

# Epistemic Arithmetic

Eric Pacuit

University of Maryland

Lecture 4, ESSLLI 2025

July 31, 2025

# The Knower Paradox

## Theorem (Montague-Kaplan 1960)

Let  $\mathbf{T}$  be an axiomatizable extension of  $\mathbf{Q}$ , with  $I(x, y)$  a formula of expressing derivability between sentences in  $\mathbf{T}$ , and  $K$  a (perhaps complex) unary predicate satisfying, for all sentences  $\varphi$  and  $\psi$ :

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

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

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

then  $\mathbf{T}$  is inconsistent.

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

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  $\varphi$  and  $\psi$ :

$$(4) \quad B(\varphi) \rightarrow B(B(\varphi))$$

$$(D) \quad B(\neg\varphi) \rightarrow \neg B(\varphi)$$

$$(\text{Nec}) \quad \text{If } \mathbf{T} \vdash \varphi, \text{ then } \mathbf{T} \vdash B(\varphi)$$

$$(\text{Re}) \quad \text{If } \mathbf{T} \vdash \varphi \leftrightarrow \psi, \text{ then } \mathbf{T} \vdash B(\varphi) \leftrightarrow B(\psi)$$

then  $\mathbf{T}$  is inconsistent.

## 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:

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

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

then  $\mathbf{T}$  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  $\varphi$ :

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

$$(U^+) \quad K(K'(\varphi) \rightarrow \varphi)$$

then **T** is inconsistent.

# 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

$$\mathbf{T}_0 = \mathbf{Q} \cup \{K_0(\ulcorner \varphi \urcorner) \rightarrow \varphi \mid \varphi \in \mathcal{L}_\omega\}$$

$$\mathbf{T}_{i+1} = \mathbf{T}_i \cup \{K_{i+1}(\ulcorner \varphi \urcorner) \rightarrow \varphi \mid \varphi \in \mathcal{L}_\omega\}$$

$$V_0(K_0(\ulcorner \varphi \urcorner)) = 1 \text{ if and only if } \mathbf{Q} \vdash \varphi$$

$$V_{i+1}(K_{i+1}(\ulcorner \varphi \urcorner)) = 1 \text{ if and only if } \mathbf{T}_i \vdash \varphi$$

$$V_0(I_0(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner)) = 1 \text{ if and only if } \mathbf{Q} \vdash \varphi \rightarrow \psi$$

$$V_{i+1}(I_{i+1}(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner)) = 1 \text{ if and only if } \mathbf{T}_i \vdash \varphi \rightarrow \psi$$

$$\mathbf{T}_\omega = \bigcup_{i \in \omega} \mathbf{T}_i.$$

# 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  $V_j(\varphi \rightarrow \psi) = 1$ .
- ▶ 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

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

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

# Anderson's Solution

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

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

$$\begin{aligned}K_i(\ulcorner \varphi \urcorner) \rightarrow K_j(\ulcorner \varphi \urcorner) \text{ for } j \geq i. \\I_i(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner) \rightarrow I_j(\ulcorner \varphi \urcorner, \ulcorner \psi \urcorner) \text{ for } j \geq i.\end{aligned}$$

## 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. | $\perp$                                                                                                  | 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. | $\perp$                                                                                                                        | 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.

Francesca 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.

Francesca 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.

$K(\ulcorner A \urcorner)$  if and only if there exists a proof  $p$  which demonstrates  $A$

“...[W]e do not, at least *ipso facto*, wish to suggest that  $K(x)$  must be analyzed in terms of provability in a specific axiom system in the manner in which Cross’s definition of  $K'(x)$  might seem to suggest. For we might also adopt the view that in  $p$  should be understood as ranging over some class of informal proofs—i.e. intuitively correct pieces of reasoning which carry conviction for the epistemic subject in question.” (Dean and Kurokawa, p. 11)

# Logic of Proofs

S. Artemov and M. Fitting (2024). *Justification Logic*. Stanford Encyclopedia of Philosophy, <http://plato.stanford.edu/entries/logic-justification/>.

