

Knowledge, Proof and the Knower

Walter Dean & Hidenori Kurokawa

(University of Warwick and CUNY) & CUNY

FEW 2009, CMU

The original Knower (Montague & Kaplan 1960)

- ▶ $\mathcal{L}_T \supseteq \mathcal{L}_a \cup \{K(x)\}$
- ▶ T an \mathcal{L}_T theory s.t. $T \supseteq Q$
- ▶ It follows that for every $\varphi(x) \in \text{Form}_{\mathcal{L}_T}$ there is γ s.t.

$$T \vdash \gamma \leftrightarrow \varphi(\ulcorner \gamma \urcorner)$$

- ▶ In particular, there is δ such that

$$\text{(FPm)} \quad T \vdash \delta \leftrightarrow K(\ulcorner \neg \delta \urcorner).$$

- ▶ Additionally $K(x)$ is taken to satisfy

$$\text{(T)} \quad K(\ulcorner \varphi \urcorner) \rightarrow \varphi$$

$$\text{(U)} \quad K(\ulcorner K(\ulcorner \varphi \urcorner) \urcorner) \rightarrow \varphi$$

$$\text{(I)} \quad (K(\ulcorner \varphi \urcorner) \wedge I(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner)) \rightarrow K(\ulcorner \psi \urcorner)$$

$$\text{where } T \vdash I(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner) \text{ iff } T \vdash \varphi \rightarrow \psi$$

Original Knower (derivation)

Proposition

T is inconsistent.

- | | | |
|-----|---|------------|
| 1) | $T \vdash \delta \leftrightarrow \neg K(\ulcorner \neg \delta \urcorner)$ | FPm |
| 2) | $T \vdash K(\ulcorner \neg \delta \urcorner) \rightarrow \neg \delta$ | T |
| 3) | $T \vdash \delta \rightarrow \neg \delta$ | 1), 2) |
| 4) | $T \vdash \neg \delta$ | 3) |
| 5) | $T \vdash I(\ulcorner K(\ulcorner \neg \delta \urcorner) \rightarrow \neg \delta \urcorner, \ulcorner \neg \delta \urcorner)$ | 1) - 4) |
| 6) | $T \vdash K(\ulcorner K(\ulcorner \neg \delta \urcorner) \rightarrow \neg \delta \urcorner)$ | U |
| 7) | $T \vdash [K(\ulcorner K(\ulcorner \neg \delta \urcorner) \rightarrow \neg \delta \urcorner)$
$\quad \wedge I(\ulcorner K(\ulcorner \neg \delta \urcorner) \rightarrow \neg \delta \urcorner, \ulcorner \neg \delta \urcorner))] \rightarrow K(\ulcorner \neg \delta \urcorner)$ | I |
| 8) | $T \vdash K(\ulcorner \neg \delta \urcorner)$ | 5), 6), 7) |
| 9) | $T \vdash \delta$ | 1), 8) |
| 10) | $T \vdash \perp$ | 4), 9) |

Diagnosing the Paradox

theorist(s)	diagnosis	cure
Myhill [60]	informal provability is not expressible in \mathcal{L}_a	Forbid iterated modalities
Montague [63]	“syntactic treatment of modality”	treat modalities as sentential operators
Maitzen [98]	epistemic closure	Reject I -- i.e. $(K(\ulcorner\varphi\urcorner) \wedge I(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)) \rightarrow K(\ulcorner\psi\urcorner)$
McGee [91], Horsten [02]	complicated	Reject T -- i.e. $K(\ulcorner\varphi\urcorner) \rightarrow \varphi$
Anderson [83], Cross [01] Égré [05]	process of elimination (?)	Reject U -- i.e. $K(\ulcorner K(\ulcorner\varphi\urcorner) \rightarrow \varphi\urcorner)$

A philosophical puzzle

T “knowledge entails truth”

$$K(\ulcorner \varphi \urcorner) \rightarrow \varphi$$

U “T is known”

$$K(\ulcorner K(\ulcorner \varphi \urcorner) \rightarrow \varphi \urcorner)$$

- ▶ Consensus view: retain **T** and **I**, reject **U**.
- ▶ **T** expresses (something like) a conceptual truth about knowledge.
- ▶ **U** appears to follow from **T** by reflecting on the meaning of “knows that.”
- ▶ So how can we accept **T** and reject **U**?

