Epistemic Arithmetic

Eric Pacuit
University of Maryland

Lecture 3, ESSLLI 2025

July 30, 2025

Derivability Conditions

A provability predicate for $\mathbf{T}$, denoted $\mathrm{Prov}_{\mathbf{T}}$, satisfies the following:

*D*1. If $\mathbf{T} \vdash A$, then $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner)$

*D*2. $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \to B \urcorner) \to (\mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to \mathrm{Prov}_{\mathbf{T}}(\ulcorner B \urcorner))$

*D*3. $\mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to \mathrm{Prov}_{\mathbf{T}}(\ulcorner \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \urcorner)$

Löb's Theorem

Theorem (Löb's Theorem)

Let $\mathbf{T}$ be an axiomatizable theory extending $\mathbf{Q}$, and suppose $\mathrm{Prov}_{\mathbf{T}}(y)$ is a formula satisfying conditions $D1$-$D3$.

$$\text{If } \mathbf{T} \vdash \mathrm{Prov}_{\mathbf{T}}(\ulcorner A \urcorner) \to A, \text{ then } \mathbf{T} \vdash A.$$

Plan

- ✓ Introduction: Smullyan's Machine
- ✓ Background
  - ✓ Formal Arithmetic
  - ✓ Gödel's Incompleteness Theorems
  - ✓ Names and Gödel numbering
  - ✓ Fixed Point Theorem
- ✓ Provability predicate and Löb's Theorem
- ▶ Provability logic
- ▶ Predicate approach to modality
- ▶ The Knower Paradox and variants
- ▶ Predicate approach to modality, continued
- ▶ A Primer on Epistemic and Doxastic Logic
- ▶ Anti-Expert Paradox, and related paradoxes
- ▶ Epistemic Arithmetic
- ▶ Gödel's Disjunction

Rineke Verbrugge (2024). *Provability Logic*. The Stanford Encyclopedia of Philosophy (Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), https://plato.stanford.edu/archives/sum2024/entries/logic-provability/.

Propositional Modal Logic

Propositional Modal Language:

$$p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box\varphi$$

where $p \in \mathrm{AT}$ (at set of atomic propositions).

The intended interpretation of $\Box\varphi$ is "there is a proof (in **PA**) of φ".

Propositional Modal Logic

Propositional Modal Language:

$$p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box\varphi$$

where $p \in \mathrm{AT}$ (at set of atomic propositions).

The intended interpretation of $\Box\varphi$ is “there is a proof (in **PA**) of φ”.

A **frame** is a tuple (W, R) such that $W \neq \emptyset$ and $R \subseteq W \times W$.

Propositional Modal Logic

Propositional Modal Language:

$$p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box\varphi$$

where $p \in \mathsf{AT}$ (at set of atomic propositions).

The intended interpretation of $\Box\varphi$ is “there is a proof (in **PA**) of φ”.

A **frame** is a tuple (W, R) such that $W \neq \emptyset$ and $R \subseteq W \times W$.

A **model** is a tuple (W, R, V) where (W, R) is a frame and $V : \mathsf{AT} \to \wp(W)$.

Truth/Validity

For a model $\mathcal{M} = (W, R, V)$ and $w \in W$, truth is defined as usual:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff $\mathcal{M}, w \not\models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$
- $\mathcal{M}, w \models \Box\varphi$ iff for all $v \in W$, if $w\,R\,v$, then $\mathcal{M}, v \models \varphi$

For a frame $\mathcal{F} = (W, R)$, φ is **valid on** $\mathcal{F}$, denoted $\mathcal{F} \models \varphi$, when $\mathcal{M}, w \models \varphi$ for all models $\mathcal{M}$ based on $\mathcal{F}$ and $w \in W$.

Provability Logic: **GL**

K	$\Box(\varphi \to \psi) \to (\Box\varphi \to \psi)$
L	$\Box(\Box\varphi \to \varphi) \to \Box\varphi$
MP	$\varphi, \varphi \to \psi \ \therefore \ \psi$
NEC	$\varphi \ \therefore \ \Box\varphi$

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.
- $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ is valid on a frame (W, R) if, and only if, R is transitive and converse well-founded (there are no infinite ascending sequences, that is sequences of the form $w_1 \; R \; w_2 \; R \; w_3 \cdots$).

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.
- $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ is valid on a frame (W, R) if, and only if, R is transitive and converse well-founded (there are no infinite ascending sequences, that is sequences of the form $w_1 \; R \; w_2 \; R \; w_3 \cdots$).
- The logic **GL** is not compact:

$$\Gamma = \{\Diamond p_0, \Box(p_0 \to \Diamond p_1), \Box(p_1 \to \Diamond p_2), \ldots, \Box(p_n \to \Diamond p_{n+1}), \ldots\}.$$

  is finitely satisfiable, but not satisfiable.

Some Results

- $\mathbf{GL} \vdash \Box\varphi \to \Box\Box\varphi$.
- $\Box(\Box\varphi \to \varphi) \to \Box\varphi$ is valid on a frame (W, R) if, and only if, R is transitive and converse well-founded (there are no infinite ascending sequences, that is sequences of the form $w_1 \; R \; w_2 \; R \; w_3 \cdots$).
- The logic **GL** is not compact:

$$\Gamma = \{\Diamond p_0, \Box(p_0 \to \Diamond p_1), \Box(p_1 \to \Diamond p_2), \ldots, \Box(p_n \to \Diamond p_{n+1}), \ldots\}.$$

  is finitely satisfiable, but not satisfiable.
- The logic **GL** is sound and weakly complete with respect to the class of frames that are transitive and converse well-founded.

