remains to be seen is what sort of independent consideration can be brought to bear on the acceptability of this principle.

3 UBF, consistency and internalization

In this section, we will discuss two distinct sorts of considerations which we take to argue *against* the epistemic acceptability of UBF. The first of these derives from an attempt to make sense of this principle relative to the informal provability interpretation of explicit modalities. The second consideration represents an attempt to understand the arithmetic significance of UBF by attempting to construct an arithmetic semantics for QLP.

Consider again the interpretation which takes an explicit modality $t : \varphi$ expresses that t is an informal proof or other form of justification of φ. This interpretation allows us to provide an epistemic motivation for axioms LP2, LP3 and LP4. For instance, if t is an informal proof of φ it then follows that φ must be true.[9] And if t denotes a proof $\varphi \rightarrow \psi$, and s a proof of φ, then it follows that there exists a proof of ψ – i.e. that compromised of

---

[8]Specifically the required form is of the form $t(y) \equiv (a_1 \cdot y) \cdot ((b \cdot (q \cdot (r(x)\forall x))) \cdot (a_2 \cdot y))$ where $r(x)$ and q are primitive proof terms which justify the formulas displayed in the derivation and a_1, a_2 and b are primitive proof terms for propositional tautologies.

[9]This of course presupposes that the notion of informal proof which we are attempting to capture is one according to which proofs provide an indefeasible form of justification for the statements which they verify. This is presumably true of mathematical proofs (both informal and formal) but not for other forms of everyday forms of evidence based on sense data, inductive or abductive inference, etc.

t, followed by s, followed by one application of MP with conclusion ψ. In QLP, such a proof is denoted by the complex proof term $t \cdot s$.[10]

Of course there is nothing special about the proof terms t and s mentioned in the preceding examples. Accordingly, many intuitively valid principles about informal proof may naturally be formulated using quantifiers – e.g.

(6)   a) $(\forall x)[x : \varphi \to \varphi]$[11]
        b) $(\forall x)(\forall y)[x : (\varphi \to \psi) \to (y : \varphi \to x \cdot y : \psi)]$

It should thus also be apparent that the quantifier axioms and rules of QLP which underwrite the derivability of such examples also express valid principles about informal proof. As such, there appears to be no conceptual difficultly in envisioning a calculus of informal provability which contains quantifiers intended to range over proofs.

Now consider the acceptability of UBF – which we reproduce here for reference – on the informal proof interpretation:

UBF $(\forall x)t(x) : \varphi(x) \to (t\forall x) : (\forall x)\varphi(x)$

Recall the gloss of this principle provided above: if for all x, the proof term $t(x)$ serves as a proof of $\varphi(x)$, then the uniform verifier term $(t\forall x)$ serves as a proof of $(\forall x)\varphi(x)$. In order to get a better impression of the meaning of the antecedent of such a statement, it is useful to consider a specific instance in which x occurs free in $\varphi(x)$. An illustrative example is provided by taking $\varphi(x) = x : P \to P$ – i.e. an instance of LP3 with P an atomic formula.

To make the case yet more concrete, suppose P symbolizes a specific empirical proposition – e.g. that a particular substance p contained cyanide. Also suppose we are considering a class of empirical tests $\mathcal{T}$ by which P might be confirmed. For instance $x_1 \in \mathcal{T}$ might test p's appearance under UV light, $x_2 \in \mathcal{T}$ p's reaction with iron sulfate, etc.[12] And suppose in addition we were concerned with verifying whether each of the tests in $\mathcal{T}$ was accurate in the sense of producing no false positives – a proposition which can be expressed in QLP as $x : P \to P$. In such a case we can presumably also envision a uniform procedure $t(x)$ which would, when presented with a specific test, produced a verification that this test was accurate. For instance, on input x_1, $t(x)$ might return a demonstration invoking principles from quantum theory, on input x_2, $t(x)$ might return a demonstrate invoking principles from electrochemistry, etc. Were such a function to exist, $(\forall x)t(x) : (x : P \to P)$ would be true.

