# Reasoning about Knowledge and Beliefs

## Lecture 4

Eric Pacuit

University of Maryland, College Park

pacuit.org
epacuit@umd.edu

September 23, 2013

# Normal Modal Logics

A polymodal logic extending propositional logic with a set $\{\Box_i\}_{i \in I}$ of unary sentential operators is *normal* iff (i) for all $i \in I$,

$$\mathsf{RK}_i \ \frac{(\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \psi}{(\Box_i \varphi_1 \wedge \cdots \wedge \Box_i \varphi_m) \rightarrow \Box_i \psi}$$

is an admissible rule and (ii) the logic is closed under uniform substitution: if $\varphi$ is a theorem, so is the result of uniformly substituting formulas for the atomic sentences in $\varphi$.

## The "Problem" of Logical Omniscience

The rule

$$\text{RK}_i \ \frac{(\varphi_1 \wedge \cdots \wedge \varphi_m) \rightarrow \psi}{(K_i\varphi_1 \wedge \cdots \wedge K_i\varphi_m) \rightarrow K_i\psi}$$

reflects so-called (*synchronic*) *logical omniscience*: the agent knows (at time $t$) all the consequences of what she knows (at $t$).

# The "Problem" of Logical Omniscience

The rule

$$\mathsf{RK}_i \ \frac{(\varphi_1 \wedge \cdots \wedge \varphi_m) \to \psi}{(K_i\varphi_1 \wedge \cdots \wedge K_i\varphi_m) \to K_i\psi}$$

reflects so-called (*synchronic*) *logical omniscience*: the agent knows (at time $t$) all the consequences of what she knows (at $t$).

Given this, there are two ways to view $K_i$: as representing either the idealized (implicit, "virtual") knowledge of ordinary agents, or the ordinary knowledge of idealized agents. For discussion, see

R. Stalnaker.
1991. "The Problem of Logical Omniscience, I," *Synthese*.
2006. "On Logics of Knowledge and Belief," *Philosophical Studies*.

# The "Problem" of Logical Omniscience

The rule

$$\mathsf{RK}_i \ \frac{(\varphi_1 \wedge \cdots \wedge \varphi_m) \to \psi}{(K_i\varphi_1 \wedge \cdots \wedge K_i\varphi_m) \to K_i\psi}$$

reflects so-called (*synchronic*) *logical omniscience*: the agent knows (at time $t$) all the consequences of what she knows (at $t$).

There is now a large literature on alternative frameworks for representing the knowledge of agents with bounded rationality, who do not always "put two and two together" and therefore lack the logical omniscience reflected by $\mathsf{RK}_i$. See, for example:

J. Y. Halpern and R. Pucella. 2011. *Dealing with Logical Omniscience: Expressiveness and Pragmatics*. Artificial Intelligence.

# Logical Omniscience

- From $\varphi \leftrightarrow \psi$ infer $K_i\varphi \leftrightarrow K_i\psi$

# Logical Omniscience

- From $\varphi \leftrightarrow \psi$ infer $K_i\varphi \leftrightarrow K_i\psi$
- From $\varphi \rightarrow \psi$ infer $K_i\varphi \rightarrow K_i\psi$

## Logical Omniscience

- From $\varphi \leftrightarrow \psi$ infer $K_i\varphi \leftrightarrow K_i\psi$
- From $\varphi \rightarrow \psi$ infer $K_i\varphi \rightarrow K_i\psi$
- $(K_i(\varphi \rightarrow \psi) \wedge K_i\varphi) \rightarrow K_i\psi$

## Logical Omniscience

- From $\varphi \leftrightarrow \psi$ infer $K_i\varphi \leftrightarrow K_i\psi$
- From $\varphi \to \psi$ infer $K_i\varphi \to K_i\psi$
- $(K_i(\varphi \to \psi) \wedge K_i\varphi) \to K_i\psi$
- From $\varphi$ infer $K_i\varphi$

