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 $p \in \mathrm{AT}$ (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 $R \subseteq W \times W$.

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

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 $\mathbf{T}$ 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

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

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

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

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

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.	$\Box(\ulcorner D \urcorner) \rightarrow D$	Truth
3.	$\Box(\ulcorner D \urcorner) \rightarrow \neg\Box(\ulcorner D \urcorner)$	PC: 1, 2
4.	$\neg\Box(\ulcorner D \urcorner)$	PC: 3

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.	$\perp$	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; $(K(D) \to D)$

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; $(K(D) \to D)$
    - 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. $(K(D) \to \neg K(D))$
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. $(\perp)$

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.	$K(\neg D) \rightarrow \neg D$	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.	$K(\neg D) \to \neg D$	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

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

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

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

(T) $K(\varphi) \to \varphi$

(U) $K(K(\varphi) \to \varphi)$

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

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

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

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.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \to F)))$	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
10.	$D \to \neg K(D \to F)$	PC: 4

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

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.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \to F)))$	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

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

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.	$D \leftrightarrow (K\neg D \vee (F \wedge \neg K(D \rightarrow F)))$	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
15.	D	PC: 1, 14
16.	$\perp$	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)	$B(\varphi) \rightarrow B(B(\varphi))$
(D)	$B(\neg\varphi) \rightarrow \neg B(\varphi)$
(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.	$\perp$	6, 8

Theorem (Cross 2001)

Let $\mathbf{T}$ be an axiomatizable theory extending $\mathbf{Q}$, 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 $\mathbf{T}$. Suppose $\mathbf{T}$ satisfies the following axiom schemata:

(T′) $K'(\varphi) \to \varphi$

(U′) $K'(K'(\varphi) \to \varphi)$

then $\mathbf{T}$ is inconsistent.

We show that $\mathbf{T} \vdash (K'(\varphi) \wedge I(\varphi, \psi)) \to K'(\psi)$, 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 $\mathbf{T} \vdash (K'(\varphi) \wedge I(\varphi, \psi)) \rightarrow K'(\psi)$, 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

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
4.	$K'(\varphi) \land I(\varphi, \psi) \leftrightarrow K'(\psi)$	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:

$$
\begin{array}{ll}
(\mathrm{T}') & K'(\varphi) \to \varphi \\
(\mathrm{U}^{+}) & K(K'(\varphi) \to \varphi)
\end{array}
$$

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 $\mathbf{T}$:

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

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

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

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

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

Blocking the Knower Paradox

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

Knower Paradox in the Quantified Logic of Proofs

W. Dean (2014). *Montague's paradox, informal provability, and explicit modal logic*. Notre Dame Journal of Formal Logic, 55(2), pp. 157 - 196.

W. Dean and H. Kurokawa (2014). *The paradox of the Knower revisited*. Annals of Pure and Applied Logic, 165(1), pp. 199 - 224.

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
- ▶ A Primer on Epistemic and Doxastic Logic
- ▶ Anti-Expert Paradox, and related paradoxes
- ▶ Predicate approach to modality, continued
- ▶ Epistemic Arithmetic
- ▶ Gödel's Disjunction

Doxastic Logic: Models

Model: $\langle W, R, V \rangle$

States/possible worlds: $W \neq \emptyset$

Quasi-partitions: $R \subseteq W \times W$ is serial, transitive and Euclidean

Doxastic Logic: Models

Model: $\langle W, R, V \rangle$

States/possible worlds: $W \neq \emptyset$

Quasi-partitions: $R \subseteq W \times W$ is serial, transitive and Euclidean

- *serial*: for all $w \in W$, there is a $v \in W$ such that $w \, R \, v$
- *transitive*: for all $w, v, u \in W$, if $w \, R \, v$ and $v \, R \, u$, then $w \, R \, u$
- *Euclidean*: for all $w, v, u \in W$, if $w \, R \, v$ and $w \, R \, u$, then $v \, R \, u$

Doxastic Logic: Models

Model: $\langle W, R, V \rangle$

States/possible worlds: $W \neq \emptyset$

Quasi-partitions: $R \subseteq W \times W$ is serial, transitive and Euclidean

- *serial*: for all $w \in W$, there is a $v \in W$ such that $w \, R \, v$
- *transitive*: for all $w, v, u \in W$, if $w \, R \, v$ and $v \, R \, u$, then $w \, R \, u$
- *Euclidean*: for all $w, v, u \in W$, if $w \, R \, v$ and $w \, R \, u$, then $v \, R \, u$

Valuation function: $V : \mathrm{At} \rightarrow \wp(W)$, where At is a set of atomic propositions.