Arithmetic Completeness

An **arithmetic translation** is a function t such that

1. For all $p \in \mathrm{At}$, $t(p)$ is a sentence of $\mathcal{L}_A$
2. t commutes with the boolean connectives: $t(\neg\varphi) = \neg t(\varphi)$, $t(\varphi \wedge \psi) = t(\varphi) \wedge t(\psi)$, etc.
3. $t(\Box\varphi) = \mathrm{Prov}_{\mathbf{PA}}(\ulcorner t(\varphi) \urcorner)$

Theorem (Solovay 1976).

$$\mathbf{GL} \vdash \varphi \text{ iff for every arithmetic translation } t,\ \mathbf{PA} \vdash t(\varphi).$$

Plan

- ✓ Introduction: Smullyan's Machine
- ✓ Background
  - ✓ Formal Arithmetic
  - ✓ Gödel's Incompleteness Theorems
  - ✓ Names and Gödel numbering
  - ✓ Fixed Point Theorem
- ✓ Provability predicate and Löb's Theorem
- ✓ Provability logic
- Predicate approach to modality
- The Knower Paradox and variants
- Predicate approach to modality, continued
- A Primer on Epistemic and Doxastic Logic
- Anti-Expert Paradox, and related paradoxes
- Epistemic Arithmetic
- Gödel's Disjunction

Predicate vs. Operator Approach to Modality

Predicate Approach	'$2+2=4$' is necessary
Operator Approach	It is necessary that $2+2=4$.

Predicate vs. Operator Approach to Modality

Predicate Approach	'$2+2=4$' is necessary
Operator Approach	It is necessary that $2+2=4$.

- We have treated 'provability' both as a predicate ($\mathrm{Prov}_{\mathbf{T}}(\cdot)$) and as a sentential operator (in **GL**)
- Truth is typically only treated as a predicate

Predicate vs. Operator Approach to Modality

Whether necessity, knowledge, belief, future and past truth, obligation, and other modalitities should be formalised by operators or by predicates was a matter of dispute up to the early sixties between two almost equally strong parties. Then two technical achievements helped the operator approach to an almost complete triumph over the predicate approach that had been advocated by illustrious philosophers like Quine. (p. 180)

Volker Halbach, Hannes Leitgeb and Philip Welch (2003). *Possible-Worlds Semantics for Modal Notions Conceived as Predicates*. Journal of Philosophical Logic, 32:2, pp. 179-223.

Operator > Predicate

1. Montague provided the first result by proving that the predicate version of the modal system **T** is inconsistent if it is combined with weak systems of arithmetic. From his result he concluded that “virtually all of modal logic...must be sacrificed”, if necessity is conceived of as a predicate of sentences.

Operator > Predicate

1. Montague provided the first result by proving that the predicate version of the modal system **T** is inconsistent if it is combined with weak systems of arithmetic. From his result he concluded that "virtually all of modal logic...must be sacrificed", if necessity is conceived of as a predicate of sentences.

2. The other technical achievement that brought about the triumph of the operator view was the emergence of possible-worlds semantic. Hintikka, Kanger and Kripke provided semantics for modal operator logics, while nothing similar seemed available for the predicate approach.

Theorem (Tarski/Gödel). Let $\mathbf{T}$ be a theory extending $\mathbf{Q}$ and T a unary predicate such that for all sentences φ:

$$\mathbf{T} \vdash T(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$$

Then, $\mathbf{T}$ is inconsistent.

Proof. By the Fixed Point Theorem, there is a sentence D such that

$$\mathbf{T} \vdash D \leftrightarrow \neg T(\ulcorner D \urcorner)$$

Theorem (Tarski/Gödel). Let **T** be a theory extending **Q** and T a unary predicate such that for all sentences φ:

$$\mathbf{T} \vdash T(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$$

Then, **T** is inconsistent.

Proof. By the Fixed Point Theorem, there is a sentence D such that

$$\mathbf{T} \vdash D \leftrightarrow \neg T(\ulcorner D \urcorner)$$

But, since $\mathbf{T} \vdash T(\ulcorner D \urcorner) \leftrightarrow D$, the contradiction is immediate.

Montague's Theorem

Theorem (Montague, 1963)

Suppose $\mathbf{T}$ is a theory and $\Box(x)$ is a formula such that for all sentences φ,

(T) $\mathbf{T} \vdash \Box(\ulcorner\varphi\urcorner) \rightarrow \varphi$

(Nec) If $\mathbf{T} \vdash \varphi$, then $\mathbf{T} \vdash \Box(\ulcorner\varphi\urcorner)$

(Q) $\mathbf{Q} \subseteq \mathbf{T}$

Then $\mathbf{T}$ is inconsistent.

R. Montague (1963). *Syntactical Treatment of Modality, with Corollaries on Reflexion Principles and Finite Axiomatizability*. Acta Philosophica Finnica, 16, pp. 153 - 167.