S. Artemov and M. Fitting. *Justification Logic: Reasoning with Reasons*. Cambridge University Press, 2019.

R. Kuznets and T. Studer (2019). *Logics of Proofs and Justifications*. College Publications.

$$t := c \mid x \mid !t \mid t \cdot s$$

$$\varphi := p \mid \neg \varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \varphi \rightarrow \psi \mid t : \varphi$$

$$t := c \mid x \mid !t \mid t \cdot s$$

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \varphi \rightarrow \psi \mid t : \varphi$$

- ▶  $t : \varphi$  means “ $t$  is a justification/proof of  $\varphi$ ”, or “ $\varphi$  is so for reason  $t$ ”.
- ▶  $x$  is a variable: an arbitrary justification
- ▶  $c$  is a constant: justifications of formulas we do not analyze further (axioms)
- ▶  $t \cdot u$  justifies  $\varphi$  whenever  $u$  justifies some formula  $\psi$  and  $t$  justifies  $\psi \rightarrow \varphi$
- ▶ if  $t$  justifies  $\varphi$ ,  $!t$  justifies that fact.

Factivity:  $t : \varphi \rightarrow \varphi$

Application:  $t : (\varphi \rightarrow \psi) \rightarrow (s : \varphi \rightarrow t \cdot s : \psi)$

Proof checker  $t : \varphi \rightarrow !t : t : \varphi$

Constant specification: a set of formulas of the form  $c_1 : c_2 : \dots : c_n : \varphi$  where  $\varphi$  is an instance of the axiom from the list above,  $n \geq 0$  and  $c_1, \dots, c_n$  are justification constants.



$$\vdash_{S4} (\Box P \vee \Box Q) \rightarrow \Box(\Box P \vee \Box Q)$$

$$\vdash_{LP} (x:P \vee y:Q) \rightarrow [a!\cdot x + b!\cdot y]:(x:P \vee y:Q)$$

*Note: 's + t' is a justification for everything that is justified by s or by t.*

# Quantified Logic of Proofs

- ▶  $(\forall x) (\forall x) ((x : (\varphi \rightarrow \psi) \wedge y : \varphi) \rightarrow x \cdot y : \psi)$
- ▶  $(\exists x) (x : \varphi \rightarrow \varphi)$
- ▶  $(\exists y) y : ((\exists x) (x : \varphi \rightarrow \varphi))$

M. Fitting (2008). *A quantified logic of evidence*. *Annals of Pure and Applied Logic*, 152(1-3), pp. 67-83.

# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$       Factivity, MP

# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$     Factivity, MP
2.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D \rightarrow \neg D$     Derivation in QLP

# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$       Factivity, MP
2.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D \rightarrow \neg D$       Derivation in QLP
3.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash \neg D$       1, 2

# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$       Factivity, MP
2.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D \rightarrow \neg D$       Derivation in QLP
3.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash \neg D$       1, 2
4.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash t(d):\neg D$       for some  $t(d)$  by Internalization

# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$       Factivity, MP
2.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D \rightarrow \neg D$       Derivation in QLP
3.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash \neg D$       1, 2
4.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash t(d):\neg D$       for some  $t(d)$  by Internalization
5.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D$       from 4, by  $\exists$ -intro

# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$       Factivity, MP
2.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D \rightarrow \neg D$       Derivation in QLP
3.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash \neg D$       1, 2
4.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash t(d):\neg D$       for some  $t(d)$  by Internalization
5.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D$       from 4, by  $\exists$ -intro
6.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D$       from 1, 5



# The Knower Paradox in the QLP

1.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D \leftrightarrow (\exists x) x:\neg D$  Factivity, MP
2.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D \rightarrow \neg D$  Derivation in QLP
3.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash \neg D$  1, 2
4.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash t(d):\neg D$  for some  $t(d)$  by Internalization
5.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash (\exists x) x:\neg D$  from 4, by  $\exists$ -intro
6.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash D$  from 1, 5
7.  $d:(D \leftrightarrow (\exists x) x:\neg D) \vdash \perp$  from 3, 6

# 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:  $V : \text{At} \rightarrow \wp(W)$ , where  $\text{At}$  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

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

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

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

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

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

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  $\mathcal{M}, v \models \varphi$ .