## Logical Omniscience

- From $\varphi \leftrightarrow \psi$ infer $K_i\varphi \leftrightarrow K_i\psi$
- From $\varphi \rightarrow \psi$ infer $K_i\varphi \rightarrow K_i\psi$
- $(K_i(\varphi \rightarrow \psi) \wedge K_i\varphi) \rightarrow K_i\psi$
- From $\varphi$ infer $K_i\varphi$
- $K_i\top$

## Logical Omniscience

- From $\varphi \leftrightarrow \psi$ infer $K_i\varphi \leftrightarrow K_i\psi$
- From $\varphi \rightarrow \psi$ infer $K_i\varphi \rightarrow K_i\psi$
- $(K_i(\varphi \rightarrow \psi) \wedge K_i\varphi) \rightarrow K_i\psi$
- From $\varphi$ infer $K_i\varphi$
- $K_i\top$
- $(K_i\varphi \wedge K_i\psi) \rightarrow K_i(\varphi \wedge \psi)$

# Dealing with Logical Omniscience

- *Syntactic approaches*: an agents knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);

# Dealing with Logical Omniscience

- *Syntactic approaches*: an agents knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);
- *Awareness*: an agent knows $\varphi$ if she is aware of $\varphi$ and $\varphi$ is true in all the worlds she considers possible;

# Dealing with Logical Omniscience

- *Syntactic approaches*: an agents knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);
- *Awareness*: an agent knows $\varphi$ if she is aware of $\varphi$ and $\varphi$ is true in all the worlds she considers possible;
- *Algorithmic knowledge*: an agent knows $\varphi$ if her knowledge algorithm returns "Yes" on a query of $\varphi$; and

# Dealing with Logical Omniscience

- *Syntactic approaches*: an agents knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);
- *Awareness*: an agent knows $\varphi$ if she is aware of $\varphi$ and $\varphi$ is true in all the worlds she considers possible;
- *Algorithmic knowledge*: an agent knows $\varphi$ if her knowledge algorithm returns "Yes" on a query of $\varphi$; and
- *Impossible worlds*: an agent may consider possible worlds that are logically inconsistent (for example, where $p$ and $\neg p$ may both be true).

# Dealing with Logical Omniscience

- *Syntactic approaches*: an agents knowledge is represented by a set of formulas (intuitively, the set of formulas she knows);
- *Awareness*: an agent knows $\varphi$ if she is aware of $\varphi$ and $\varphi$ is true in all the worlds she considers possible;
- *Algorithmic knowledge*: an agent knows $\varphi$ if her knowledge algorithm returns "Yes" on a query of $\varphi$; and
- *Impossible worlds*: an agent may consider possible worlds that are logically inconsistent (for example, where $p$ and $\neg p$ may both be true).

*Non-Normal Modal Logics*

## Dealing with Logical Omniscience

- *Syntactic approaches*: $\mathcal{M}, w \models K_i\varphi$ iff $\varphi \in \mathcal{C}_i(w)$

# Dealing with Logical Omniscience

- *Syntactic approaches*: $\mathcal{M}, w \models K_i\varphi$ iff $\varphi \in \mathcal{C}_i(w)$
- *Awareness structures*: $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $wR_iv$ then $\mathcal{M}, v \models \varphi$ and $\varphi \in \mathcal{A}_i(w)$

# Dealing with Logical Omniscience

- *Syntactic approaches*: $\mathcal{M}, w \models K_i\varphi$ iff $\varphi \in \mathcal{C}_i(w)$
- *Awareness structures*: $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $wR_iv$ then $\mathcal{M}, v \models \varphi$ and $\varphi \in \mathcal{A}_i(w)$
- *Algorithmic knowledge*: $\mathcal{M}, w \models K_i\varphi$ iff $\mathsf{A}_i(w, \varphi) = \mathrm{Yes}$