1. $D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$ FPT (using Q)

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth
3.	$\Box(\ulcorner D \urcorner) \rightarrow \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \to D$	Truth
3.	$\Box(\ulcorner D \urcorner) \to \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2
4.	$\neg\Box(\ulcorner D \urcorner)$	PC: 3

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth
3.	$\Box(\ulcorner D \urcorner) \rightarrow \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2
4.	$\neg\Box(\ulcorner D \urcorner)$	PC: 3
5.	D	PC: 1, 4

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth
3.	$\Box(\ulcorner D \urcorner) \rightarrow \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2
4.	$\neg\Box(\ulcorner D \urcorner)$	PC: 3
5.	D	PC: 1, 4
6.	$\Box(\ulcorner D \urcorner)$	Nec: 5

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth
3.	$\Box(\ulcorner D \urcorner) \rightarrow \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2
4.	$\neg\Box(\ulcorner D \urcorner)$	PC: 3
5.	D	PC: 1, 4
6.	$\Box(\ulcorner D \urcorner)$	Nec: 5
7.	$\bot$	3, 6

1.	$D \leftrightarrow \neg\Box(\ulcorner D \urcorner)$	FPT (using Q)
2.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth
3.	$\Box(\ulcorner D \urcorner) \rightarrow \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2
4.	$\neg\Box(\ulcorner D \urcorner)$	PC: 3
5.	D	PC: 1, 4
6.	$\Box(\ulcorner D \urcorner)$	Nec: 5
7.	$\bot$	3, 6

T. Tymoczko (1984). *An unsolved puzzle about knowledge*. Philosophical Quarterly 34, pp. 437-458.

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$
   - therefore, 'nobody knows this statement to be true' is true, that is nobody knows this statement to be true.

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$
   - therefore, 'nobody knows this statement to be true' is true, that is nobody knows this statement to be true.
3. I have shown that if someone knows this statement to be true, nobody knows this statement to be true. $(K(D) \rightarrow \neg K(D))$

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$
   - therefore, 'nobody knows this statement to be true' is true, that is nobody knows this statement to be true.
3. I have shown that if someone knows this statement to be true, nobody knows this statement to be true. $(K(D) \rightarrow \neg K(D))$
4. So nobody knows this statement to be true. $(\neg K(D))$

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$
   - therefore, 'nobody knows this statement to be true' is true, that is nobody knows this statement to be true.
3. I have shown that if someone knows this statement to be true, nobody knows this statement to be true. $(K(D) \rightarrow \neg K(D))$
4. So nobody knows this statement to be true. $(\neg K(D))$
5. This is what the statement says, hence it is true. (D)

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$
   - therefore, 'nobody knows this statement to be true' is true, that is nobody knows this statement to be true.
3. I have shown that if someone knows this statement to be true, nobody knows this statement to be true. $(K(D) \rightarrow \neg K(D))$
4. So nobody knows this statement to be true. $(\neg K(D))$
5. This is what the statement says, hence it is true. (D)
6. But hold on! I have just proved this statement to be true. Hence someone (at least me) knows this statement to be true! $(K(D))$

1. Nobody knows this statement to be true. $(D \leftrightarrow \neg K(D))$
2. Suppose someone knows this statement to be true;
   - then this statement is true, otherwise it couldn't be known; $(K(D) \rightarrow D)$
   - therefore, 'nobody knows this statement to be true' is true, that is nobody knows this statement to be true.
3. I have shown that if someone knows this statement to be true, nobody knows this statement to be true. $(K(D) \rightarrow \neg K(D))$
4. So nobody knows this statement to be true. $(\neg K(D))$
5. This is what the statement says, hence it is true. (D)
6. But hold on! I have just proved this statement to be true. Hence someone (at least me) knows this statement to be true! $(K(D))$
7. Now this contradicts what has just been established. $(\bot)$

The Knower Paradox

Theorem (Montague-Kaplan 1960)

Let **T** be an axiomatizable extension of **Q**, with $I(x, y)$ a formula of expressing derivability between sentences in **T**, and K a (perhaps complex) unary predicate satisfying, for all sentences φ and ψ:

(T) $K(\varphi) \rightarrow \varphi$

(U) $K(K\varphi \rightarrow \varphi)$

(I) $(K(\varphi) \wedge I(\varphi, \psi)) \rightarrow K(\psi)$

then T is inconsistent.

1. $D \leftrightarrow K(\neg D)$ FPT (using Q)

1. $D \leftrightarrow K(\neg D)$ FPT (using Q)

2. $K(\neg D) \to \neg D$ Truth

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \rightarrow \neg D)$	U

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \rightarrow \neg D)$	U
7.	$(K(K(\neg D) \rightarrow \neg D) \wedge I(K(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K(\neg D)$	I

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \rightarrow \neg D)$	U
7.	$(K(K(\neg D) \rightarrow \neg D) \wedge I(K(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K(\neg D)$	I
8.	$K(\neg D)$	PC, MP: 5 & 6, 7

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \rightarrow \neg D)$	U
7.	$(K(K(\neg D) \rightarrow \neg D) \wedge I(K(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K(\neg D)$	I
8.	$K(\neg D)$	PC, MP: 5 & 6, 7
9.	D	PC: 1, 8

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \rightarrow \neg D)$	U
7.	$(K(K(\neg D) \rightarrow \neg D) \wedge I(K(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K(\neg D)$	I
8.	$K(\neg D)$	PC, MP: 5 & 6, 7
9.	D	PC: 1, 8
10.	$\bot$	4, 9

1.	$D \leftrightarrow K(\neg D)$	FPT (using Q)
2.	$K(\neg D) \to \neg D$	Truth
3.	$D \to \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \to \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \to \neg D)$	U
7.	$(K(K(\neg D) \to \neg D) \wedge I(K(\neg D) \to \neg D, \neg D)) \to K(\neg D)$	I
8.	$K(\neg D)$	PC, MP: 5 & 6, 7
9.	D	PC: 1, 8
10.	$\bot$	4, 9

Surprise Exam

A schoolmaster announces to his pupils:

> Unless you know this statement to be false, you will have an exam tomorrow, but you can't know from this statement that you will have an exam tomorrow.