$$\mathcal{M}, w \models B\varphi \text{ iff } R(w) \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$$

# Doxastic Logic: Language and Semantics

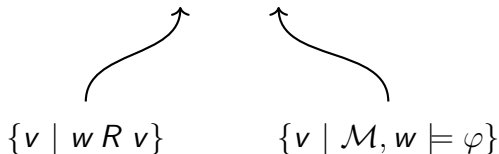
$$p \mid \varphi \wedge \psi \mid \neg \varphi \mid B\varphi$$

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  $\mathcal{M}, v \models \varphi$ .

$$\mathcal{M}, w \models B\varphi \text{ iff } R(w) \subseteq \llbracket \varphi \rrbracket^{\mathcal{M}}$$



# Doxastic Logic: **KD45**

$$K \quad B(\varphi \rightarrow \psi) \rightarrow (B\varphi \rightarrow B\psi)$$

$$D \quad B\varphi \rightarrow \neg B\neg\varphi$$

$$4 \quad B\varphi \rightarrow BB\varphi$$

$$5 \quad \neg B\varphi \rightarrow B\neg B\varphi$$

## Doxastic Logic: **KD45**

$$K \quad B(\varphi \rightarrow \psi) \rightarrow (B\varphi \rightarrow B\psi)$$

$$D \quad B\varphi \rightarrow \neg B\neg\varphi$$

$$4 \quad B\varphi \rightarrow BB\varphi$$

$$5 \quad \neg B\varphi \rightarrow B\neg B\varphi$$

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  $\varphi$  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:  $(B\varphi \wedge B\psi) \rightarrow B(\varphi \wedge \psi)$
- ▶ consistency:  $\neg B\perp$
- ▶ monotonicity: From  $\varphi \rightarrow \psi$  infer  $B\varphi \rightarrow B\psi$

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

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

**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:  $\neg B\perp$
- ▶ 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 ( $\neg B_a \varphi \rightarrow B_a \neg B_a \varphi$ ), we have that  $B_a \neg B_a q$ . But since  $q$  is  $\neg B_a q$ , we have  $B_a q$ . Contradiction.
2. Suppose  $B_a q$ . By the 4 axiom ( $B_a \varphi \rightarrow B_a B_a \varphi$ ), we have that  $B_a B_a q$ . By the D axioms ( $B_a \varphi \rightarrow \neg B_a \neg \varphi$ ), 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.

T. Raleigh (2021). *A New Anti-Expertise Dilemma*. *Synthese*, 199, pp. 5551 - 5569.

# Decision Instability

(DD)  $S$  Chooses  $\phi \leftrightarrow \phi$  does not maximize utility for  $S$

## Decision Instability

(DD)  $S$  Chooses  $\phi \leftrightarrow \phi$  does not maximize utility for  $S$

**Death in Damascus:** Death works from an appointment book which states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions. An appointment on the next day has been inscribed for him.

## Decision Instability

(DD)  $S$  Chooses  $\phi \leftrightarrow \phi$  does not maximize utility for  $S$

**Death in Damascus:** Death works from an appointment book which states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions. An appointment on the next day has been inscribed for him. Suppose, on this basis, the man would take his being in Damascus the next day as strong evidence that his appointment with Death is in Damascus, and would take his being in Aleppo the next day as strong evidence that his appointment is in Aleppo...



# Decision Instability

(DD)  $S$  Chooses  $\phi \leftrightarrow \phi$  does not maximize utility for  $S$

**Death in Damascus:** Death works from an appointment book which states time and place; a person dies if and only if the book correctly states in what city he will be at the stated time. The book is made up weeks in advance on the basis of highly reliable predictions. An appointment on the next day has been inscribed for him. Suppose, on this basis, the man would take his being in Damascus the next day as strong evidence that his appointment with Death is in Damascus, and would take his being in Aleppo the next day as strong evidence that his appointment is in Aleppo...If he decides to go to Aleppo, he then has strong grounds for expecting that Aleppo is where Death already expects him to be, and hence it is rational for him to prefer staying in Damascus. Similarly, deciding to stay in Damascus would give him strong grounds for thinking that he ought to go to Aleppo. (Gibbard & Harper, 1978, p. 373)

# Anti-Expert

$$(AE) \quad S \text{ Believes } p \leftrightarrow \neg p$$

# Anti-Expert

$$(AE) \quad S \text{ Believes } p \leftrightarrow \neg p$$

“I’m a neurologist, and know there’s a device that has been shown to induce the following state in people: they believe that their brains are in state  $\Sigma$  iff their brains are not in state  $\Sigma$ . I watch many trials with the device, and become extremely confident that it’s extremely reliable.