Plan

- I) a simplified Knower
- II) reconstruction in modal logic (S4)
- III) reconstruction in **explicit** modal logic (QLP)
- IV) isolation of a new principle (UBF) **needed to derive U from T**
- V) arguments against UBF

A simplified Knower (\approx Montague 1963)

- ▶ $\mathcal{L}_T \supseteq \mathcal{L}_a \cup \{K(x)\}$
- ▶ T an \mathcal{L}_T theory s.t. $T \supseteq \text{PA}$
- ▶ There is δ such that
(FP) $T \vdash \delta \leftrightarrow \neg K(\ulcorner \delta \urcorner)$
- ▶ For ease of reference
(FP1) $T \vdash \neg K(\ulcorner \delta \urcorner) \rightarrow \delta$
(FP2) $T \vdash K(\ulcorner \delta \urcorner) \rightarrow \neg \delta$
- ▶ We additionally suppose $K(x)$ satisfies
(T) $T \vdash K(\ulcorner \varphi \urcorner) \rightarrow \varphi$
(Int) If $T \vdash \varphi$, then $T \vdash K(\ulcorner \varphi \urcorner)$
- ▶ We will refer to **Int** as an **internalization principle**.
- ▶ NB: It resembles a traditional modal necessitation rule.

A simplified Knower (derivation)

Proposition

\mathcal{T} is inconsistent.

- | | | |
|----|--|---------------|
| 1) | $\mathcal{T} \vdash \neg K(\ulcorner \delta \urcorner) \rightarrow \delta$ | FP1 |
| 2) | $\mathcal{T} \vdash K(\ulcorner \delta \urcorner) \rightarrow \neg \delta$ | FP2 |
| 3) | $\mathcal{T} \vdash K(\ulcorner \delta \urcorner) \rightarrow \delta$ | T |
| 4) | $\mathcal{T} \vdash \neg K(\ulcorner \delta \urcorner)$ | 2), 3) |
| 5) | $\mathcal{T} \vdash \delta$ | 1), 4) |
| 6) | $\mathcal{T} \vdash K(\ulcorner \delta \urcorner)$ | Int 5) |
| 7) | $\mathcal{T} \vdash \perp$ | 4), 6) |

Reasoning about knowledge as an operator

- ▶ The reasoning used in the Knower can be reconstructed if K is treated as a **sentential operator** rather than a predicate.
- ▶ This is the conventional approach of *epistemic logic* (Hintikka, Fagin et al.).
- ▶ Modal analogues to **T**, **U**, **I** and **Int**:
 - T** $\Box F \rightarrow F$
 - U** $\Box(\Box F \rightarrow F)$
 - K** $\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G)$
 - Nec** $\vdash F \quad \therefore \vdash \Box F$
- ▶ But what is the status of **FP** on the operator interpretation?

$\Box\varphi \Leftrightarrow \varphi$ is knowable



$\Box\varphi \Leftrightarrow \varphi$ is provable

formal (i.e. in PA)

informal

GL (Smorynski, Boolos)

S4 (Gödel)

K: $\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G)$

4: $\Box F \rightarrow \Box\Box F$

L: $\Box(\Box F \rightarrow F) \rightarrow \Box F$

Nec: $\vdash F \therefore \vdash \Box F$

Reflection: **NO**

GL + **T** is inconsistent

Fixed points: **YES**

If P is modalized in Θ , then $\exists D$ s.t.