Doxastic Logic: Language and Semantics

$$p \mid \varphi \wedge \varphi \mid \neg\varphi \mid B\varphi$$

Doxastic Logic: Language and Semantics

$$p \mid \varphi \wedge \varphi \mid \neg\varphi \mid B\varphi$$

Boolean connectives:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff it is not the case that $\mathcal{M}, w \models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$

Doxastic Logic: Language and Semantics

$$p \mid \varphi \wedge \varphi \mid \neg\varphi \mid B\varphi$$

Boolean connectives:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff it is not the case that $\mathcal{M}, w \models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$

Belief operators: $\mathcal{M}, w \models B\varphi$ iff for all v, if $w\, R\, v$, then $\mathcal{M}, v \models \varphi$.

Doxastic Logic: Language and Semantics

$$p \mid \varphi \wedge \varphi \mid \neg\varphi \mid B\varphi$$

Boolean connectives:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff it is not the case that $\mathcal{M}, w \models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$

Belief operators: $\mathcal{M}, w \models B\varphi$ iff for all v, if $w\, R\, v$, then $\mathcal{M}, v \models \varphi$.

$$\mathcal{M}, w \models B\varphi \text{ iff } R(w) \subseteq [\![\varphi]\!]^{\mathcal{M}}$$

Doxastic Logic: Language and Semantics

$$p \mid \varphi \wedge \varphi \mid \neg\varphi \mid B\varphi$$

Boolean connectives:

- $\mathcal{M}, w \models p$ iff $w \in V(p)$
- $\mathcal{M}, w \models \neg\varphi$ iff it is not the case that $\mathcal{M}, w \models \varphi$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$

Belief operators: $\mathcal{M}, w \models B\varphi$ iff for all v, if $w\, R\, v$, then $\mathcal{M}, v \models \varphi$.

$$\mathcal{M}, w \models B\varphi \text{ iff } \underbrace{R(w)}_{\{v \mid w\, R\, v\}} \subseteq \underbrace{[\![\varphi]\!]^{\mathcal{M}}}_{\{v \mid \mathcal{M}, w \models \varphi\}}$$

Doxastic Logic: **KD45**

K	$B(\varphi \to \psi) \to (B\varphi \to B\psi)$
D	$B\varphi \to \neg B \neg \varphi$
4	$B\varphi \to BB\varphi$
5	$\neg B\varphi \to B\neg B\varphi$

Doxastic Logic: **KD45**

$$
\begin{array}{ll}
K & B(\varphi \to \psi) \to (B\varphi \to B\psi) \\
D & B\varphi \to \neg B \neg \varphi \\
4 & B\varphi \to BB\varphi \\
5 & \neg B\varphi \to B\neg B\varphi
\end{array}
$$

The logic **KD45** adds the above axiom schemes to an axiomatization of classical propositional logic with the rules Modus Ponens, Substitution of Equivalents, and Necessitation (from φ infer $B\varphi$).

KD45 is sound and strongly complete with respect to all quasi-partition frames.

Exercise: Show that the following axiom schemes and rules are valid on quasi-partition models and are theorems of **KD45**:

- agglomeration: $(B\varphi \wedge B\psi) \to B(\varphi \wedge \psi)$
- consistency: $\neg B\bot$
- monotonicity: From $\varphi \to \psi$ infer $B\varphi \to B\psi$

Exercise: Show that the following axiom schemes and rules are valid on quasi-partition models and are theorems of **KD45**:

- agglomeration: $(B\varphi \wedge B\psi) \to B(\varphi \wedge \psi)$
- consistency: $\neg B\bot$
- monotonicity: From $\varphi \to \psi$ infer $B\varphi \to B\psi$
- secondary-reflexivity: for all $w, v \in W$, if $w\ R\ v$ then $v\ R\ v$
  $B(B\varphi \to \varphi)$

Exercise: Show that the following axiom schemes and rules are valid on quasi-partition models and are theorems of **KD45**:

- agglomeration: $(B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$
- consistency: $\neg B\bot$
- monotonicity: From $\varphi \rightarrow \psi$ infer $B\varphi \rightarrow B\psi$
- secondary-reflexivity: for all $w, v \in W$, if $w\,R\,v$ then $v\,R\,v$
  $B(B\varphi \rightarrow \varphi)$
- correctness of own beliefs:
  $B\neg B\varphi \rightarrow \neg B\varphi$
  for all w, there is a v such that $w\,R\,v$ and for all z if $v\,R\,z$ then $w\,R\,z$
  $BB\varphi \rightarrow B\varphi$
  density: for all w and v if $w\,R\,v$ then there is a z such that $w\,R\,z$ and $z\,R\,v$

Buridan-Burge Paradox I

Suppose that q is the statement that $\neg B_a q$.

Buridan-Burge Paradox I

Suppose that q is the statement that $\neg B_a q$. Now, either $B_a q$ or $\neg B_a q$.

Buridan-Burge Paradox I

Suppose that q is the statement that $\neg B_a q$. Now, either $B_a q$ or $\neg B_a q$.

1. Suppose $\neg B_a q$. Then by the 5 axiom ($\neg B_a \varphi \rightarrow B_a \neg B_a \varphi$), we have that $B_a \neg B_a q$. But since q is $\neg B_a q$, we have $B_a q$. Contradiction.

Buridan-Burge Paradox I

Suppose that q is the statement that $\neg B_a q$. Now, either $B_a q$ or $\neg B_a q$.

1. Suppose $\neg B_a q$. Then by the 5 axiom ($\neg B_a \varphi \to B_a \neg B_a \varphi$), we have that $B_a \neg B_a q$. But since q is $\neg B_a q$, we have $B_a q$. Contradiction.
2. Suppose $B_a q$. By by the 4 axiom ($B_a \varphi \to B_a B_a \varphi$), we have that $B_a B_a q$. By the D axioms ($B_a \varphi \to \neg B_a \neg \varphi$), we have that $\neg B_a \neg B_a q$. But since $\neg B_a q$ is q, we have $\neg B_a q$. Contradiction.

Tyler Burge (1984). *Epistemic paradox*. Journal of Philosophy, 81(1), pp. 5 - 29.

Buridan-Burge Paradox II

Of course, "q is the statement that $\neg B_a q$" is not a sentence of the modal logic of beliefs.

What we have shown is that $\neg B_a(q \leftrightarrow \neg B_a q)$ is a theorem of **KD45**.

This is a paradox only if it should be possible for an ideally rational agent to believe that $q \leftrightarrow \neg B_a q$.

Wolfgang Lenzen (1981). *Doxastic Logic and the Burge-Buridan-Paradox*. Philosophical Studies, 39(1), pp. 43 - 49.

Michael Caie (2012). *Belief and indeterminacy*. The Philosophical Review, 121(1), pp. 1 - 54.

Propositional Quantifiers

While we naturally quantify over propositions in both ordinary and philosophical discussion of beliefs, the addition of propositional quantifiers is not given much attention in the literature.

Propositional Quantifiers

While we naturally quantify over propositions in both ordinary and philosophical discussion of beliefs, the addition of propositional quantifiers is not given much attention in the literature. Consider the following examples:

- "One believes that everything one believes is true": $B\forall p(Bp \to p)$
- "If no matter what p stands for, one believes that φ, then one believes that no matter what p stands for, φ": $\forall pB\varphi \to B\forall p\varphi$
- "There is a proposition that the agent takes to be consistent and to settle everything": $\exists q(Bq \wedge \forall p(B(q \to p) \vee B(q \to \neg p))$

See the course by Peter Fritz.

Immodest Beliefs

Immod: "One believes that everything one believes is true": $B\forall p(Bp \to p)$

- Even for idealized agents or idealized beliefs, as axiomatized by **KD45**, it seems that Immod should not be included in a logic of belief.

Immodest Beliefs

Immod: "One believes that everything one believes is true": $B\forall p(Bp \to p)$

- Even for idealized agents or idealized beliefs, as axiomatized by **KD45**, it seems that Immod should not be included in a logic of belief.
- `Immod` should be distinguished from "for every proposition p, one believes that if she believes that p then p": $\forall p(B(Bp \to p))$.