# Anti-Expert

$$(AE) \quad S \text{ Believes } p \leftrightarrow \neg p$$

“I’m a neurologist, and know there’s a device that has been shown to induce the following state in people: they believe that their brains are in state  $\Sigma$  iff their brains are not in state  $\Sigma$ . I watch many trials with the device, and become extremely confident that it’s extremely reliable. I’m also confident that my brain is not in state  $\Sigma$ .

# Anti-Expert

$$(AE) \quad S \text{ Believes } p \leftrightarrow \neg p$$

“I’m a neurologist, and know there’s a device that has been shown to induce the following state in people: they believe that their brains are in state  $\Sigma$  iff their brains are not in state  $\Sigma$ . I watch many trials with the device, and become extremely confident that it’s extremely reliable. I’m also confident that my brain is not in state  $\Sigma$ . Then the device is placed on my head and switched on. My confidence that my brain is not in state  $\Sigma$ ....well, it’s not clear here what should happen here.” (Christensen 2010, drawn from Conee 1982)

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

# Immodest Beliefs

Immod: “One believes that everything one believes is true”:  $B\forall p(Bp \rightarrow p)$

- ▶ 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”:  $B\forall p(Bp \rightarrow p)$

- ▶ 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  $p$ ”:  $\forall p(B(Bp \rightarrow p))$ .

# Immodest Beliefs

Immod: “One believes that everything one believes is true”:  $B\forall p(Bp \rightarrow p)$

- ▶ 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  $p$ ”:  $\forall p(B(Bp \rightarrow p))$ .

Consider an agent who has credences about a real number  $x$  randomly generated from the interval  $[0, 1]$ . For all measurable  $X \subseteq [0, 1]$ , 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  $\bigcap_{a \in [0, 1]} ([0, 1] \setminus \{a\}) = \emptyset$ , 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**,  $B\forall p(Bp \rightarrow p)$  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

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

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.*  $CK(dUpCdpNp)(NUpCdpNp)(EpKdpp)$  – substitution in  $CdqEpdp$ .
- T7.*  $C(dUpCdpNp)(EpKdpp)$  – syllogistically from *T5* and *T6*.
- T8.*  $C(dUpCdpNp) K(EpKdpp)(EpKdpp)$  – from *T4*, *T7* and  $CCpqCCprCpKqr$ .

# Prior's Theorem

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

# 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.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow (\forall p(Qp \rightarrow \neg p) \rightarrow \neg \forall p(Qp \rightarrow \neg p))$   
 $((a \rightarrow (b \rightarrow c)) \rightarrow (b \rightarrow (a \rightarrow c)))$



# 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.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \neg \forall p(Qp \rightarrow \neg p)$   
 $((a \rightarrow (b \rightarrow \neg b)) \rightarrow (a \rightarrow \neg b))$

## 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.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \neg \forall p(Qp \rightarrow \neg p)$   
 $((a \rightarrow (b \rightarrow \neg b)) \rightarrow (a \rightarrow \neg b))$
4.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \exists p(Qp \wedge p)$   
 $(\neg \forall p \varphi \leftrightarrow \exists p \neg \varphi \text{ and } \neg(a \rightarrow \neg b) \leftrightarrow (a \wedge b))$

# 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.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \neg \forall p(Qp \rightarrow \neg p)$   
 $((a \rightarrow (b \rightarrow \neg b)) \rightarrow (a \rightarrow \neg b))$
4.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \exists p(Qp \wedge p)$   
 $(\neg \forall p \varphi \leftrightarrow \exists p \neg \varphi \text{ and } \neg(a \rightarrow \neg b) \leftrightarrow (a \wedge b))$
5.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow (Q(\forall p(Qp \rightarrow \neg p)) \wedge \neg \forall p(Qp \rightarrow \neg p))$   
 $((a \rightarrow b) \rightarrow (a \rightarrow (a \wedge b)))$