$GL \vdash D \leftrightarrow \Theta[D/P]$

K: $\Box(F \rightarrow G) \rightarrow (\Box F \rightarrow \Box G)$

4: $\Box F \rightarrow \Box\Box F$

T: $\Box F \rightarrow F$

Nec: $\vdash F \therefore \vdash \Box F$

Reflection: **YES**

T is an axiom

Fixed points: **NO**

$S4 \vdash \neg\Box(F \leftrightarrow \neg\Box F)$ for all F

The simplified Krower in S4

0)	$\Box(F \leftrightarrow \neg\Box F)$	$\vdash F \leftrightarrow \neg\Box F$	T
1)	"	$\vdash \neg\Box F \rightarrow F$	
2)	"	$\vdash \Box F \rightarrow \neg F$	
3)	"	$\vdash \Box F \rightarrow F$	T
4)	"	$\vdash \neg\Box F$	2), 3)
5)	"	$\vdash F$	1), 4)
6)	"	$\vdash \Box F$	Nec 5)
7)	"	$\vdash \perp$	4), 6)
8)		$\vdash \neg\Box(F \leftrightarrow \neg\Box F)$	0) - 7)

Proposition

For all F , $S4 \vdash \neg\Box(F \leftrightarrow \neg\Box F)$

Knowledge and justification

- ▶ Classically: F is known (by agent i) \Leftrightarrow
 - i) F is true
 - ii) i believes F
 - iii) i is **justified** in believing F
 - iv) ...
- ▶ Claim: the notion of knowledge relevant to the Knower is **knowledge in virtue of proof**.
- ▶ Why? Because derivability in T is intended to represent i 's own (idealized) deductive capacities.
- ▶ This is implicit in the **I** and **Int**.
- ▶ Question: What happens to the Knower if we try to represent justifications explicitly?

Making justification explicit

- ▶ Q: What are “**justifications**”?
- ▶ A: In the context of the Knower, it is reasonable to identify them with *mathematical proofs*.
- ▶ t justifies $\varphi \iff \text{Proof}_T(\ulcorner t \urcorner, \ulcorner \varphi \urcorner)$.
- ▶ **Explicit modalities:** $t : F \iff$ “ t justifies F ”
- ▶ Arithmetic interpretation: $(t : F)^* = \text{Proof}_T(\ulcorner t^* \urcorner, \ulcorner F^* \urcorner)$
- ▶ For $T = \text{PA}$, this interpretation leads to the **Logic of Proofs** [LP] (Artemov [01]).
- ▶ We’ll work in a quantified extension of this system know as the **Quantified Logic of Proof** [QLP] (Fitting [04],[05]).

QLP (language)

- ▶ The class of QLP proof terms $Term_{QLP}$ is given by

$$t := x_i \mid a_i(x_{k_1}, \dots, x_{k_n}) \mid !t \mid t_1 \cdot t_2 \mid t_1 + t_2 \mid \langle t \forall x \rangle$$

- ▶ x_1, x_2, x_3, \dots are *proof variables*
 - ▶ $a_1(\vec{x}), a_2(\vec{x}), a_3(\vec{x}), \dots$ are *primitive proof terms*
 - ▶ $!, \cdot, +$ and $\langle \cdot \forall \cdot \rangle$ are *proof operations*
- ▶ The class $Form_{QLP}$ of QLP formulas is given by

$$\varphi := P_i \mid F \wedge G \mid F \vee G \mid F \rightarrow G \mid \neg F \mid t : F \mid (\forall x)F \mid (\exists x)F$$

- ▶ Some characteristic formulas:

- ▶ $a : ((F \wedge G) \rightarrow G)$ (a justifies $F \wedge G \rightarrow G$)
- ▶ $!a : (a : ((F \wedge G) \rightarrow G))$ ($!a$ justifies $a : (F \wedge G \rightarrow G)$)
- ▶ $x : (F \rightarrow G) \rightarrow (y : F \rightarrow x \cdot y : G)$
- ▶ $b(x, y) : (x : (F \rightarrow G) \rightarrow (y : F \rightarrow x \cdot y : G))$
- ▶ $(\forall x)(\forall y)[x : (F \rightarrow G) \rightarrow (y : F \rightarrow x \cdot y : G)]$

QLP (axioms)

LP1 all tautologies of classical propositional logic

LP2 $t : (F \rightarrow G) \rightarrow (s : F \rightarrow t \cdot s : G)$

LP3 $t : F \rightarrow F$

LP4 $t : F \rightarrow !t : t : F$

LP5 $t : F \rightarrow t + s : F$ and $s : F \rightarrow t + s : F$

QLP1 $(\forall x)F(x) \rightarrow F(t)$

QLP2 $(\forall x)(F \rightarrow G(x)) \rightarrow (F \rightarrow (\forall x)G(x))$