Immodest Beliefs

Immod: "One believes that everything one believes is true": $B\forall p(Bp \to p)$

- Even for idealized agents or idealized beliefs, as axiomatized by **KD45**, it seems that Immod should not be included in a logic of belief.
- `Immod` should be distinguished from "for every proposition p, one believes that if she believes that p then p": $\forall p(B(Bp \to p))$.

Consider an agent who has credences about a real number x randomly generated from the interval $[0, 1]$. For all measurable $X \subseteq [0, 1]$, then the agent's credence that $x \in X$ is just the measure of X. Suppose that the agent outright believes precisely those propositions with credence 1. Then, for all $a \in [0, 1]$, the agent believes that $x \in [0, 1] \setminus \{a\}$ since $[0, 1] \setminus \{a\}$ is measure 1. However, the agent does not believe that for all $a \in [0, 1]$, $x \in [0, 1] \setminus \{a\}$ since $\bigcap_{a \in [0,1]} ([0, 1] \setminus \{a\}) = \varnothing$, which is not measure 1. Hence the agent in this situation does not believe that all her beliefs are true.

Yifeng Ding (2021). *On the Logic of Belief and Propositional Quantification*. Journal of Philosophical Logic, 50, pp. 1143 - 1198.

In any possible world semantics for **KD45**, $B\forall p(Bp \to p)$ is valid on any frame. So, any logic validating **KD45** must validate Immod. Algebraic semantics is needed for logics that do not validate Immod.

Yifeng Ding (2021). *On the Logic of Belief and Propositional Quantification*. Journal of Philosophical Logic, 50, pp. 1143 - 1198.

Also, see:

Jeremy Goodman (2020). *I'm mistaken*. manuscript.

Prior's Theorem

$$Q(\forall p(Qp \to \neg p)) \to (\exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p))$$

is a derivable using Universal Instantiation and propositional reasoning.

A. N. Prior. *On a family of paradoxes*. Notre Dame Journal of Formal Logic, 2(1), pgs. 16 - 32, 1961.

Prior's Theorem

$T1$. $C(UpCdpNp)\,C(dUpCdpNp)(NUpCdpNp)$ – from $CUpdpdq$ by substitution.

$T2$. $C(dUpCdpNp)\,C(UpCdpNp)\,(NUpCdpNp)$ – from $T1$ and $CCpCqrCqCpr$.

$T3$. $C(dUpCdpNp)\,(NUpCdpNp)$ – from $T2$ and $CCpCqNqCpNq$.

$T4$. $C(dUpCdpNp)(EpKdpp)$ – from $T3$ and equivalence of 'not-none' and 'some', i.e. of 'not-all-not' and 'some'.

$T5$. $C(dUpCdpNp)\,K(dUpCdpNp)\,(NUpCdpNp)$ – from $T3$ and $CCpqCpKpq$.

$T6$. $CK(dUpCdpNp)(NUpCdpNp)(EpKdpNp)$ – substitution in $CdqEpdp$.

$T7$. $C(dUpCdpNp)(EpKdpNp)$ – syllogistically from $T5$ and $T6$.

$T8$. $C(dUpCdpNp)\,K(EpKdpp)(EpKdpNp)$ – from $T4$, $T7$ and $CCpqCCprCpKqr$.

Prior's Theorem

1. $\forall p (Qp \to \neg p) \to (Q(\forall p(Qp \to \neg p)) \to \neg \forall p(Qp \to \neg p))$
   $(\forall p \varphi(p) \to \varphi[p/q])$

Prior's Theorem

1. $\forall p(Qp \to \neg p) \to (Q(\forall p(Qp \to \neg p)) \to \neg\forall p(Qp \to \neg p))$

   $(\forall p \varphi(p) \to \varphi[p/q])$

2. $Q(\forall p(Qp \to \neg p)) \to (\forall p(Qp \to \neg p) \to \neg\forall p(Qp \to \neg p))$

   $((a \to (b \to c)) \to (b \to (a \to c)))$

Prior's Theorem

1. $\forall p\,(Qp \to \neg p) \to (Q(\forall p(Qp \to \neg p)) \to \neg \forall p(Qp \to \neg p))$

   $(\forall p \varphi(p) \to \varphi[p/q])$

2. $Q(\forall p(Qp \to \neg p)) \to (\forall p(Qp \to \neg p) \to \neg\forall p(Qp \to \neg p))$