If UBF expressed a valid principle about informal proof, then in the situation just described there must exist a single proof denoted by $(t\forall x)$ which would verify $(\forall x)(x : P \to P)$. But it certainly does not seem to follow from what we *mean* when we speak about informal proofs that such a higher-order verification must exists, even given that we have individually

---

[10]For more on the use of LP as a medium for reasoning about the informal notion of justification, see[3] and [2].

[11]Note that $(\exists x)x : \varphi \to \varphi \equiv \mathbf{T}_q$ follows from (6a) in the manner illustrated by i)-iv) above. Since the informal gloss of this statement (i.e. "If φ is provable, then φ is true") may seem fundamental to our conception of informal than which can be given for LP3, one might reasonably wonder why it is the latter principle and not the former which is adopted as an axion of QLP. The arithmetic examples below provide some insight why it is the latter which we take as to be the axiom and the former the theorem.

[12]Cf. [23].

verified that all of the tests in $\mathcal{T}$ are accurate. On this basis, it appears that UBF lacks the same degree of self evidence possessed by the other axioms of QLP.

Another sort of case can be built against UBF by considering in more detail why this principle leads to inconsistency in the presence of self-reference while the other axioms of QLP do not. In order to do this, we will now demonstrate that while there is a natural way of arithmetically interpreting the language of QLP$^-$ so that all its theorems all come out as true in the standard model N, adjoining UBF to QLP$^-$ allows us to prove statements whose interpretations are not arithmetically valid. An arithmetic interpretation is a mapping $(\cdot)^\circ$ from $\mathcal{L}_{\mathsf{QLP}\,-}$ into an arithmetic language $\mathcal{L}_a$ such that QLP$^-$ proof terms are interpreted as codes of arithmetic proofs, QLP$^-$ quantifiers are interpreted as ranging over such codes, and such that statements of the form $t : \varphi$ are interpreted as asserting that t° is the code of an arithmetic proof of φ°.

Our goal will be to formulate $(\cdot)^\circ$ in such a way that the images of the theorems of QLP$^-$ are true in the standard model N. In order to do this, we now embark on a series of definitions, borrowing as many details as possible from [4] and [6]. We assume that $\mathcal{L}_a$ contains terms $\boldsymbol{f_0}, \boldsymbol{f}_1, \ldots$ which represent all primitive recursive functions. We will fix $Z = \mathsf{PA}$ as the background arithmetic theory and also fix a multi-conclusion proof predicate $\mathrm{Prf}(x, y)$ which is provably Δ_1 in PA and such that for every arithmetic formula φ

(7) $\mathsf{PA} \vdash \varphi$ if and only if for some $n \in \mathbb{N}$, $N \models \mathrm{Prf}(\overline{n}, \overline{\ulcorner\varphi\urcorner})$.

As described in [4] we will also assume that $T(n) = \{k \mid N \models \mathrm{Prf}(\overline{n}, \overline{k})\}$ is finite and that for n, m there is a k such that $T(n) \cup T(m) \subseteq T(k)$.

It follows from the fact that $\mathrm{Prf}(x, y)$ is provably Δ_1 that the set of $\mathcal{P} = \{n \mid N \models \exists y \mathrm{Prf}(\overline{n}, y)\}$ is recursive. There is hence a computable function $p : \mathbb{N} \to \mathbb{N}$ which enumerates $\mathcal{P}$ injectively – i.e. $\mathcal{P} = \{p(n) \mid n \in \mathbb{N}\}$ and if $p(n) = p(m)$, then $n = m$. Let $\boldsymbol{p}$ be a name for a $\mathcal{L}_a$-representation of this function. Such a function will be used to interpret proof variables. In particular, if we assume that the official names for proof variables in $\mathcal{L}_{\mathsf{QLP}^-}$ are $x_0, x_1, \ldots$ and the official names for arithmetic variables in $\mathcal{L}_a$ are $y_0, y_1, \ldots$, then we set $(x_i)^\circ = \boldsymbol{p}(y_i)$. In the case that x_i occurs in a QLP$^-$ formula $\varphi(x_i)$, this will mean that y_i will occur free in $(\varphi(x_i))^\circ$ inside an arithmetic term of the form $\boldsymbol{p}(y_i)$.