Surprise Exam

A schoolmaster announces to his pupils:

> Unless you know this statement to be false, you will have an exam tomorrow, but you can't know from this statement that you will have an exam tomorrow.

$$\mathbf{T} \vdash D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$$

Theorem

Let **T** be an axiomatizable extension of **Q**, with $I(x, y)$ a formula expressing derivability between sentences in **T**, and K a (perhaps complex) unary predicate, such that **T** satisfies the axiom schemata:

(T) $K(\varphi) \rightarrow \varphi$

(U) $K(K(\varphi) \rightarrow \varphi)$

(I) $(K(\varphi) \wedge I(\varphi, F)) \rightarrow K(F)$

(R) $K(T' \wedge U' \wedge I')$ (where T', U', and I' any instance of T, U, I)

Then **T** is inconsistent.

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2
4.	$D \rightarrow (F \wedge \neg K(D \rightarrow F))$	PC: 1, 3

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \to F)))$	FPT (using Q)
2.	$K(\neg D) \to \neg D$	T, call it T'
3.	$D \to \neg K(\neg D)$	PC: 2
4.	$D \to (F \wedge \neg K(D \to F))$	PC: 1, 3
5.	$D \to F$	PC: 4

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2
4.	$D \rightarrow (F \wedge \neg K(D \rightarrow F))$	PC: 1, 3
5.	$D \rightarrow F$	PC: 4
6.	$I(T', D \rightarrow F)$	2-5

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2
4.	$D \rightarrow (F \wedge \neg K(D \rightarrow F))$	PC: 1, 3
5.	$D \rightarrow F$	PC: 4
6.	$I(T', D \rightarrow F)$	2-5
7.	$K(T')$	U, call it U'

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2
4.	$D \rightarrow (F \wedge \neg K(D \rightarrow F))$	PC: 1, 3
5.	$D \rightarrow F$	PC: 4
6.	$I(T', D \rightarrow F)$	2-5
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \to F)))$	FPT (using Q)
2.	$K(\neg D) \to \neg D$	T, call it T'
3.	$D \to \neg K(\neg D)$	PC: 2
4.	$D \to (F \wedge \neg K(D \to F))$	PC: 1, 3
5.	$D \to F$	PC: 4
6.	$I(T', D \to F)$	2-5
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \to F) \to K(D \to F)$	I, call it I'
9.	$K(D \to F)$	PC: 6, 7, 8

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2
4.	$D \rightarrow (F \wedge \neg K(D \rightarrow F))$	PC: 1, 3
5.	$D \rightarrow F$	PC: 4
6.	$I(T', D \rightarrow F)$	2-5
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'
9.	$K(D \rightarrow F)$	PC: 6, 7, 8
10.	$D \rightarrow \neg K(D \rightarrow F)$	PC: 4

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
3.	$D \rightarrow \neg K(\neg D)$	PC: 2
4.	$D \rightarrow (F \wedge \neg K(D \rightarrow F))$	PC: 1, 3
5.	$D \rightarrow F$	PC: 4
6.	$I(T', D \rightarrow F)$	2-5
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'
9.	$K(D \rightarrow F)$	PC: 6, 7, 8
10.	$D \rightarrow \neg K(D \rightarrow F)$	PC: 4

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'
11.	$\neg D$	PC: 9, 10

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'
11.	$\neg D$	PC: 9, 10
12.	$I(T' \wedge U' \wedge I', \neg D)$	2-11

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'
11.	$\neg D$	PC: 9, 10
12.	$I(T' \wedge U' \wedge I', \neg D)$	2-11
13.	$K(T' \wedge U' \wedge I')$	R

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	FPT (using Q)
2.	$K(\neg D) \rightarrow \neg D$	T, call it T'
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \rightarrow F) \rightarrow K(D \rightarrow F)$	I, call it I'
11.	$\neg D$	PC: 9, 10
12.	$I(T' \wedge U' \wedge I', \neg D)$	2-11
13.	$K(T' \wedge U' \wedge I')$	R
14.	$K(\neg D)$	12, 13, I

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \to F)))$	FPT (using Q)
2.	$K(\neg D) \to \neg D$	T, call it T'
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \to F) \to K(D \to F)$	I, call it I'
11.	$\neg D$	PC: 9, 10
12.	$I(T' \wedge U' \wedge I', \neg D)$	2-11
13.	$K(T' \wedge U' \wedge I')$	R
14.	$K(\neg D)$	12, 13, I
15.	D	PC: 1, 14

1.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \to F)))$	FPT (using Q)
2.	$K(\neg D) \to \neg D$	T, call it T'
7.	$K(T')$	U, call it U'
8.	$K(T') \wedge I(T', D \to F) \to K(D \to F)$	I, call it I'
11.	$\neg D$	PC: 9, 10
12.	$I(T' \wedge U' \wedge I', \neg D)$	2-11
13.	$K(T' \wedge U' \wedge I')$	R
14.	$K(\neg D)$	12, 13, I
15.	D	PC: 1, 14
16.	$\bot$	11, 15

Theorem (Montague 1963)

Let **T** be a theory extending **Q**, with K a (perhaps complex) unary predicate, satisfying, for all sentences φ and ψ:

(T)	$K(\varphi) \to \varphi$
(U)	$K(K(\varphi) \to \varphi)$
(I)	$(K(\varphi) \wedge I(\varphi, \psi)) \to K(\psi)$
(Log)	$K(\alpha)$, if α is a logical axiom of first-order logic with identity
(Strong)	If $\mathbf{T} \vdash K(\varphi \to \psi)$ and $\mathbf{T} \vdash K(\varphi)$, then $\mathbf{T} \vdash K(\psi)$

then T is inconsistent.