   $((a \to (b \to c)) \to (b \to (a \to c)))$

3. $Q(\forall p(Qp \to \neg p)) \to \neg\forall p(Qp \to \neg p)$

   $((a \to (b \to \neg b)) \to (a \to \neg b))$

Prior's Theorem

1. $\forall p\,(Qp \rightarrow \neg p) \rightarrow (Q(\,\forall p(Qp \rightarrow \neg p)\,) \rightarrow \neg\,\forall p(Qp \rightarrow \neg p)\,)$

   $(\forall p \varphi(p) \rightarrow \varphi[p/q])$

2. $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow (\,\forall p(Qp \rightarrow \neg p) \rightarrow \neg\forall p(Qp \rightarrow \neg p)\,)$

   $((\,a \rightarrow (b \rightarrow c)) \rightarrow (b \rightarrow (\,a \rightarrow c)))$

3. $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \boxed{\neg\forall p(Qp \rightarrow \neg p)}$

   $((a \rightarrow (b \rightarrow \neg b)) \rightarrow (a \rightarrow \neg b))$

4. $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \boxed{\exists p(Qp \wedge p)}$

   ($\boxed{\neg\forall p\varphi \leftrightarrow \exists p\neg\varphi}$ and $\boxed{\neg(a \rightarrow \neg b) \leftrightarrow (a \wedge b)}$)

Prior's Theorem

1. $\forall p\,(Qp \to \neg p) \to (Q(\forall p(Qp \to \neg p)) \to \neg\forall p(Qp \to \neg p))$

   $(\forall p \varphi(p) \to \varphi[p/q])$

2. $Q(\forall p(Qp \to \neg p)) \to (\forall p(Qp \to \neg p) \to \neg\forall p(Qp \to \neg p))$

   $((a \to (b \to c)) \to (b \to (a \to c)))$

3. $Q(\forall p(Qp \to \neg p)) \to \neg\forall p(Qp \to \neg p)$

   $((a \to (b \to \neg b)) \to (a \to \neg b))$

4. $Q(\forall p(Qp \to \neg p)) \to \exists p(Qp \wedge p)$

   $(\neg\forall p\varphi \leftrightarrow \exists p\neg\varphi$ and $\neg(a \to \neg b) \leftrightarrow (a \wedge b))$

5. $Q(\forall p(Qp \to \neg p)) \to (Q(\forall p(Qp \to \neg p)) \wedge \neg\forall p(Qp \to \neg p))$

   $((a \to b) \to (a \to (a \wedge b)))$

Prior's Theorem

1. $\forall p\,(Qp \to \neg p) \to (Q(\forall p(Qp \to \neg p)) \to \neg\forall p(Qp \to \neg p))$
   $(\forall p\varphi(p) \to \varphi[p/q])$

2. $Q(\forall p(Qp \to \neg p)) \to (\forall p(Qp \to \neg p) \to \neg\forall p(Qp \to \neg p))$
   $((a \to (b \to c)) \to (b \to (a \to c)))$

3. $Q(\forall p(Qp \to \neg p)) \to \neg\forall p(Qp \to \neg p)$
   $((a \to (b \to \neg b)) \to (a \to \neg b))$

4. $Q(\forall p(Qp \to \neg p)) \to \exists p(Qp \wedge p)$
   $(\neg\forall p\varphi \leftrightarrow \exists p\neg\varphi$ and $\neg(a \to \neg b) \leftrightarrow (a \wedge b))$

5. $Q(\forall p(Qp \to \neg p)) \to (Q(\forall p(Qp \to \neg p)) \wedge \neg\forall p(Qp \to \neg p))$
   $(a \to b) \to (a \to (a \wedge b))$

6. $Q(\forall p(Qp \to \neg p)) \to \exists p(Qp \wedge \neg p)$
   $((Q\varphi \wedge \neg\varphi) \to \exists p(Qp \wedge \neg p))$

Prior's Theorem

4. $Q(\forall p(Qp \to \neg p)) \to \boxed{\exists p(Qp \land p)}$

   $(\ \neg\forall p\varphi \leftrightarrow \exists p\neg\varphi$ and $\neg(a \to \neg b) \leftrightarrow (a \land b))$