Our next task is to specify an arithmetic interpretation for primitive proof terms. Recall that we are working with a fixed, axiomatically injective primitive term specification $\mathcal{F}$. Thus for each instance of a QLP$^-$ axiom φ, there is some primitive proof term a such that $\varphi \in \mathcal{F}(a)$ and $\mathrm{FV}(a) = \mathrm{FV}(\varphi)$. We wish statements of the form $a(x_1, \ldots, x_n) : \varphi(x_1, \ldots, x_n)$ to get mapped to arithmetic open sentences which are true for all numerical substitution instances of their free variables. This means that the image of a primitive proof term must itself be a function mapping numbers into codes of proofs. For instance, if $\varphi(x_1, \ldots, x_n)$ is an instance of an axiom of QLP$^-$ and we have defined $(\cdot)^\circ$ so that $(\varphi(x_1, \ldots, x_n))^\circ$ is an open sentence of $\mathcal{L}_a$ provable in PA, then the interpretation of $a(x_1, \ldots, x_n)$ must be a function $a_{\varphi(x_1, \ldots, x_n)} \varphi(x_1^\circ, \ldots, x_n^\circ)$ such that for every assignment of QLP$^-$ terms to $x_1, \ldots, x_n$, $a_{\varphi(x_1, \ldots, x_n)}$ returns the code of an arithmetic proof of the image of the corresponding substitution instance of $\varphi(x_1, \ldots, x_n)$. We will refer to such functions as *axiom functions*. Once we have verified that the axioms of QLP$^-$ are themselves valid under $(\cdot)^\circ$, then it is routine to find a family of primitive recursive functions with the desired properties.

11

Finally, we must provide a means of interpreting the proof operations $\cdot, +$ and $!$ of $\mathcal{L}_{\mathsf{QLP}^-}$. This can be accomplished in the same manner as Artemov's original arithmetic interpretation of LP via the use of three functions $m(x,y)$, $p(x,y)$ and $c(x)$. These functions are defined to operate on codes of proofs so that $m(t^\circ, s^\circ)$ returns the least godel-number of a proof containing all instances of ψ such that $\ulcorner \varphi \to \psi \urcorner \in T(t^\circ)$ and $\ulcorner \varphi \urcorner \in T(s^\circ)$, $p(t^\circ, s^\circ)$ returns the least godel number of a proof containing $T(t^\circ)$ and $T(s^\circ)$ and $c(t^\circ)$ returns the least godel-number of a proof k which has $\mathrm{Prf}(\overline{t^\circ}, \ulcorner \varphi \urcorner) \in T(k)$ for all $\ulcorner \varphi \urcorner \in T(t^\circ)$. As is shown in [4], these functions are primitive recursive and will hence be represented by arithmetic terms $\boldsymbol{m}, \boldsymbol{a}$ and $\boldsymbol{c}$.

Putting these pieces together, we may now specify how $(\cdot)^\circ$ is defined on proof terms:
- $x_i^\circ = \boldsymbol{p}(y_i)$
- if $\varphi(x_1, \ldots, x_n) \in \mathcal{F}(a(x_1, \ldots, x_n))$, then $(a(x_1, \ldots, x_n))^\circ = \boldsymbol{a}_{\varphi(x_1, \ldots, x_n)}(x_1^\circ, \ldots, x_n^\circ)$
- $(t \cdot s)^\circ = \boldsymbol{m}(t^\circ, s^\circ), (t + s)^\circ = \boldsymbol{a}(t^\circ, s^\circ), (!t)^\circ = \boldsymbol{c}(t^\circ)$