QLP3 $F(t) \rightarrow (\exists x)F(x)$

QLP4 $(\forall x)(F(x) \rightarrow G) \rightarrow ((\exists x)F(x) \rightarrow G)$

UBF $(\forall x)t(x) : F(x) \rightarrow \langle t\forall x \rangle : (\forall x)F(x)$

- ▶ Usual definition of $FV(F)$.
- ▶ Usual free-variable restrictions for QLP1-QLP4.

QLP (rules)

- ▶ A *primitive term specification* is a mapping \mathcal{F} s.t. $\mathcal{P}(a(\vec{x}))$ is set of formulas $\mathcal{P}(a)$ such that if $F(\vec{x}) \in \mathcal{P}(a(\vec{x}))$, then $FV(a) = FV(\varphi)$. For all axioms $F(\vec{x})$, there is $a(\vec{x})$ s.t. $F(\vec{x}) \in \mathcal{P}(a(\vec{x}))$.
- ▶ Idea: if $F(\vec{x}) \in \mathcal{P}(a(\vec{x}))$, $a(\vec{x})$ serves a **name** for the axiom $F(\vec{x})$.
- ▶ QLP rules:
 - ▶ Modus Ponens
 - ▶ Axiom Internalization: if $F(\vec{x}) \in \mathcal{P}(a(\vec{x}))$, then $\vdash a(\vec{x}) : F(\vec{x})$
 - ▶ Universal Generalization: $\vdash F(x) \quad \therefore \vdash (\forall x)F(x)$.
- ▶ A derived rule:
 - ▶ **JUG**: $\vdash t(x) : F(x) \quad \therefore \vdash \langle t(x)\forall x \rangle : (\forall x)F(x)$

Necessitation vs. constructive internalization

- ▶ Traditional necessitation:
 - ▶ $\vdash F \therefore \vdash \Box F$
 - ▶ idea: if F is derivable, then F is knowable/necessary/true
- ▶ Necessitation in QLP:
 - ▶ necessitation rule only for axioms

Proposition (Constructive Internalization Theorem)

If $\vdash F$, then there exists $t \in \mathit{Term}_{QLP}$ s.t. $\vdash t : F$.

- ▶ idea: if F is derivable in QLP, then there exists of F we can construct internally in QLP
- ▶ the axiom $t : (F \rightarrow G) \rightarrow (s : F \rightarrow t \cdot s : G)$ is used to internalize MP step
- ▶ UBF (or JUG) is used to internalize UG steps

Reconstructing the Knower in QLP (1)

Theorem (Realization (Artemov))

If $S4 \vdash F$, then there is an r s.t. $LP \vdash (F)^r$

where $(\cdot)^r$ uniformly replaces \Box s with terms $t \in \text{Term}_{LP}$.

Theorem (Embedding (Fitting))

If $S4 \vdash F$, then $QLP \vdash (F)^\exists$

where $(\Box F)^\exists = (\exists x)x : F^\exists$.

- ▶ So we should expect QLP to be incompatible with FPs.
- ▶ $S4 \vdash \neg\Box(F \leftrightarrow \neg\Box F) \implies$
 $QLP \vdash \neg(\exists y)y : [F \leftrightarrow \neg(\exists x)x : F]$

Reconstructing the Knower in QLP (2)

0)	$y : (F \leftrightarrow \neg(\exists x)x : F)$	$\vdash F \leftrightarrow \neg(\exists x)x : F$	LP3
1)	“	$\vdash \neg(\exists x)x : F \rightarrow F$	
2)	“	$\vdash (\exists x)x : F \rightarrow \neg F$	
3)	“	$\vdash (\exists x)x : F \rightarrow F$	derivable in QLP
4)	“	$\vdash \neg(\exists x)x : F$	2), 3)
5)	“	$\vdash F$	1), 4)
6)	“	$\vdash t(y) : F$	for some $t(y)$ (via CIT)
6')	“	$\vdash (\exists x)x : F$	QLP3
7)	“	$\vdash \perp$	4), 6')
8)		$\vdash \neg y : (F \leftrightarrow \neg(\exists x)x : F)$	0) - 7)
9)		$\vdash (\forall y)\neg y : [F \leftrightarrow \neg(\exists x)x : F]$	UG
10)		$\vdash \neg(\exists y)y : [F \leftrightarrow \neg(\exists x)x : F]$	