5. $Q(\forall p(Qp \to \neg p)) \to (\ Q(\forall p(Qp \to \neg p)) \land \neg\forall p(Qp \to \neg p)\ )$

   $(a \to b) \to (a \to (a \land b))$

6. $Q(\forall p(Qp \to \neg p)) \to \boxed{\exists p(Qp \land \neg p)}$

   $((Q\varphi \land \neg\varphi) \to \exists p(Qp \land \neg p))$

7. $Q(\forall p(Qp \to \neg p)) \to (\boxed{\exists p(Qp \land p)} \land \boxed{\exists p(Qp \land \neg p)})$

   $(((a \to \boxed{b}) \land (a \to \boxed{c})) \to (a \to (\boxed{b} \land \boxed{c})))$

$$Q(\forall p(Qp \to \neg p)) \to (\exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p))$$

$$Q(\forall p(Qp \to \neg p)) \to (\exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p))$$

- $Q\varphi :=$ Ann believes that φ

  If Ann believes that everything that Ann believes is wrong, then Ann believes something true and Ann believes something wrong.

$$Q(\forall p(Qp \to \neg p)) \to (\exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p))$$

- $Q\varphi :=$ Ann believes that φ

  If Ann believes that everything that Ann believes is wrong, then Ann believes something true and Ann believes something wrong.

- $Q\varphi :=$ Ann says that φ

  If Ann says that everything that Ann says is wrong, then Ann says something true and Ann says something wrong.

$$Q(\forall p(Qp \to \neg p)) \to (\exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p))$$

- $Q\varphi :=$ Ann believes that φ

  If Ann believes that everything that Ann believes is wrong, then Ann believes something true and Ann believes something wrong.

- $Q\varphi :=$ Ann says that φ

  If Ann says that everything that Ann says is wrong, then Ann says something true and Ann says something wrong.

- $Q\varphi :=$ Ann wrote on the board at midnight that φ

  If Ann wrote on the board at midnight that everything that Ann wrote on the board at midnight is wrong, then Ann wrote a true thing on the board at midnight and Ann wrote a false thing on the board at midnight.

A. Bacon, J. Hawthorne and G. Uzquiano. *Higher-Order Free Logic and the Prior-Kaplan Paradox*. Forthcoming in *Williamson on Modality*.

A. Bacon and G. Uzquiano. *Some results on the limits of thought*. Journal of Philosophical Logic, 2018.

R. H. Thomason and D. Tucker. *Paradoxes of Intensionality*. Review of Symbolic Logic, 4, pgs. 394 - 411, 2011.

S5

K	$K(\varphi \to \psi) \to (K\varphi \to K\psi)$
T	$K\varphi \to \varphi$
4	$K\varphi \to KK\varphi$
5	$\neg K\varphi \to K\neg K\varphi$

The logic **S5** adds the above axiom schemes to an axiomatization of classical propositional logic with the rules Modus Ponens, Substitution of Equivalents, and Necessitation (from φ infer $K\varphi$).

S5 is sound and strongly complete with respect to all partition frames.

S4

K	$K(\varphi \to \psi) \to (K\varphi \to K\psi)$
T	$K\varphi \to \varphi$
4	$K\varphi \to KK\varphi$

The logic **S4** adds the above axiom schemes to an axiomatization of classical propositional logic with the rules Modus Ponens, Substitution of Equivalents, and Necessitation (from φ infer $K\varphi$).

S4 is sound and strongly complete with respect to all reflexive and transitive frames.

A Problem with the Operator Approach

The operator approach suffers from a severe drawback: it restricts the expressive power of the language in a dramatic way because it rules out quantification in the following sense:

There is no direct formalisation of a sentence like

“All tautologies of propositional logic are necessary.”

- Substitutional quantification: $\forall A(P(A) \to \Box A)$, where P is a predicate and $\Box$ is an operator.

- Substitutional quantification: $\forall A(P(A) \to \Box A)$, where P is a predicate and $\Box$ is an operator. However, this quantification does not come with a semantics, only rules and axioms. Also, why are the following sentences formalized using different types of quantification?
  - "All Σ_1 sentences are provable"
  - "All Σ_1 sentences are necessary"

- Substitutional quantification: $\forall A(P(A) \to \Box A)$, where P is a predicate and $\Box$ is an operator. However, this quantification does not come with a semantics, only rules and axioms. Also, why are the following sentences formalized using different types of quantification?
  - "All Σ_1 sentences are provable"
  - "All Σ_1 sentences are necessary"

- Rather than "x is necessary", say "x is necessarily true". Thus, $\Box x$ is replaced by $\Box Tx$, where T is a truth predicate.