# 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.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \neg \forall p(Qp \rightarrow \neg p)$   
 $((a \rightarrow (b \rightarrow \neg b)) \rightarrow (a \rightarrow \neg b))$
4.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \exists p(Qp \wedge p)$   
 $(\neg \forall p \varphi \leftrightarrow \exists p \neg \varphi \text{ and } \neg(a \rightarrow \neg b) \leftrightarrow (a \wedge b))$
5.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow (Q(\forall p(Qp \rightarrow \neg p)) \wedge \neg \forall p(Qp \rightarrow \neg p))$   
 $(a \rightarrow b) \rightarrow (a \rightarrow (a \wedge b))$
6.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \exists p(Qp \wedge \neg p)$   
 $((Q\varphi \wedge \neg \varphi) \rightarrow \exists p(Qp \wedge \neg p))$

# Prior's Theorem

4.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \exists p(Qp \wedge p)$   
(  $\neg \forall p \varphi \leftrightarrow \exists p \neg \varphi$  and  $\neg(a \rightarrow \neg b) \leftrightarrow (a \wedge b)$  )
5.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow ( Q(\forall p(Qp \rightarrow \neg p)) \wedge \neg \forall p(Qp \rightarrow \neg p) )$   
(  $a \rightarrow b \rightarrow (a \rightarrow (a \wedge b))$  )
6.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow \exists p(Qp \wedge \neg p)$   
(  $(Q\varphi \wedge \neg \varphi) \rightarrow \exists p(Qp \wedge \neg p)$  )
7.  $Q(\forall p(Qp \rightarrow \neg p)) \rightarrow ( \exists p(Qp \wedge p) \wedge \exists p(Qp \wedge \neg p) )$   
(  $((a \rightarrow b) \wedge (a \rightarrow c)) \rightarrow (a \rightarrow (b \wedge c))$  )

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

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

►  $Q\varphi :=$  Ann believes that  $\varphi$

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

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

- $Q\varphi := \text{Ann believes that } \varphi$

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

- $Q\varphi := \text{Ann says that } \varphi$

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



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

- ▶  $Q\varphi :=$  Ann believes that  $\varphi$

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  $\varphi$

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  $\varphi$

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

$$K \quad K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$$

$$T \quad K\varphi \rightarrow \varphi$$

$$4 \quad K\varphi \rightarrow KK\varphi$$

$$5 \quad \neg K\varphi \rightarrow K\neg K\varphi$$

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  $\varphi$  infer  $K\varphi$ ).

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

# S4

$$K \quad K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$$

$$T \quad K\varphi \rightarrow \varphi$$

$$4 \quad K\varphi \rightarrow KK\varphi$$

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  $\varphi$  infer  $K\varphi$ ).

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

# 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

## 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:  $\forall A(P(A) \rightarrow \Box A)$ , where  $P$  is a predicate and  $\Box$  is an operator.

- ▶ Substitutional quantification:  $\forall A(P(A) \rightarrow \Box A)$ , 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  $\Sigma_1$  sentences are provable”
  - ▶ “All  $\Sigma_1$  sentences are necessary”



- ▶ Substitutional quantification:  $\forall A(P(A) \rightarrow \Box A)$ , 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  $\Sigma_1$  sentences are provable”
  - ▶ “All  $\Sigma_1$  sentences are necessary”
- ▶ Rather than “ $x$  is necessary”, say “ $x$  is necessarily true”. Thus,  $\Box x$  is replaced by  $\Box T x$ , where  $T$  is a truth predicate.

- ▶ Substitutional quantification:  $\forall A(P(A) \rightarrow \Box A)$ , 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  $\Sigma_1$  sentences are provable”
  - ▶ “All  $\Sigma_1$  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 inherit all the semantical paradoxes.

# Operator > Predicate

- ✓ 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.
- ⇒ 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.

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  $\mathcal{L}_\square$  such that:

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

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  $\mathcal{L}_\square$  such that:

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

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.