How should we solve this paradox? Should knowledge entail truth? Should we accept the epistemic closure principle or not? Should the syntax be changed in such a way that statements that lead to paradoxes are eliminated?

Theorem (Koons, Turner)

Let $\mathbf{T}$ be a theory extending $\mathbf{Q}$, with B a (perhaps complex) unary predicate, such that $\mathbf{T}$ satisfies, for all sentences φ and ψ:

(4)	$B(\varphi) \to B(B(\varphi))$
(D)	$B(\neg\varphi) \to \neg B(\varphi)$
(Nec)	If $\mathbf{T} \vdash \varphi$, then $\mathbf{T} \vdash B(\varphi)$
(Re)	If $\mathbf{T} \vdash \varphi \leftrightarrow \psi$, then $\mathbf{T} \vdash B(\varphi) \leftrightarrow B(\psi)$

then T is inconsistent.

1. $F \leftrightarrow \neg B(F)$ FPT

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \rightarrow \neg B(B(F))$	D

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \rightarrow \neg B(B(F))$	D
4.	$B(F) \rightarrow \neg B(B(F))$	PC: 2, 3

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \to \neg B(B(F))$	D
4.	$B(F) \to \neg B(B(F))$	PC: 2, 3
5.	$B(F) \to B(B(F))$	4

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \rightarrow \neg B(B(F))$	D
4.	$B(F) \rightarrow \neg B(B(F))$	PC: 2, 3
5.	$B(F) \rightarrow B(B(F))$	4
6.	$\neg B(F)$	PC: 4, 5

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \rightarrow \neg B(B(F))$	D
4.	$B(F) \rightarrow \neg B(B(F))$	PC: 2, 3
5.	$B(F) \rightarrow B(B(F))$	4
6.	$\neg B(F)$	PC: 4, 5
7.	F	PC: 1, 6

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \rightarrow \neg B(B(F))$	D
4.	$B(F) \rightarrow \neg B(B(F))$	PC: 2, 3
5.	$B(F) \rightarrow B(B(F))$	4
6.	$\neg B(F)$	PC: 4, 5
7.	F	PC: 1, 6
8.	$B(F)$	Nec, 7

1.	$F \leftrightarrow \neg B(F)$	FPT
2.	$B(F) \leftrightarrow B(\neg B(F))$	Re, 1
3.	$B(\neg B(F)) \rightarrow \neg B(B(F))$	D
4.	$B(F) \rightarrow \neg B(B(F))$	PC: 2, 3
5.	$B(F) \rightarrow B(B(F))$	4
6.	$\neg B(F)$	PC: 4, 5
7.	F	PC: 1, 6
8.	$B(F)$	Nec, 7
9.	$\bot$	6, 8

Theorem (Cross 2001)

Let **T** be an axiomatizable theory extending **Q**, with K a (perhaps complex) predicate. Let $K'(x)$ be the predicate defined by the formula:

$$\exists y(K(y) \wedge I(y,x))$$

where $I(y,x)$ is a predicate expressing derivability between sentences in **T**. Suppose **T** satisfies the following axiom schemata:

(T′) $K'(\varphi) \rightarrow \varphi$

(U′) $K'(K'(\varphi) \rightarrow \varphi)$

then **T** is inconsistent.

We show that $\mathbf{T} \vdash (K'(\varphi) \wedge I(\varphi, \psi)) \rightarrow K'(\psi)$, for every sentences φ and ψ.

We show that $\mathbf{T} \vdash (K'(\varphi) \wedge I(\varphi, \psi)) \rightarrow K'(\psi)$, for every sentences φ and ψ.

1. $K'(\varphi) \leftrightarrow \exists y(K(y) \wedge I(y, \varphi))$ Definition of K'

We show that $\mathbf{T} \vdash (K'(\varphi) \land I(\varphi, \psi)) \to K'(\psi)$, for every sentences φ and ψ.

1.	$K'(\varphi) \leftrightarrow \exists y(K(y) \land I(y, \varphi))$	Definition of K'
2.	$K'(\varphi) \land I(\varphi, \psi) \leftrightarrow \exists y(K(y) \land I(y, \varphi) \land I(\varphi, \psi))$	PC: 1

We show that $\mathbf{T} \vdash (K'(\varphi) \land I(\varphi, \psi)) \to K'(\psi)$, for every sentences φ and ψ.

1.	$K'(\varphi) \leftrightarrow \exists y(K(y) \land I(y, \varphi))$	Definition of K'
2.	$K'(\varphi) \land I(\varphi, \psi) \leftrightarrow \exists y(K(y) \land I(y, \varphi) \land I(\varphi, \psi))$	PC: 1
3.	$K'(\varphi) \land I(\varphi, \psi) \leftrightarrow \exists y(K(y) \land I(y, \psi))$	Transitivity of I

We show that $\mathbf{T} \vdash (K'(\varphi) \wedge I(\varphi, \psi)) \rightarrow K'(\psi)$, for every sentences φ and ψ.