Proposition

For all F , QLP $\vdash \neg(\exists y)y : [F \leftrightarrow \neg(\exists x)x : F]$

Reconstructing the Knower in QLP (3)

0) $y : (F \leftrightarrow \neg(\exists x)x : F) \vdash F \leftrightarrow \neg(\exists x)x : F$

1) “ $\vdash \neg(\exists x)x : F \rightarrow F$

2) “ $\vdash (\exists x)x : F \rightarrow \neg F$

3) “ $\vdash (\exists x)x : F \rightarrow F$ derivable in QLP

4) “ $\vdash \neg(\exists x)x : F$

5) “ $\vdash F$

6) “ $\vdash t(y) : F$ for some $t(y)$ (via CIT)

6') “ $\vdash (\exists x)x : F$

7) “ $\vdash \perp$

8) $\vdash \neg y : (F \leftrightarrow \neg(\exists x)x : F)$

9) $\vdash (\forall y)\neg y : [F \leftrightarrow \neg(\exists x)x : F]$

10) $\vdash \neg(\exists y)y : [F \leftrightarrow \neg(\exists x)x : F]$

Reconstructing the Knower in QLP (4)

1) $\vdash x : F \rightarrow F$	LP3 (explicit reflection)
2) $\vdash (\forall x)(x : F \rightarrow F)$	UG
3) $\vdash (\forall x)(x : F \rightarrow F) \rightarrow ((\exists x)x : F \rightarrow F)$	QLP4
4) $\vdash (\exists x)(x : F \rightarrow F)$	

1) $\vdash x : F \rightarrow F$	LP3
2) $\vdash r(x) : (x : F \rightarrow F)$	axiom nec.
3) $\vdash (\forall x)r(x) : (x : F \rightarrow F)$	UG
4) $\vdash (\forall x)r(x) : (x : F \rightarrow F) \rightarrow \langle r(x)\forall x \rangle : (\forall x)(x : F \rightarrow F)$	UBF
5) $\vdash \langle r(x)\forall x \rangle : (\forall x)(x : F \rightarrow F)$	3),4) ¹
6) $\vdash q : (\forall x)(x : F \rightarrow F) \rightarrow ((\exists x)x : F \rightarrow F)$	axiom nec.
7) $\vdash q \cdot \langle r(x)\forall x \rangle : ((\exists x)x : F \rightarrow F)$	LP2 5), 6)

¹We can get 5) from 2) via JUG:

$\vdash r(x) : (x : F \rightarrow F) \quad \therefore \vdash \langle r(x)\forall x \rangle : (\forall x)(x : F \rightarrow F)$

Reconstructing the Knower in QLP (6)

- ▶ We need to find $t(y)$ s.t. $y : (F \leftrightarrow \neg(\exists x)x : F) \vdash t(y) : F$.
- ▶ From above: $\vdash q \cdot \langle r(x)\forall x \rangle : ((\exists x)x : F \rightarrow F)$.
- ▶ We can take $t(y) \equiv (a_1 \cdot y) \cdot ((b \cdot (q \cdot \langle r(x)\forall x \rangle))) \cdot (a_2 \cdot y)$
(a_1, a_2 and b are constants for tautologies).

- ▶ Parallels:

$$\mathbf{T}_q \quad (\exists x)x : F \rightarrow F \quad \approx \quad \Box F \rightarrow F \quad (\text{i.e. } \mathbf{T})$$

$$\mathbf{U}_q \quad (\exists y)y : ((\exists x)x : F \rightarrow F) \quad \approx \quad \Box(\Box F \rightarrow F) \quad (\text{i.e. } \mathbf{U})$$

- ▶ In the arithmetic and modal settings, **U** and **Int/Nec** are primitive.
- ▶ In QLP, \mathbf{U}_q must be derived by constructive internalization.
- ▶ This appears to **require UBF** (or JUG) ...

