# Philosophical Logic

Johns Hopkins University, Spring 2015

## Course Information

| | |
|---|---|
| **Instructor** | Justin Bledin<br>Assistant Professor<br>Philosophy Department<br>Gilman 206<br>`jbledin@jhu.edu` |
| **Office Hours** | Tu 3:15pm-4:15pm & by appt |
| **Class Code** | AS.150.467 |
| **Class Time** | MW 4:30pm-5:45pm |
| **Class Location** | Gilman 288 |

## Course Description

This course is a survey of various topics in philosophical logic. We begin with a review of the model theory of classical first-order logic. In our first unit, we will then move beyond the standard existential and universal quantifiers and consider generalized quantifiers, substitutional quantifiers, and plural quantification. In our second unit, we will investigate the theory of propositional modal logic, considering its syntax, semantics, proof theory, and some of its applications. In our third unit, we will investigate various formal approaches to defining truth. In our fourth unit, we will inquire into the nature and normativity of logical validity.

## Prerequisites

A solid understanding of classical first-order logic is essential. Students should also be familiar with basic methods of mathematical proof.

## Readings

We will work through the following texts during the semester:

- Jon Barwise and Robin Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4(2): 159–219, 1981.
- Justin Bledin. Logic informed. *Mind*, 123(490): 277–316, 2014.
- George Boolos. For every $A$ there is a $B$. *Linguistic Inquiry*, 12(3): 465–467, 1981.
- George Boolos. To be is to be a value of a variable (or to be some value of some variables). *Journal of Philosophy*, 81(8): 430–449, 1984a.
- George Boolos. Nonfirstorderizability again. *Linguistic Inquiry*, 15(2): 343, 1984b.
- John P. Burgess. No requirement of relevance. In Stewart Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 727–750. Oxford University Press, 2005.

- John Etchemendy. *The Concept of Logical Consequence*. Harvard University Press, Cambridge, 1990.
- Hartry Field. Truth and the unprovability of consistency. *Mind*, 115(459): 567–606, 2006.
- Hartry Field. What is the normative role of logic? *Proceedings of the Aristotelian Society Supplementary Volume*, LXXXIII: 251–268, 2009.
- Hartry Field. What is logical validity? In Colin Caret and Ole Hjortland, editors, *Foundations of Logical Consequence*. Oxford University Press, forthcoming.
- Gilbert Harman. *Change in View: Principles of Reasoning*. MIT Press, Cambridge, 1986.
- Richard G. Heck. Kripke's theory of truth. Unpublished manuscript.
- Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72(19): 690–716, 1975.
- Hannes Leitgeb. What theories of truth should be like (but cannot be). *Philosophical Compass*, 2(2): 276–290, 2007.
- David Lewis. Logic for equivocators. *Noûs*, 16(3): 431–441, 1982.
- Leonard Linsky. Two concepts of quantification. *Noûs*, 6(3): 224–239, 1972.
- John MacFarlane. In what sense (if any) is logic normative for thought? Unpublished manuscript.
- Florian Steinberger. Explosion and the normativity of logic. Forthcoming in *Mind*.
- Alfred Tarski. Der wahrheitsbegriff in den formalisierten sprachen. *Studia Philosophica*, 1, 1936a. Translated as 'The Concept of Truth in Formalized Languages' in *Logic, Semantics, Metamathematics*, pp. 152–278, 1956.
- Alfred Tarski. O pojciu wynikania logicznego. *Przeglad Filozoficzny*, 39, 1936b. Translated as 'On the Concept of Logical Consequence' in *Logic, Semantics, Metamathematics*, pp. 409–420, 1956.
- Seth Yalcin. Epistemic modals. *Mind*, 116(464): 983–1026, 2007.


These texts will be posted on Blackboard.

## Schedule

The following schedule projects the lectures over the course of the semester. It is subject to revision as the semester progresses. I want to keep things fairly flexible and allocate our time based on which topics you find the most interesting.

**Introduction** Jan 26

**Logic Review** Jan 28, Feb 2

**1. Quantifiers**

| | | |
|---|---|---|
| 1.1 Generalized Quantifiers | Barwise and Cooper 1981 | Feb 4 |
| 1.2 Substitutional Quantifiers | Linsky 1972 | Feb 9 |
| 1.3 Plural Quantifiers | Boolos 1981, 1984a & 1984b | Feb 11 |

**2. Modals**

| | | |
|---|---|---|
| 2.1 Propositional Modal Logic | Lecture Notes Only | Feb 16 & 18 |
| 2.2 The Modal Zoo | Lecture Notes Only | Feb 23 & 25 |
| 2.3 Temporal Logic | Lecture Notes Only | Mar 2 |
| 2.4 Deontic Logic | Lecture Notes Only | Mar 4 |
| 2.5 Epistemic Logic | Lecture Notes Only | Mar 9 & 11 |