1.	$K'(\varphi) \leftrightarrow \exists y(K(y) \wedge I(y, \varphi))$	Definition of K'
2.	$K'(\varphi) \wedge I(\varphi, \psi) \leftrightarrow \exists y(K(y) \wedge I(y, \varphi) \wedge I(\varphi, \psi))$	PC: 1
3.	$K'(\varphi) \wedge I(\varphi, \psi) \leftrightarrow \exists y(K(y) \wedge I(y, \psi))$	Transitivity of I
4.	$K'(\varphi) \wedge I(\varphi, \psi) \leftrightarrow K'(\psi)$	Definition of K'

We show that $\mathbf{T} \vdash (K'(\varphi) \wedge I(\varphi, \psi)) \rightarrow K'(\psi)$, for every sentences φ and ψ.

1.	$K'(\varphi) \leftrightarrow \exists y(K(y) \wedge I(y, \varphi))$	Definition of K'
2.	$K'(\varphi) \wedge I(\varphi, \psi) \leftrightarrow \exists y(K(y) \wedge I(y, \varphi) \wedge I(\varphi, \psi))$	PC: 1
3.	$K'(\varphi) \wedge I(\varphi, \psi) \leftrightarrow \exists y(K(y) \wedge I(y, \psi))$	Transitivity of I
4.	$K'(\varphi) \wedge I(\varphi, \psi) \leftrightarrow K'(\psi)$	Definition of K'

Call this property I′: It depends only on the definition K' and I. Hence, by Montague-Kaplan's theorem, $\mathbf{T}$ is inconsistent.

Theorem (Cross's 'Knowledge-Plus Knower')

Let **T** be an axiomatizable theory extending **Q**, with K and K' defined as previously, and such that **T** satisfies, for every sentence φ:

$$\text{(T}'\text{)} \quad K'(\varphi) \to \varphi$$

$$\text{(U}^+\text{)} \quad K(K'(\varphi) \to \varphi)$$

then **T** is inconsistent.

Let $T'_{\neg D}$, $U'_{\neg D}$, and $U^{+}_{\neg D}$ denote instances of T′, U′ and U^{+} using $\neg D$, where:

$$\mathbf{T} \vdash D \leftrightarrow K'(\neg D)$$

Let $T'_{\neg D}$, $U'_{\neg D}$, and $U^{+}_{\neg D}$ denote instances of T′, U′ and U⁺ using $\neg D$, where:

$$\mathbf{T} \vdash D \leftrightarrow K'(\neg D)$$

By the definition of K', the following is provable in **T**:

$$(K(K'(\neg D) \to \neg D) \wedge I(K'(\neg D) \to \neg D, K'(\neg D) \to \neg D)) \to K'(K'(\neg D) \to \neg D)$$

Let $T'_{\neg D}$, $U'_{\neg D}$, and $U^{+}_{\neg D}$ denote instances of T′, U′ and U^{+} using $\neg D$, where:

$$\mathbf{T} \vdash D \leftrightarrow K'(\neg D)$$

By the definition of K', the following is provable in **T**:

$$(K(\, K'(\neg D) \to \neg D \,) \wedge I(\, K'(\neg D) \to \neg D \,,\, K'(\neg D) \to \neg D \,)) \to K'(\, K'(\neg D) \to \neg D \,)$$

This is equivalent to:

$$(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \to U'_{\neg D}$$

1. $(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \rightarrow U'_{\neg D}$ From above formula

1. $(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \rightarrow U'_{\neg D}$ From above formula

2. $I(T'_{\neg D}, T'_{\neg D})$ Every sentence is derivable from itself

1.	$(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \rightarrow U'_{\neg D}$	From above formula
2.	$I(T'_{\neg D}, T'_{\neg D})$	Every sentence is derivable from itself
3.	$U^{+}_{\neg D} \rightarrow U'_{\neg D}$	PC: 1, 2

1. $(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \rightarrow U'_{\neg D}$ From above formula
2. $I(T'_{\neg D}, T'_{\neg D})$ Every sentence is derivable from itself
3. $U^{+}_{\neg D} \rightarrow U'_{\neg D}$ PC: 1, 2
4. $(T'_{\neg D} \wedge U'_{\neg D}) \rightarrow \bot$ Cross Theorem

1.	$(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \rightarrow U'_{\neg D}$	From above formula
2.	$I(T'_{\neg D}, T'_{\neg D})$	Every sentence is derivable from itself
3.	$U^{+}_{\neg D} \rightarrow U'_{\neg D}$	PC: 1, 2
4.	$(T'_{\neg D} \wedge U'_{\neg D}) \rightarrow \bot$	Cross Theorem
5.	$(T'_{\neg D} \wedge U^{+}_{\neg D}) \rightarrow \bot$	PC: 3, 4

1.	$(U^{+}_{\neg D} \wedge I(T'_{\neg D}, T'_{\neg D})) \rightarrow U'_{\neg D}$	From above formula
2.	$I(T'_{\neg D}, T'_{\neg D})$	Every sentence is derivable from itself
3.	$U^{+}_{\neg D} \rightarrow U'_{\neg D}$	PC: 1, 2
4.	$(T'_{\neg D} \wedge U'_{\neg D}) \rightarrow \bot$	Cross Theorem
5.	$(T'_{\neg D} \wedge U^{+}_{\neg D}) \rightarrow \bot$	PC: 3, 4

Anderson's Solution

C. Anthony Anderson (1983). *The Paradox of the Knower*. The Journal of Philosophy, 80, 6, pp. 338-355.

Anderson's Solution

$\mathcal{L}_0$: the smallest extension of $\mathcal{L}_A$ such that
if $\varphi, \psi \in \mathcal{L}_A$, then $K_0(\varphi), I_0(\varphi, \psi) \in \mathcal{L}_0$,
closed under Boolean operators.