UBF, \mathbf{U}_q and self-reference

- ▶ $\mathbf{U}_q \quad (\exists y)y : ((\exists x)x : F \rightarrow F) \approx \Box(\Box F \rightarrow F) \quad (\text{i.e. } \mathbf{U})$
- ▶ $\text{QLP}^- := \text{QLP} - \text{UBF}, \mathcal{L}_{\text{QLP}^-} = \mathcal{L}_{\text{QLP}} - \langle \cdot \forall \cdot \rangle$
- ▶ Using *Fitting semantics* for QLP we can show:

Proposition

If $\text{QLP} \not\vdash F$, then $\text{QLP}^- \not\vdash (\exists y)y : ((\exists x)x : F \rightarrow F)$

Proposition

For any propositional letter P ,

$\text{QLP}^- + (\exists y)y : (P \leftrightarrow \neg(\exists x)x : P)$ is consistent.

The arithmetic case against UBF and JUG

- (1) $\vdash r(x) : (x : \perp \rightarrow \perp)$
- (2) $\vdash (\forall x)r(x) : (x : \perp \rightarrow \perp)$
- (3) $\vdash (\forall x)r(x) : (x : \perp \rightarrow \perp) \rightarrow \langle r(x)\forall x \rangle : (\forall x)(x : \perp \rightarrow \perp)$ UBF
- (4) $\vdash \langle r(x)\forall x \rangle : (\forall x)(x : \perp \rightarrow \perp)$ 2), 3) via UBF or 1) via JUG

► Arithmetic interpretation of 1-2):

$$1^*) \quad \forall x[\text{Proof}(r^*(x), \ulcorner \text{Proof}(\dot{x}, \ulcorner \perp \urcorner) \rightarrow \perp \urcorner)]$$

► “For every $n \in \mathbb{N}$, we can prove that n is not a proof of \perp .”

► 1^* is **true** in N (presuming Z is **consistent**).

► Arithmetic interpretation of 3):

$$4^*) \quad \text{Proof}(\langle r(x)\forall x \rangle^*, \ulcorner \forall x[\text{Proof}(x, \ulcorner \perp \urcorner) \rightarrow \perp] \urcorner)$$

► “The number $\langle r(x)\forall x \rangle^*$ **is a proof that Z is consistent.**”

► 4^* is **false** in N because of Gödel’s Second Incompleteness Theorem.

The conceptual case against UBF

- 1) $x : P \rightarrow P$
- 2) $r(x) : (x : P \rightarrow P)$
- 3) $(\forall x)r(x) : (x : P \rightarrow P)$
- 4) $\langle r(x)\forall x \rangle : (\forall x)(x : P \rightarrow P)$

- ▶ Claim: 1)-3) can all be true and 4) false.
- ▶ $P =$ Substance s contains cyanide.
- ▶ Proof variables denote chemical tests in a domain \mathcal{D}_1 .
- ▶ It could be that all $t \in \mathcal{D}_1$ are **truth entailing** – i.e. such that $t : P \rightarrow P$ holds – and that we can prove this for each test we encounter.
- ▶ But there might be a larger domain $\mathcal{D}_2 \supsetneq \mathcal{D}_1$ of non-truth entailing tests (i.e. “non-standard” ones).
- ▶ If we can't **prove** that our quantifiers range over \mathcal{D}_1 and not over \mathcal{D}_2 , then 4) should be false.