**3. Truth Predicates**

| | | |
|---|---|---|
| 3.1 Tarski | Tarski 1936a | Mar 23 & 25 |
| 3.2 Kripke | Kripke 1975 / Heck *ms.* | Mar 30 |
| 3.3 Contemporary Truth Theories | Leitgeb 2007 | Apr 1 |

**4. Logical Consequence**

| | | |
|---|---|---|
| 4.1 Truth Preservation View of Logic | Tarski 1936b / Etchemendy 1990 | Apr 6 & 8 |
| 4.2 Relevance Logic | Lewis 1982 / Burgess 2005 | Apr 13 |
| 4.3 Intuitionistic Logic | Lecture Notes Only | Apr 15 |
| 4.4 Informational View of Logic | Yalcin 2007 / Bledin 2014 | Apr 20 |
| 4.5 Field Against Truth Preservation | Field 2006 | Apr 22 |
| 4.6 Normativity of Logic | Harman 1986 / MacFarlane *ms.* / Field 2009 & forthcoming / Steinberger forthcoming | Apr 27 & 29 |

## Requirements

There are three requirements for taking this course.

The first requirement is to complete a series of four exercise sets (worth 40% of your final grade).

The second requirement is to write a ≤8 page double-spaced final paper on a topic of your choice (worth 25% of your final grade). I will meet with each of you individually to discuss your topic and the paper should be submitted on May 9.

The third requirement is to take a final in-class exam on May 9 between 2:00pm-5:00pm (worth 35% of your final grade).

Attendance is strongly encouraged.

## Group Work

Feel free to discuss the exercise sets in groups. Logic is best enjoyed with others! However, you should write up your solutions independently. Also, you should attempt the exercises on your own before meeting with your group. If you cannot do the exercises, you will not do well on the exam.

## Academic Integrity

Please do not cheat. This would be depressing. Cheating hurts the Johns Hopkins community by undermining academic and personal integrity, creating mistrust, and fostering unfair competition. Ethical violations include cribbing on exams, plagiarism, reuse of assignments, improper use of the internet and electronic devices, unauthorized collaboration, and alteration of graded assignments. Cheaters may receive a grade of F in the course and can face direr consequences in extreme cases.

Report any violations you witness. You may consult the associate dean of student affairs and/or the chairman of the Ethics Board beforehand. For more information, see the guide *Academic Ethics for Undergraduates* and/or the Ethics Board website:

`http://e-catalog.jhu.edu/undergrad-students/student-life-policies/#UAEB`

## Disability Accommodations

If you are a student with a disability or believe that you might have a disability that requires special accommodations, please contact Student Disability Services to obtain a letter from a specialist:

Garland 385
410.516.4720
`studentdisabilityservices@jhu.edu`

The terms of this letter will be honored.

**Enjoy the course!**

# Philosophical Logic

Johns Hopkins University, Spring 2015

| | |
|---|---|
| **Instructor** | Justin Bledin |
| | Assistant Professor |
| | Philosophy Department |
| | Gilman 206 |
| | `jbledin@jhu.edu` |
| **Office Hours** | Tu 3:15pm-4:15pm & by appt |

# Prerequisites

A solid understanding of classical first-order logic is essential. Students should also be familiar with basic methods of mathematical proof.

# Schedule

| | | |
|---|---|---|
| **Introduction** | | Jan 26 |
| **Logic Review** | | Jan 28, Feb 2 |
| **1. Quantifiers** | | |
| 1.1 Generalized Quantifiers | Barwise and Cooper 1981 | Feb 4 |
| 1.2 Substitutional Quantifiers | Linsky 1972 | Feb 9 |
| 1.3 Plural Quantifiers | Boolos 1981, 1984a&b | Feb 11 |

# Schedule

## 2. Modals

| | | |
|---|---|---|
| 2.1 Propositional Modal Logic | Lecture Notes Only | Feb 16 & 18 |
| 2.2 The Modal Zoo | Lecture Notes Only | Feb 23 & 25 |
| 2.3 Temporal Logic | Lecture Notes Only | Mar 2 |
| 2.4 Deontic Logic | Lecture Notes Only | Mar 4 |
| 2.5 Epistemic Logic | Lecture Notes Only | Mar 9 |

# Schedule

## 3. Truth Predicates

| | | |
|---|---|---|
| 3.1 Tarski | Tarski 1936a | Mar 23 |
| 3.2 Kripke | Kripke 1975 / Heck *ms.* | Mar 25 |
| 3.3 Contemporary Theories | Leitgeb 2007 / TBD | Mar 30, Apr 1 |