- Substitutional quantification: $\forall A(P(A) \to \Box A)$, where P is a predicate and $\Box$ is an operator. However, this quantification does not come with a semantics, only rules and axioms. Also, why are the following sentences formalized using different types of quantification?
  - "All Σ_1 sentences are provable"
  - "All Σ_1 sentences are necessary"

- Rather than "x is necessary", say "x is necessarily true". Thus, $\Box x$ is replaced by $\Box Tx$, where T is a truth predicate. However, there is the question about why should truth and necessity be treated differently at the syntactic level; and, this would mean that the theory of necessity would inherent all the semantical paradoxes.

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.

A **frame** is a tuple (W, R) where W is a nonempty set and R is a relation on W.

A **frame** is a tuple (W, R) where W is a nonempty set and R is a relation on W.

A **PW-model** is a triple (W, R, V) such that (W, R) is a frame and V assigns to every $w \in W$ as subset of $\mathcal{L}_\Box$ such that:

$$V(w) = \{A \in \mathcal{L}_\Box \mid \text{ for all } u \text{, if } w \, R \, u \text{, then } V(u) \models A\}$$

A **frame** is a tuple (W, R) where W is a nonempty set and R is a relation on W.

A **PW-model** is a triple (W, R, V) such that (W, R) is a frame and V assigns to every $w \in W$ as subset of $\mathcal{L}_\Box$ such that:

$$V(w) = \{A \in \mathcal{L}_\Box \mid \text{ for all } u, \text{ if } w\, R\, u, \text{ then } V(u) \models A\}$$

If (W, R, V) is a model, we say that the frame (W, R) **supports** the model (W, R, V) or that (W, R, V) is **based on** (W, R).

A frame **admits a valuation** if there is a valuation V such that (W, R, V) is model.

$$V(w) \models \Box\ulcorner A \urcorner \text{ iff for all } v \in W\text{, if } w\,R\,v\text{, then } V(v) \models A$$

$$V(w) \models \Box\ulcorner A \urcorner \text{ iff for all } v \in W, \text{ if } w\,R\,v, \text{ then } V(v) \models A$$

Characterization Problem: Which frames support PW-models?

$$V(w) \models \Box\ulcorner A \urcorner \text{ iff for all } v \in W, \text{ if } w\,R\,v, \text{ then } V(v) \models A$$

Characterization Problem: Which frames support PW-models?

Lemma (Normality). Suppose (W, R, V) is a PW-model, $w \in W$ and $A, B \in \mathcal{L}_\Box$. Then the following holds:

- If $V(u) \models A$ for all $u \in W$, then $V(w) \models \Box\ulcorner A \urcorner$.
- $V(w) \models \Box(\ulcorner A \to B \urcorner) \to (\Box\ulcorner A \urcorner \to \Box\ulcorner B \urcorner)$

$$\forall x \forall y((\mathsf{Sent}(x) \wedge \mathsf{Sent}(y)) \rightarrow (\Box \ulcorner x \dot{\rightarrow} y \urcorner \rightarrow (\Box x \rightarrow \Box y)))$$

Fact (Tarski). The above frame with one world that sees itself does not admit a valuation.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Assume (W, R, V) is a PW-model based on (W, R) which is reflexive.

- We have $\mathbf{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A \urcorner$, and so it holds at every world.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Assume (W, R, V) is a PW-model based on (W, R) which is reflexive.

- We have $\mathbf{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A \urcorner$, and so it holds at every world.
- If $V(w) \models \neg A$, then $V(w) \models \Box\ulcorner A \urcorner$.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Assume (W, R, V) is a PW-model based on (W, R) which is reflexive.

- We have $\mathbf{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A \urcorner$, and so it holds at every world.
- If $V(w) \models \neg A$, then $V(w) \models \Box\ulcorner A \urcorner$.
- So, by reflexivity, $V(w) \models A$. Contradiction.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Assume (W, R, V) is a PW-model based on (W, R) which is reflexive.