The provenance of UBF

- The original Barcan Formula:

$$(BF) \quad \forall x \Box \varphi(x) \rightarrow \Box \forall x \varphi(x)$$

- If we take $\Box \varphi(x)$ as $\exists y \text{Proof}(y, \ulcorner \varphi(x) \urcorner)$, BF corresponds to

$$(\omega\text{-rule}) \quad \forall x \exists y \text{Proof}(y, \ulcorner \varphi(\dot{x}) \urcorner) \rightarrow (\exists y) \text{Proof}(y, \ulcorner \forall x \varphi(x) \urcorner)$$

- If we take $\varphi(x) = \neg \text{Proof}(x, \ulcorner \perp \urcorner)$, then ω -rule is **invalid** in N .
- Direct QLP analogue of ω -rule

$$(UBF') \quad (\forall x)(\exists y)y : F(x) \rightarrow (\exists y)y : (\forall x)F(x)$$

- UBF' is **not** provable in QLP. UBF has a stronger antecedent:

$$(UBF) \quad (\forall x)t : F(x) \rightarrow \langle t \forall x \rangle : (\forall x)F(x)$$

- But UBF already leads to **arithmetical unsoundness** . . .

Arithmetical semantics for QLP^- (1)

- ▶ Fix an injective primitive term specification \mathcal{P} .
- ▶ An arithmetic interpretation is a function $(\cdot)^* : \mathcal{L}_{QLP^-} \rightarrow \mathcal{L}_a$.
- ▶ Let $\text{Proof}(x, y)$ be a PA proof predicate satisfying the usual GL conditions, $P = \{n \mid N \models \exists y < n(\text{Proof}(\bar{n}, y))\}$,
 $p : \mathbb{N} \rightarrow \mathbb{N}$ a name for a p.r. function which enumerates P .
- ▶ $(\cdot)^*$ is defined inductively on proof terms as follows:
 - ▶ $(x_i)^* = p(x_i)$
 - ▶ if $F(\vec{x}) \in \mathcal{P}(a(\vec{x}))$, then $(a(\vec{x}))^* = a_F(\vec{x}^*) : \vec{\mathbb{N}} \rightarrow \mathbb{N}$
 - ▶ $(t \cdot s)^* = m(t^*, s^*)$, $(t + s)^* = b(t^*, s^*)$, $(!t)^* = c(t^*)$
- ▶ $(\cdot)^*$ is defined inductively on formulas as follows:
 - ▶ $P^* \in \text{Sent}_{\mathcal{L}_a}$ (P atomic), $\perp^* = \perp$
 - ▶ $(F \rightarrow G)^* = F^* \rightarrow G^*$, $(\neg F)^* = \neg F^*$
 - ▶ $((\forall x)F)^* = (\forall x)F^*$, $((\exists x)F)^* = (\exists x)F^*$
 - ▶ $\text{Proof}(t^*, F(x_{k_1}, \dots, x_{k_m})) =$
 $\text{Proof}(t^*, su(x_{k_m}^*, \bar{k}_m, \dots, su(x_2^*, \bar{k}_2, su(x_1^*, \bar{k}_1, \ulcorner F(x_{k_1}^*, \dots, x_{k_m}^* \urcorner))) \dots$

Arithmetical semantics for QLP^- (2)

- ▶ $(\cdot)^*$ is a \mathcal{P} -interpretation if $F \in \mathcal{P}(a) \implies N \models (a : F)^*$.
- ▶ We say that a formula F is \mathcal{P} -valid if for all \mathcal{P} -interpretations $(\cdot)^*$, $N \models (F)^*$.
- ▶ If $FV(F) = \{x_1, \dots, x_n\}$, then $FV(F)^* = \{x_1, \dots, x_n\}$.
- ▶ \mathcal{P} -validity extends to open formulas as follows:

$$N \models (F(x_1, \dots, x_n))^* \text{ iff for all } v, N \models_v (F(x_1, \dots, x_n))^*$$

Proposition

There is a \mathcal{P} -interpretation $(\cdot)^*$ such that $N \models (QLP^-)^*$.

Proof: E.g. $(x : F \rightarrow F)^* = \text{Proof}(p(x), \ulcorner F^* \urcorner) \rightarrow F$. Suppose $N \models_v \text{Proof}(p(x), \ulcorner F^* \urcorner)$. Then $N \models_v \text{Proof}(p(x), \ulcorner F^* \urcorner)[v(x)/x]$. Hence if $m = p(v(x))$, then $PA \vdash \text{Proof}(\bar{m}, \ulcorner F^* \urcorner)$. Thus $N \models F^*$.