Such an interpretation may be extended to the entire language of $\mathcal{L}_{\mathsf{QLP}^-}$ as follows:
- $(P_i)^\circ$ is some fixed closed sentence of $\mathcal{L}_a$
- $(\varphi \to \psi)^\circ = \varphi^\circ \to \psi^\circ, (\neg \varphi)^\circ = \neg \varphi^\circ$
- $((\forall x)\varphi)^\circ = (\forall x)\varphi^\circ, ((\exists x)\varphi)^\circ = (\exists x)\varphi^\circ$
- $t : \varphi(x_{k_1}, \ldots, x_{k_m}) =$
$$\mathrm{Prf}(t^\circ, su(x_{k_m}^\circ, \overline{k}_m, \ldots, su(x_2^\circ, \overline{k}_2, su(x_1^\circ, \overline{k}_1, \ulcorner \varphi(x_{k_1}^\circ, \ldots, x_{k_m}^\circ)\urcorner)) \ldots))$$
See [6] for the definition of $su(y_1, \ldots, y_n)$.

In order to illustrate the manner in which the axiom functions are defined, it will be useful to consider an example. Let P be a propositional letter and consider again the following instance of axiom LP3:

(8) $x : P \to P$

First note that the image of (8) under $(\cdot)^\circ$ is

(9) $\mathrm{Prf}(\boldsymbol{p}(y), \overline{\ulcorner P^\circ \urcorner}) \to P^\circ$

where y is an arithmetic variable. Although this is an open arithmetic sentence, it is easy to see that for all $n \in \mathbb{N}$,

(10) $\mathsf{PA} \vdash \mathrm{Prf}(\boldsymbol{p}(\overline{n}), \ulcorner P^\circ \urcorner) \to P^\circ$

For note that we have either i) $N \models \mathrm{Prf}(\boldsymbol{p}(\overline{n}), \ulcorner P^\circ \urcorner)$ or ii) $N \not\models \mathrm{Prf}(\boldsymbol{p}(\overline{n}), \ulcorner P^\circ \urcorner)$. In the first case, $\mathrm{Prf}(\boldsymbol{p}(\overline{n}), \ulcorner P^\circ \urcorner)$ is a true Δ_1-sentence and hence $\boldsymbol{p}(\overline{n})^N$ is indeed the godel number of a proof of P° in PA. Hence $\mathsf{PA} \vdash P^\circ$, meaning that (10) holds. In the second case, $\mathrm{Prf}(\boldsymbol{p}(\overline{n}), \ulcorner P^\circ \urcorner)$ is a false Δ_1-sentence and hence $\mathsf{PA} \vdash \neg \mathrm{Prf}(\boldsymbol{p}(\overline{n}), \ulcorner P^\circ \urcorner)$ and (10) again holds.

The foregoing reasoning is uniform in n in the sense that not only is each substitution instance of (10) provable, but the argument just rehearsed demonstrating this fact can itself be formalized in PA. From this it follows that there is a computable function which for every n, returns the godel number of a proof demonstrating (10) in PA. It is this function which we take as the axiom function $a_{x:P \to P}$. Using the foregoing example as a model, it is now routine to show that the interpretations of the other QLP^- axioms are all true in N and that the rules preserve truth. We thus have following:

Proposition 3.1. (Arithmetic soundness of QLP^-) If $\mathsf{QLP}^- \vdash \varphi$, then $N \models \varphi^\circ$.