# Dealing with Logical Omniscience

- *Syntactic approaches*: $\mathcal{M}, w \models K_i\varphi$ iff $\varphi \in \mathcal{C}_i(w)$
- *Awareness structures*: $\mathcal{M}, w \models K_i\varphi$ iff for all $v \in W$, if $wR_iv$ then $\mathcal{M}, v \models \varphi$ and $\varphi \in \mathcal{A}_i(w)$
- *Algorithmic knowledge*: $\mathcal{M}, w \models K_i\varphi$ iff $\mathsf{A}_i(w, \varphi) = \mathrm{Yes}$
- *Impossible worlds*: $\mathcal{M}, w \models K_i\varphi$ iff if $w \in N$, then for all $v \in W$, if $wR_iv$ and $v \in N$ then $\mathcal{M}, v \models \varphi$

  $\mathcal{M}, w \models K_i\varphi$ iff if $w \notin N$, then $\varphi \in \mathcal{C}_i(w)$

# Justification Logic (1)

$t:\varphi$: "$t$ is a *justification/proof* for $\varphi$"

S. Artemov and M. Fitting. *Justification logic*. The Stanford Encyclopedia of Philosophy, 2012.

S. Artemov. *Explicit provability and constructive semantics*. The Bulletin of Symbolic Logic 7 (2001) 1 36.

M. Fitting. *The logic of proofs, semantically*. Annals of Pure and Applied Logic 132 (2005) 1 25.

# Justification Logic (2)

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

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

# Justification Logic (2)

$$t \ := c \mid t+s \mid \,!t \mid t \cdot s$$

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

Justification Logic:

- $t:\varphi \to \varphi$
- $t:(\varphi \to \psi) \to (s:\varphi \to t \cdot s:\psi)$
- $t:\varphi \to (t+s):\varphi$
- $t:\varphi \to (s+t):\varphi$
- $t:\varphi \to !t:t:\varphi$

# Justification Logic (2)

$$t := c \mid t + s \mid\, !t \mid t \cdot s$$

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

Justification Logic:

- $t : \varphi \rightarrow \varphi$
- $t : (\varphi \rightarrow \psi) \rightarrow (s : \varphi \rightarrow t \cdot s : \psi)$
- $t : \varphi \rightarrow (t + s) : \varphi$
- $t : \varphi \rightarrow (s + t) : \varphi$
- $t : \varphi \rightarrow !t : t : \varphi$

**Internalization**: if $\vdash_{JL} \varphi$ then there is a proof polynomial $t$ such that $\vdash_{JL} t : \varphi$
**Realization Theorem**: if $\vdash_{\mathbf{S4}} \varphi$ then there is a proof polynomial $t$ such that $\vdash_{JL} t : \varphi$

# Justification Logic (3)

**Fitting Semantics**: $\mathcal{M} = \langle W, R, \mathcal{E}, V \rangle$

- $W \neq \emptyset$
- $R \subseteq W \times W$
- $\mathcal{E} : W \times \text{ProofTerms} \to \wp(\mathcal{L}_{JL})$
- $V : \text{At} \to \wp(W)$

$\mathcal{M}, w \models t : \varphi$ iff for all $v$, if $wRv$ then $\mathcal{M}, v \models \varphi$ and $\varphi \in \mathcal{E}(w, t)$

# Justification Logic (3)

**Monotonicity** For all $w, v \in W$, if $wRv$ then for all proof polynomials $t$, $\mathcal{E}(w, t) \subseteq \mathcal{E}(v, t)$.

**Application** For all proof polynomials $s$, $t$ and for each $w \in W$, if $\varphi \to \psi \in \mathcal{E}(w, t)$ and $\varphi \in \mathcal{E}(w, s)$, then $\psi \in \mathcal{E}(w, t \cdot s)$

**Proof Checker** For all proof polynomials $t$ and for each $w \in W$, if $\varphi \in \mathcal{E}(w, t)$, then $t : \varphi \in \mathcal{E}(w, !t)$.

**Sum** For all proof polynomials $s, t$ and for each $w \in W$, $\mathcal{E}(w, s) \cup \mathcal{E}(w, t) \subseteq \mathcal{E}(w, s + t)$.