Arithmetical soundness and UBF

- ▶ Let QLP° = all theorems of QLP not containing $\langle \cdot \forall \cdot \rangle$.
- ▶ NB: $QLP^- \subsetneq QLP^\circ$.
- ▶ The language of $QLP^\circ = \mathcal{L}_{QLP^-}$. So it makes sense to consider $(QLP^\circ)^* = \{F^* \mid F \in QLP^\circ\}$.
- ▶ As another diagnostic about UBF we have:

Proposition

For any \mathcal{P} -interpretation $(\cdot)^*$, $PA \cup (QLP^\circ)^*$ is inconsistent.

$PA \cup (QLP^\circ)^*$ is inconsistent.

Proof: Taking $F \equiv \perp$ above, note that

- 1) $QLP \vdash (\exists x)x : \perp \rightarrow \perp$
- 2) $QLP \vdash (\exists y)y : ((\exists x)x : \perp \rightarrow \perp)$

Note that since they do not contain $\langle \cdot, \forall \cdot \rangle$, 1), 2) $\in QLP^\circ$.

- 3) $((\exists x)x : \perp \rightarrow \perp)^* = (\exists x)\text{Proof}(p(x), \ulcorner \perp \urcorner) \rightarrow \perp$
- 4) $(\exists y)y : ((\exists x)x : \perp \rightarrow \perp)^* =$
 $(\exists y)[\text{Proof}(p(y), \ulcorner (\exists x)(\text{Proof}(p(x), \ulcorner \perp \urcorner) \rightarrow \perp) \urcorner)]$
- 5) $PA \vdash (\exists y)[\text{Proof}(p(y), \ulcorner (\exists x)(\text{Proof}(p(x), \ulcorner \perp \urcorner) \rightarrow \perp) \urcorner)] \rightarrow$
 $(\exists x)\text{Proof}(p(x), \ulcorner \perp \urcorner)$ (Löb)
- 6) $PA \cup (QLP^\circ)^* \vdash (\exists x)\text{Proof}(p(x), \ulcorner \perp \urcorner)$ 2), 4), 5)
- 7) $PA \cup (QLP^\circ)^* \vdash \perp$ 3), 6)

Morals about the Knower

- ▶ If we take \Box to express knowability, we cannot drop **T**.
- ▶ We still get a paradox even if we drop **I** (or **K**) [Cross].
- ▶ Anderson, Érgé: Don't internalize **T** to yield **U**.
- ▶ **But what's wrong with U?**
- ▶ In QLP **U is not an axiom**.
- ▶ UBF or JUG is required to derive \mathbf{U}_q from \mathbf{T}_q .
- ▶ We've seen four reasons to be suspicious of UBF / JUG:
 - 1) non-converstativeness
 - 2) consistency of $\text{QLP}^- + (\exists y)y : (P \leftrightarrow \neg(\exists x)x : P)$
 - 3) inconsistency of $\text{QLP}^\circ \cup \text{PA}$
 - 4) the possibility that our proof quantifiers range over non-truth entailing justifications
- ▶ Upshot: The problem does lie with internalization, **but only of UG inferences**.

UBF and Gödel's "Lecture at Zilsel's"

- ▶ Gödel [38a] considers how to axiomatize the relation

$$zB\varphi \Leftrightarrow z \text{ is a derivation of } \varphi$$

- ▶ Here B must express the "absolute" proof relation.
- ▶ In particular, Gödel appears to claim
 - 4) $\vdash aB((\forall u)\neg uB(0 = 1))$
 - i.e. a is a derivation of "no u is a derivation of $(0 = 1)$."
- ▶ His other axioms allow him to derive
 - 1) $\vdash \neg uB(0 = 1)$
 - 2) $\vdash bB(\neg uB(0 = 1))$
 - 3) $\vdash (\forall u)bB(\neg uB(0 = 1))$ (?no UG rule is stated?)
- ▶ The step from 3) to 4) appears to require

(UBFg)

$(\forall u)zB\varphi \rightarrow vB((\forall u)\varphi)$ for some derivation v