The foregoing example also foreshadows why it is not possible to extend an arithmetic interpretation to all of QLP. For note that taking $\varphi \equiv \bot$ in the derivation v)-xi) above shows that

(11) $\mathsf{QLP} \vdash t : (\forall x)(x : \bot \to \bot)$

where $t \equiv q \cdot (a_{x:\bot \to \bot} \forall x)$ contains the uniform verifier symbol. Were we able to extend the given interpretation of QLP^- to the full language of QLP including the uniform verifier symbol, then this statement would get mapped to the $\mathcal{L}_a$ statement

(12) $\mathrm{Prf}(t^\circ, \ulcorner (\forall y) \neg \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner) \urcorner)$

where t° corresponded to a closed arithmetic term. (12) would hence express that $(t^\circ)^N \in \mathbb{N}$ is the godel number of a proof of $(\forall y) \neg \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner)$ in PA. But assuming that PA is consistent, the existence of such a number in the standard model would violate Gödel's Second Incompleteness Theorem.

These observations can readily be put together with the internal provability of Löb's Theorem in PA to show the following:

Theorem 3.1. Let $S = \{\varphi^\circ \mid \varphi \in \mathcal{L}_{\mathsf{QLP}^-}$ and $\mathsf{QLP} \vdash \varphi\}$. $\mathsf{PA} \cup S$ is inconsistent.

Proof. It follows from (11) that

(13) $\mathsf{QLP} \vdash (\exists y) y : (\forall x) \neg x : \bot$

The arithmetic interpretation of (13) – i.e.

(14) $(\exists y) \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner (\forall y) \neg \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner) \urcorner)$

is thus in S. However, an easy modification of the traditional proof of Löb's Theorem (cf., e.g, [6], p. 56) shows that

(15) $\mathsf{PA} \vdash (\exists y) \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner (\forall y) \neg \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner) \urcorner) \to \exists y \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner)$

Since (14) is thus also in S, it follows that $\mathsf{PA} \cup S \vdash \exists y \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner)$. But since $\mathsf{QLP} \vdash (\exists x) x : \bot \to \bot$, $S \vdash ((\exists x) x : \bot \to \bot)^\circ = \exists y \mathrm{Prf}(\boldsymbol{p}(y), \ulcorner \bot \urcorner) \to \bot$. Putting these two facts together, we have $\mathsf{PA} \cup S \vdash \bot$.

$\square$

This result demonstrates that even if we restrict attention to statements not containing the uniform verifier symbol itself, the arithmetic interpretations of statements provable with the aid of UBF is still inconsistent with PA (or, more generally, any arithmetic theory strong enough to derive Löb's Theorem).

Theorem 3.1 is obviously based on the same sorts of arithmetic considerations which motivate Anderson's resolution of the Knower discussed in Section 1. As such we do not take this result to condemn UBF on epistemic grounds alone. However, the proof of this result does shine some light on the relationship which this principle bears to consistency statements. For consider the following instance of UBF:

13

(16) $(\forall x)r(x) : (x : \bot \to \bot) \to \langle r(x)\forall x \rangle : (\forall x)(x : \bot \to \bot)$

As we have seen, the antecedent of this conditional is derivable in QLP^- and receives the arithmetical interpretation

(17) $\forall y[\mathrm{Prf}(a_{x:\bot\to\bot}(y), \ulcorner \mathrm{Prf}(y, \ulcorner \bot \urcorner) \to \bot \urcorner]$

Assuming that PA is consistent, this statement expresses a fact which is *true* in the standard model – i.e. for all codes of proofs y in the standard model, there exists a proof (which may be found uniformly), that y is not a proof of $\bot$. However, the consequent of (16) receives (12) as its arithmetic interpretation. As we have seen, this statement expresses that PA proves its own consistency and is hence *false* in the standard model. Thus while UBF does not itself express either consistency or the provability of consistency itself, it is responsible for mediating an inference from an expression of one to an expression of the other in the language of QLP.