- We have $\mathbf{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A\urcorner$, and so it holds at every world.
- If $V(w) \models \neg A$, then $V(w) \models \Box\ulcorner A\urcorner$.
- So, by reflexivity, $V(w) \models A$. Contradiction.
- Thus, $V(w) \models A$.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Assume (W, R, V) is a PW-model based on (W, R) which is reflexive.

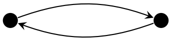
- We have $\mathbf{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A \urcorner$, and so it holds at every world.
- If $V(w) \models \neg A$, then $V(w) \models \Box\ulcorner A \urcorner$.
- So, by reflexivity, $V(w) \models A$. Contradiction.
- Thus, $V(w) \models A$.
- Hence, $V(w) \models \neg\Box\ulcorner A \urcorner$; and so, there is some u such that $w\,R\,u$ and $V(u) \models \neg A$.

Fact (Montague's Theorem). If (W, R) admits a valuation, then (W, R) is not reflexive.

Assume (W, R, V) is a PW-model based on (W, R) which is reflexive.

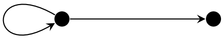
- We have $\mathbf{PA} \vdash A \leftrightarrow \neg\Box\ulcorner A \urcorner$, and so it holds at every world.
- If $V(w) \models \neg A$, then $V(w) \models \Box\ulcorner A \urcorner$.
- So, by reflexivity, $V(w) \models A$. Contradiction.
- Thus, $V(w) \models A$.
- Hence, $V(w) \models \neg\Box\ulcorner A \urcorner$; and so, there is some u such that $w\, R\, u$ and $V(u) \models \neg A$.
- Again, using the same argument as above, $V(u) \models A$. Contradiction.

1. The following frame does not admit a valuation:



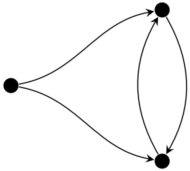
   Use the fixed point: $A \leftrightarrow \neg\Box\ulcorner\Box\ulcorner A\urcorner\urcorner$

2. The following frame does not admit a valuation:



   Use the fixed point: $A \leftrightarrow (\Box\ulcorner A\urcorner \rightarrow \Box\ulcorner\neg A\urcorner)$

3. The following frame does not admit a valuation:



Use the fixed point: $A \leftrightarrow (\neg\Box\ulcorner\Box\ulcorner A\urcorner\urcorner \wedge \neg\Box\ulcorner A\urcorner)$

4. The following frame $(\mathbb{N}, succ)$ does not admit a valuation:

$$0 \longrightarrow 1 \longrightarrow 2 \longrightarrow \cdots$$

Use the fixed point: $A \leftrightarrow \neg\forall x \Box \underset{\cdot}{h}(x, \ulcorner A \urcorner)$

where $\underset{\cdot}{h}$ represents a function that applies n-boxes to B:

$$h(n) = \ulcorner \Box \cdots \ulcorner \Box \ulcorner B \urcorner \urcorner \cdots \urcorner$$

V. McGee (1985). *How truthlike can a predicate be? A negative result*. Journal of Philosophical Logic, 14, pp. 399-410.

A. Visser (1989). *Semantics and the Liar paradox*. in Handbook of Philosophical Logic, Vol. 4, Reidel, Dordrecht.

Lemma. Let (W, R, V) be a PW-model based on a transitive frame. Then,

$$\Box\ulcorner A\urcorner \to \Box\ulcorner\Box\ulcorner A\urcorner\urcorner$$

obtains for all $w \in W$ and sentences $A \in \mathcal{L}_\Box$.

Löb's Theorem For every world w in a PW-model based on a transitive frame and every sentence $A \in \mathcal{L}_\Box$, the following holds:

$$\Box(\ulcorner\Box\ulcorner A\urcorner \to A\urcorner) \to \Box\ulcorner A\urcorner$$

Fact. In transitive frame admitting a valuation every world is either a dead end state or it can see a dead end state.

Fact. In transitive frame admitting a valuation every world is either a dead end state or it can see a dead end state.

Proof. Since the frame is transitive, Löb's Theorem holds.

Applying Löb's Theorem to $\perp$, we obtain:

$$V(w) \models \Box\ulcorner\perp\urcorner \vee \Diamond\ulcorner\Box\ulcorner\perp\urcorner\urcorner$$

Predicate Approaches to Modality

Johannes Stern (2016). *Toward Predicate Approaches to Modality*. Springer.