# Approaches

- Lack of awareness
- Lack of computational power
- Imperfect understanding of the model

# Epistemic Closure & the Skeptical Paradox

The problem of logical omniscience must be distinguished from the problem of epistemic closure, which arises even if we assume that our agents are perfect logicians who always "put two and two together" and deduce the consequence of what they know.

# Epistemic Closure & the Skeptical Paradox

The problem of logical omniscience must be distinguished from the problem of epistemic closure, which arises even if we assume that our agents are perfect logicians who always "put two and two together" and deduce the consequence of what they know.

The problem of epistemic closure is raised by the so-called *Skeptical Paradox*.

S. Cohen. 1988. "How to be a Fallibilist," *Philosophical Perspectives*.

K. DeRose. 1995. "Solving the Skeptical Problem," *Philosophical Review*.

## Epistemic Closure & the Skeptical Paradox

Let $p$ be a mundane proposition, e.g., Eric was born in the U.S., that we think our agent knows.

# Epistemic Closure & the Skeptical Paradox

Let $p$ be a mundane proposition, e.g., Eric was born in the U.S., that we think our agent knows.

Let $SH$ be a "skeptical hypothesis" (or a disjunction of hypotheses) incompatible with the truth of $p$, but according to which everything would be indistinguishable from the actual world for the agent, e.g., Russell's hypothesis that the world was created 5 minutes ago with everyone having false memories of a long past.

# Epistemic Closure & the Skeptical Paradox

Let $p$ be a mundane proposition, e.g., Eric was born in the U.S., that we think our agent knows.

Let $SH$ be a "skeptical hypothesis" (or a disjunction of hypotheses) incompatible with the truth of $p$, but according to which everything would be indistinguishable from the actual world for the agent, e.g., Russell's hypothesis that the world was created 5 minutes ago with everyone having false memories of a long past.

The skeptic argues that since the agent doesn't know $\neg SH$, but she does know the obvious fact that $p \to \neg SH$, it follows by RK$_i$ that she doesn't know $p$; i.e., $(Kp \wedge K(p \to \neg SH)) \to K\neg SH$ implies

$$(\neg K\neg SH \wedge K(p \to \neg SH)) \to \neg Kp.$$

Since $p$ was basically arbitrary, you don't know much of anything.

# Epistemic Closure & the Skeptical Paradox

The skeptic argues that since the agent doesn't know $\neg SH$, but she does know the obvious fact that $p \to \neg SH$, it follows by $\mathsf{RK}_i$ that she doesn't know $p$; i.e., $(Kp \wedge K(p \to \neg SH)) \to K\neg SH$ implies

$$(\neg K\neg SH \wedge K(p \to \neg SH)) \to \neg Kp.$$

Since $p$ was basically arbitrary, you don't know much of anything.

Three responses in defense of knowledge:

- Mooreanism: actually, you do know $\neg SH$ (How? One answer: because you know $p$ and know that $p \to \neg SH$. Too cheap?)
- Deny closure: $\mathsf{RK}_i$ is invalid; for the strange case of $SH$ versus $p$, we have $Kp$, $\neg K\neg SH$, and $K(p \to \neg SH)$.
- Contextualism: in a context where we're not worried about skepticism, we can truly claim $Kp$; in a context where we are, we can truly claim $\neg K\neg SH$; in every fixed context, $\mathsf{RK}_i$ holds.

# Epistemic Closure & the Skeptical Paradox

(1) Mooreanism: actually, you do know $\neg SH$ (How? One answer: because you know $p$ and know that $p \to \neg SH$. Too cheap?)

(2) Deny closure: $\mathrm{RK}_i$ is invalid; for the strange case of $SH$ against $p$, the truth is $Kp$, $\neg K \neg SH$, $K(p \to \neg SH)$.

(3) Contextualism: in a context where we're not worried about skepticism, we can truly claim $Kp$; in a context where we are, we can truly claim $\neg K \neg SH$; in every fixed context, $\mathrm{RK}_i$ holds.

The third option leads naturally to questions about how context is supposed to change as we consider skeptical possibilities. For modeling of this in the framework of dynamic epistemic logic, see:

Wes Holliday (http://philosophy.berkeley.edu/holliday). 2012.