Anderson's Solution

$\mathcal{L}_0$: the smallest extension of $\mathcal{L}_A$ such that if $\varphi, \psi \in \mathcal{L}_A$, then $K_0(\varphi), I_0(\varphi, \psi) \in \mathcal{L}_0$, closed under Boolean operators.

$\mathcal{L}_{i+1}$: the smallest extension of $\mathcal{L}_i$ such that if $\varphi, \psi \in \mathcal{L}_i$, then $K_{i+1}(\varphi), I_{i+1}(\varphi, \psi) \in \mathcal{L}_{i+1}$, closed under Boolean operators.

Anderson's Solution

$\mathcal{L}_0$: the smallest extension of $\mathcal{L}_A$ such that if $\varphi, \psi \in \mathcal{L}_A$, then $K_0(\varphi), I_0(\varphi, \psi) \in \mathcal{L}_0$, closed under Boolean operators.

$\mathcal{L}_{i+1}$: the smallest extension of $\mathcal{L}_i$ such that if $\varphi, \psi \in \mathcal{L}_i$, then $K_{i+1}(\varphi), I_{i+1}(\varphi, \psi) \in \mathcal{L}_{i+1}$, closed under Boolean operators.

$\mathcal{L}_\omega$: $\bigcup_{i \in \omega} \mathcal{L}_i$

Anderson's Solution

$\mathcal{L}_0$: the smallest extension of $\mathcal{L}_A$ such that
if $\varphi, \psi \in \mathcal{L}_A$, then $K_0(\varphi), I_0(\varphi, \psi) \in \mathcal{L}_0$,
closed under Boolean operators.

$\mathcal{L}_{i+1}$: the smallest extension of $\mathcal{L}_i$ such that
if $\varphi, \psi \in \mathcal{L}_i$, then $K_{i+1}(\varphi), I_{i+1}(\varphi, \psi) \in \mathcal{L}_{i+1}$,
closed under Boolean operators.

$\mathcal{L}_\omega$: $\bigcup_{i \in \omega} \mathcal{L}_i$

K_i indicates a certain level of knowledge. Anderson gives an "intuitive motivation": Some sentence that cannot be in a set of statements known at level i can still be provable. By understanding the proof of such a statement, one knows this sentence at level $i + 1$.

Anderson's Solution

$gn(\mathcal{L}_\omega) = \{gn(\alpha) \mid \alpha \in \mathcal{L}_\omega\}$ is the set of Gödel numbers of each formula in $\mathcal{L}_\omega$.
Suppose that V_p is an interpretation of $\mathcal{L}_A$:

- V_0 extends V_p to $\mathcal{L}_0$
- V_{i+1} extends V_i to $\mathcal{L}_{i+1}$
- $V_i(K_i) \subseteq gn(\mathcal{L}_\omega)$
- $V_i(I_i) \subseteq gn(\mathcal{L}_\omega) \times gn(\mathcal{L}_\omega)$
- $V = \bigcup_{i \in \omega} V_i$

Anderson's Solution

$$
\begin{aligned}
\mathbf{T}_0 &= \mathbf{Q} \cup \{K_0(\ulcorner \varphi \urcorner) \to \varphi \mid \varphi \in \mathcal{L}_\omega\} \\
\mathbf{T}_{i+1} &= \mathbf{T}_i \cup \{K_{i+1}(\ulcorner \varphi \urcorner) \to \varphi \mid \varphi \in \mathcal{L}_\omega\} \\
V_0(K_0(\ulcorner \varphi \urcorner)) &= 1 \text{ if and only if } \mathbf{Q} \vdash \varphi \\
V_{i+1}(K_{i+1}(\ulcorner \varphi \urcorner)) &= 1 \text{ if and only if } \mathbf{T}_i \vdash \varphi \\
V_0(I_0(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner)) &= 1 \text{ if and only if } \mathbf{Q} \vdash \varphi \to \psi \\
V_{i+1}(I_{i+1}(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner)) &= 1 \text{ if and only if } \mathbf{T}_i \vdash \varphi \to \psi \\
\mathbf{T}_\omega &= \bigcup_{i \in \omega} \mathbf{T}_i .
\end{aligned}
$$

Anderson's Solution

- $V_i(K_i) \subseteq V_{i+1}(K_{i+1})$.
- $V_i(I_i) \subseteq V_{i+1}(I_{i+1})$.
- If $n = gn(\varphi) \in V_i(K_i)$, then $\exists j \geq i$ such that $V_j(\varphi) = 1$.
- If $n = gn(\varphi), m = gn(\psi), (n, m) \in V_i(I_i)$, then $\exists j \geq i$ such that $V_j(\varphi \to \psi) = 1$.
- If $(n, m) \in V_i(I_i), n \in V_i(K_i)$, then $m \in V_i(K_i)$.