We take the foregoing considerations to suggest that dropping UBF from the system we use to reason about informal provability is a well motivated means of resolving the Knower Paradox. As is clear from Theorem 2.3, the cost of this omission is losing the general validity of the Constructive Internalization Theorem, a fact which was already noted by Fitting [10]. However, even without UBF in the system, it is easily verified that that Constructive Internalization still applies for all statements derivable in QLP^- without the Universal Generalization rule RQLP3. In addition it may be shown that the existence of (provably) self-referential formulas having the form of (1ii) is compatible with QLP^- in the following sense:

Proposition 3.2. The theory

$$\mathsf{QLP}^- + \{(\exists y)y : (P \leftrightarrow \neg(\exists x)x : P) \mid P \text{ atomic}\}$$

is consistent.[13]

Abandoning UBF thus allows us to formulate a system in which factivity and self-reference are preserved, while retaining internalization for a wide range of formulas.

References

[1] C. A. Anderson. The paradox of the knower. *The Journal of Philosophy*, 80(6):338–355, 1983.

[2] S. Artemov. Justification logic. Technical report 2007019, CUNY Ph.D. Program in Computer Science, 2007.

[3] S. Artemov and E. Nogina. On epistemic logic with justification. In R. Meyden, editor, *Theoretical Aspects of Rationality and Knowledge. Proceedings of the Tenth Conference (TARK 2005)*, pages 279–294, 2005.

---

[13]The proof of this Proposition, like that of Proposition 2.3, required the use of a relational semantics for QLP presented by Fitting [10], [11] which we have elected not to reproduce here. We expect to extend to Proposition 3.2 by showing that fixed point formulas for arbitrary statements can be consistently added to QLP^- via a device like Smorynski's diagonalization operator [20].

[4] S. N. Artemov. Explicit provability and constructive semantics. *The Bulletin of Symbolic Logic*, 7(1):1–36, 2001.

[5] A. Bezboruah and J. C. Shepherdson. Gödel's second incompleteness theorem for Q. *The Journal of Symbolic Logic*, 41(2):503–512, 1976.

[6] G. Boolos. *The Logic of Provability*. Cambridge University Press, 1993.

[7] T. Burge. Epistemic paradox. *Journal of Philosophy*, 81:5–29, 1984.

[8] C. Cross. The paradox of the knower without epistemic closure. *Mind, New Series*, 110(438):319–333, 2001.

[9] P. Égré. The knower paradox in the light of provability interpretations of modal logic. *Journal of Logic, Language, and Information*, 14(1):13–48, 2005.

[10] M. Fitting. Quantified LP. Technical report, CUNY Ph.D. Program in Computer Science Technical Report TR-2004019, 2004.

[11] M. Fitting. The logic of proofs, semantically. *Ann. Pure Appl. Logic*, 132(1):1–25, 2005.

[12] H. Friedman and M. Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.

[13] K. Gödel. An interpertation of intuitionistic propositional calculus. In S. Feferman et al., editors, *Kurt Gödel. Collected Works*, volume I, pages 144–195. Oxford University Press, 1986.

[14] R. Koons. *Paradoxes of belief and strategic rationality*. Cambridge University Press, 1992.

[15] S. Maitzen. The knower paradox and epistemic closure. *Synthese*, 114(337-354), 1998.

[16] V. McGee. How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14:399–410, 1985.

[17] V. McGee. *Truth, Vaugeness and Paradox*. Hackett Publishing Company, 1991.

[18] R. Montague. Syntactic treatment of modalities, with corollaries on reflexion princples and finite axiomatizability. In R. Thomason, editor, *Formal Philosophy*. Yale University Press, 1974.

[19] R. Montague and D. Kaplan. A paradox regained. *Notre Dame Journal of Formal Logic*, 1:79–90, 1960.

[20] C. Smorynski. *Self-Reference and Modal Logic*. Springer-Verlag, Berlin, 1985.

[21] R. Thomason. A note of the syntactic treatment of modality. *Synthese*, 44:371–395, 1980.

[22] T. Tymoczko. An unsolved puzzel about knowledge. *Philosophical Quarterly*, 34:437–458, 1984.

[23] Wikipedia. Cyanide. http://en.wikipedia.org/wiki/Cyanide, 2009.