"Epistemic Logic, Relevant Alternatives, and the Dynamics of Context."

# Epistemic Closure & the Skeptical Paradox

(2) Deny closure: $\mathsf{RK}_i$ is invalid; for the strange case of $SH$ against $p$, the truth is $Kp$, $\neg K \neg SH$, $K(p \to \neg SH)$.

The second option leads naturally to questions about what closure principles do hold, if closure under known implication does not. For example, shouldn't $K(\varphi \wedge \psi) \to K\varphi$ still be valid?

# Epistemic Closure & the Skeptical Paradox

(2) Deny closure: $\mathsf{RK}_i$ is invalid; for the strange case of $SH$ against $p$, the truth is $Kp$, $\neg K \neg SH$, $K(p \to \neg SH)$.

The second option leads naturally to questions about what closure principles do hold, if closure under known implication does not. For example, shouldn't $K(\varphi \wedge \psi) \to K\varphi$ still be valid? See:

W. Holliday (http://philosophy.berkeley.edu/holliday). 2012.

"Epistemic Closure and Epistemic Logic I:

Relevant Alternatives and Subjunctivism."

# Epistemic Closure & the Skeptical Paradox

(2) Deny closure: $\mathsf{RK}_i$ is invalid; for the strange case of $SH$ against $p$, the truth is $Kp$, $\neg K \neg SH$, $K(p \to \neg SH)$.

The second option leads naturally to questions about what closure principles do hold, if closure under known implication does not. For example, shouldn't $K(\varphi \wedge \psi) \to K\varphi$ still be valid? See:

W. Holliday (http://philosophy.berkeley.edu/holliday). 2012.
"Epistemic Closure and Epistemic Logic I:
Relevant Alternatives and Subjunctivism."

There are open questions about the complete logics of some famous theories of knowledge. See Problem 8.12 in the above.

## The Lottery and Preface Paradoxes

Another challenge to $\mathsf{RK}_i$ comes from the so-called *lottery paradox* and *preface paradox*.

H.E. Kyburg, Jr. 1961. *Probability and the Logic of Rational Belief*. Wesleyan University Press.

D.C. Makinson. 1965. "The Paradox of the Preface," *Analysis*.

## The Lottery Paradox

Let $\Box_i$ stand for **it is rational for *i* to believe that**. Consider:

$$(\Box_i p \wedge \Box_i q) \rightarrow \Box_i (p \wedge q),$$

which is obviously derivable from the RK rule for $\Box_i$.

# The Lottery Paradox

Let $\Box_i$ stand for **it is rational for *i* to believe that**. Consider:

$$(\Box_i p \wedge \Box_i q) \rightarrow \Box_i (p \wedge q),$$

which is obviously derivable from the RK rule for $\Box_i$.

There is a lottery with $n$ tickets, of which one will be drawn.

- Let $l_1$ stand for ‘lottery ticket 1 is the winning ticket’.
- Let $l_2$ stand for ‘lottery ticket 2 is the winning ticket’, etc.

# The Lottery Paradox

Let $\Box_i$ stand for **it is rational for *i* to believe that**. Consider:

$$(\Box_i p \wedge \Box_i q) \rightarrow \Box_i (p \wedge q),$$

which is obviously derivable from the RK rule for $\Box_i$.

There is a lottery with $n$ tickets, of which one will be drawn.

- Let $l_1$ stand for 'lottery ticket 1 is the winning ticket'.
- Let $l_2$ stand for 'lottery ticket 2 is the winning ticket', etc.

If $n$ is large, it seems rational to believe of any individual ticket $k$ that it is not the winning ticket: $\Box_i \neg l_k$.

## The Lottery Paradox

Let $\Box_i$ stand for **it is rational for $i$ to believe that**. Consider:

$$(\Box_i p \wedge \Box_i q) \rightarrow \Box_i (p \wedge q),$$

which is obviously derivable from the RK rule for $\Box_i$.

There is a lottery with $n$ tickets, of which one will be drawn.