Anderson's Solution

$$\begin{aligned} V(K_i(\ulcorner\varphi\urcorner) \rightarrow \varphi) &= 1 \\ V([I_i(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \wedge K_i(\ulcorner\varphi\urcorner)] \rightarrow K_i(\ulcorner\psi\urcorner)) &= 1 \\ V(K_{i+1}(\ulcorner K_i(\ulcorner\varphi\urcorner) \rightarrow \varphi\urcorner)) &= 1 \end{aligned}$$

Anderson's Solution

$$\begin{aligned} V(K_i(\ulcorner\varphi\urcorner) \to \varphi) &= 1 \\ V([I_i(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \wedge K_i(\ulcorner\varphi\urcorner)] \to K_i(\ulcorner\psi\urcorner)) &= 1 \\ V(K_{i+1}(\ulcorner K_i(\ulcorner\varphi\urcorner) \to \varphi\urcorner)) &= 1 \end{aligned}$$

$$K_{i+1}(\ulcorner K_i(\ulcorner\varphi\urcorner) \to \varphi\urcorner) \quad \textit{vs.} \quad K_i(\ulcorner K_i(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$$

Anderson's Solution

$$
\begin{aligned}
V(K_i(\ulcorner\varphi\urcorner) \to \varphi) &= 1 \\
V([I_i(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \wedge K_i(\ulcorner\varphi\urcorner)] \to K_i(\ulcorner\psi\urcorner)) &= 1 \\
V(K_{i+1}(\ulcorner K_i(\ulcorner\varphi\urcorner) \to \varphi\urcorner)) &= 1
\end{aligned}
$$

$$K_{i+1}(\ulcorner K_i(\ulcorner\varphi\urcorner) \to \varphi\urcorner) \quad vs. \quad K_i(\ulcorner K_i(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$$

$K_i(\ulcorner\varphi\urcorner) \to K_j(\ulcorner\varphi\urcorner)$ for $j \geq i$.

$I_i(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner) \to I_j(\ulcorner\varphi\urcorner, \ulcorner\psi\urcorner)$ for $j \geq i$.

Blocking the Knower Paradox

1.	$D \leftrightarrow K(\neg D)$	FPT
2.	$K(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I(K(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K(K(\neg D) \rightarrow \neg D)$	U
7.	$(K(K(\neg D) \rightarrow \neg D) \wedge I(K(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K(\neg D)$	I
8.	$K(\neg D)$	PC: 5, 6, 7
9.	D	PC: 1, 8
10.	$\bot$	4, 9

Blocking the Knower Paradox

1.	$D \leftrightarrow K_i(\neg D)$	FPT
2.	$K_i(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5.	$I_i(K_i(\neg D) \rightarrow \neg D, \neg D)$	2-4

Blocking the Knower Paradox

1.	$D \leftrightarrow K_i(\neg D)$	FPT
2.	$K_i(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5′.	$I_{i+1}(K_i(\neg D) \rightarrow \neg D, \neg D)$	2-4

Blocking the Knower Paradox

1.	$D \leftrightarrow K_i(\neg D)$	FPT
2.	$K_i(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5′.	$I_{i+1}(K_i(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K_{i+1}(K_i(\neg D) \rightarrow \neg D)$	
7.	$(K_{i+1}(K_i(\neg D) \rightarrow \neg D) \wedge I_{i+1}(K_i(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K_{i+1}(\neg D)$	I
8.	$K_{i+1}(\neg D)$	PC: 5, 6, 7

Blocking the Knower Paradox

1.	$D \leftrightarrow K_i(\neg D)$	FPT
2.	$K_i(\neg D) \rightarrow \neg D$	Truth
3.	$D \rightarrow \neg D$	PC: 1, 2
4.	$\neg D$	PC: 3
5′.	$I_{i+1}(K_i(\neg D) \rightarrow \neg D, \neg D)$	2-4
6.	$K_{i+1}(K_i(\neg D) \rightarrow \neg D)$	
7.	$(K_{i+1}(K_i(\neg D) \rightarrow \neg D) \wedge I_{i+1}(K_i(\neg D) \rightarrow \neg D, \neg D)) \rightarrow K_{i+1}(\neg D)$	I
8.	$K_{i+1}(\neg D)$	PC: 5′, 6, 7
9.	D	PC: 1, 8
10.	$\bot$	4, 9

Solutions to the Knower Paradox

Paul Égré (2005). *The Knower Paradox in the Light of Provability Interpretations of Modal*. Journal of Logic, Language and Information, 14, pp. 13 - 48.

Solutions to the Knower Paradox

Paul Égré (2005). *The Knower Paradox in the Light of Provability Interpretations of Modal*. Journal of Logic, Language and Information, 14, pp. 13 - 48.

Francesa Poggiolesi (2007). *Three Different Solutions to the Knower Paradox*. Annali del Dipartimento di Filosofia, 13(1), pp. 147 - 163.

Solutions to the Knower Paradox

Paul Égré (2005). *The Knower Paradox in the Light of Provability Interpretations of Modal*. Journal of Logic, Language and Information, 14, pp. 13 - 48.

Francesa Poggiolesi (2007). *Three Different Solutions to the Knower Paradox*. Annali del Dipartimento di Filosofia, 13(1), pp. 147 - 163.

Mirjam de Vos, Rineke Verbrugge, and Barteld Kooi (2023). *Solutions to the Knower Paradox in the Light of Haack's Criteria*. Journal of Philosophical Logic, 52, pp. 1101 - 1132.