- Let $l_1$ stand for 'lottery ticket 1 is the winning ticket'.
- Let $l_2$ stand for 'lottery ticket 2 is the winning ticket', etc.

If $n$ is large, it seems rational to believe of any individual ticket $k$ that it is not the winning ticket: $\Box_i \neg l_k$. Then according to the principle above, it's rational to believe that no ticket will win: $\Box_i \neg l_1 \wedge \cdots \wedge \Box_i \neg l_n$.

# The Lottery Paradox

Let $\Box_i$ stand for **it is rational for $i$ to believe that**. Consider:

$$(\Box_i p \wedge \Box_i q) \rightarrow \Box_i(p \wedge q),$$

which is obviously derivable from the RK rule for $\Box_i$.

There is a lottery with $n$ tickets, of which one will be drawn.

- Let $l_1$ stand for ‘lottery ticket 1 is the winning ticket’.
- Let $l_2$ stand for ‘lottery ticket 2 is the winning ticket’, etc.

If $n$ is large, it seems rational to believe of any individual ticket $k$ that it is not the winning ticket: $\Box_i \neg l_k$. Then according to the principle above, it's rational to believe that no ticket will win: $\Box_i(\neg l_1 \wedge \cdots \wedge \neg l_n)$.

## The Lottery Paradox

Let $\Box_i$ stand for **it is rational for *i* to believe that**. Consider:

$$(\Box_i p \wedge \Box_i q) \rightarrow \Box_i(p \wedge q),$$

which is obviously derivable from the RK rule for $\Box_i$.

There is a lottery with $n$ tickets, of which one will be drawn.

- Let $l_1$ stand for 'lottery ticket 1 is the winning ticket'.
- Let $l_2$ stand for 'lottery ticket 2 is the winning ticket', etc.

If $n$ is large, it seems rational to believe of any individual ticket $k$ that it is not the winning ticket: $\Box_i \neg l_k$. Then according to the principle above, it's rational to believe that no ticket will win: $\Box_i(\neg l_1 \wedge \cdots \wedge \neg l_n)$. But it's certain that one will win!

# Reasoning about High Probability

An uncontroversial example of a non-normal operator is **it is highly probable that**, for which $(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$ is invalid.

# Reasoning about High Probability

An uncontroversial example of a non-normal operator is **it is highly probable that**, for which $(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$ is invalid.

Again, for any lottery ticket $i$, it is highly probably that $i$ will loose. But then by repeated use of $(\Box p \wedge \Box q) \rightarrow \Box(p \wedge q)$, we could derive that it is highly probable that *all* tickets will loose, contradicting the fact that it is certain that one ticket will win.

# The Preface Paradox

An author writes a book with $n$ claims, $c_1, \ldots, c_n$, each of which the author checked carefully and therefore believes. Yet the author has written books before and realizes that errors are inevitable in any book; thus, in the preface he says something to the effect of "I thank ... for their help; but all the errors that remain are mine."

# The Preface Paradox

An author writes a book with $n$ claims, $c_1, \ldots, c_n$, each of which the author checked carefully and therefore believes. Yet the author has written books before and realizes that errors are inevitable in any book; thus, in the preface he says something to the effect of "I thank . . . for their help; but all the errors that remain are mine."

It seems that we have here a situation in which

$$\Box_i c_1 \land \cdots \land \Box_i c_n$$

and

$$\Box_i \neg (c_1 \land \cdots \land c_n),$$

which means that the set of propositions believed is *inconsistent*.

## The Preface Paradox

An author writes a book with $n$ claims, $c_1, \ldots, c_n$, each of which the author checked carefully and therefore believes. Yet the author has written books before and realizes that errors are inevitable in any book; thus, in the preface he says something to the effect of "I thank . . . for their help; but all the errors that remain are mine."

It seems that we have here a situation in which

$$\Box_i c_1 \wedge \cdots \wedge \Box_i c_n$$

and

$$\Box_i \neg(c_1 \wedge \cdots \wedge c_n),$$

which means that the set of propositions believed is *inconsistent*.

But is there anything irrational about the author so described?