# Schedule

## 4. Logical Consequence

| | | |
|---|---|---|
| 4.1 Truth Preservation | Tarski 1936b / Etchemendy 1990 | Apr 6 & 8 |
| 4.2 Relevance Logic | Lewis 1982 / Burgess 2005 | Apr 13 |
| 4.3 Intuitionistic Logic | Lecture Notes Only | Apr 15 |
| 4.4 Informational View | Yalcin 2007 / Bledin 2014 | Apr 20 |
| 4.5 Field Against TP | Field 2006 | Apr 22 |
| 4.6 Normativity of Logic | Harman 1986 / MacFarlane *ms.* / Field 2009 & forthcoming / Steinberger forthcoming | Apr 27 & 29 |

# Requirements

There are three requirements for taking this course:

- Exercise sets (40% of final grade)
- $\leq$8 page final paper due on May 9 (25% of final grade)
- In-class final exam on May 9 (35% of final grade)

# Group Work

Feel free to discuss the exercise sets in groups. Logic is best enjoyed with others! However, you should write up your solutions independently. Also, you should attempt the exercises on your own before meeting with your group. If you cannot do the exercises, you will not do well on the exam.

# Academic Integrity

Please do not cheat. This would be depressing. Cheating hurts the Johns Hopkins community by undermining academic and personal integrity, creating mistrust, and fostering unfair competition. Ethical violations include cribbing on exams, plagiarism, reuse of assignments, improper use of the internet and electronic devices, unauthorized collaboration, and alteration of graded assignments. Cheaters may receive a grade of F in the course and can face direr consequences in extreme cases.

Report any violations you witness. You may consult the associate dean of student affairs and/or the chairman of the Ethics Board beforehand. For more information, see the guide *Academic Ethics for Undergraduates* and/or the Ethics Board website:

`http://e-catalog.jhu.edu/undergrad-students/student-life-policies/#UAEB`

# Disability Accommodations

If you are a student with a disability or believe that you might have a disability that requires special accommodations, please contact Student Disability Services to obtain a letter from a specialist:

Garland 385
410.516.4720
`studentdisabilityservices@jhu.edu`

The terms of this letter will be honored.

# Philosophical Logic

## 0.1 A Bit of Set Theory

Johns Hopkins University, Spring 2015

**Def 0.1.1.** A *set* $S$ is a collection of objects drawn from some domain.

**Examples:**

The set of past and present Hopkins students.

The set of Hopkins students enrolled in Philosophical Logic during the Spring 2015 term.

The set of natural numbers.

**Notation:**

$$\begin{aligned} A &= \{x : x \text{ is or was a Hopkins student}\} \\ &= \{\text{Michael Bloomberg}, \text{John Astin}, \text{Wes Craven}, \text{Wolf Blitzer}, \ldots\} \end{aligned}$$

$$\begin{aligned} B &= \{x : x \text{ is a Hopkins student enrolled in Phil Logic in Spring 2015}\} \\ &= \{\text{Nikola Andonovski}, \text{Caleb Baechtold}, \ldots\} \end{aligned}$$

$$\begin{aligned} \mathbb{N} &= \{x : x \text{ is a natural number}\} \\ &= \{0, 1, 2, 3, \ldots\} \end{aligned}$$

**More notation:**

$x \in S$ designates that object $x$ is a *member* or *element* of set $S$.

**Examples:**

Michael Bloomberg $\in A$.

Michael Bloomberg $\notin B$.

$3 \notin B$.

$3 \in \mathbb{N}$.

**Def 0.1.2.** $S$ is a *subset* of $T$ ($S \subseteq T$) just in case $\forall x(x \in S \supset x \in T)$.

That is, $S \subseteq T$ just in case each element of $S$ is an element of $T$.

**Examples:**

$A \nsubseteq B$.

$B \subseteq A$.

A set is uniquely determined by its members:

**Axiom of Extensionality.** $S = T$ just in case $\forall x(x \in S \equiv x \in T)$.

That is, $S = T$ just in case $S \subseteq T \wedge T \subseteq S$.

**Def 0.1.3.** $S$ is a *strict subset* of $T$ ($S \subset T$) just in case $S \subseteq T \wedge S \neq T$.

**Example:** $B \subset A$.

**Def 0.1.4.** The *union* of $S$ and $T$ ($S \cup T$) is the set $\{x : x \in S \vee x \in T\}$.

**Def 0.1.5.** The *intersection* of $S$ and $T$ ($S \cap T$) is the set $\{x : x \in S \wedge x \in T\}$.

**Def 0.1.6.** The *set-theoretic difference* of $S$ with $T$ ($S \setminus T$) is the set $\{x : x \in S \wedge x \notin T\}$.

**Examples:**

$\mathbb{N} = \{0, 1, 2, 3, ...\}$.

$\mathbb{N}^- = \{0, -1, -2, -3, ...\}$.

$\mathbb{N} \cup \mathbb{N}^- =$ ?

$\mathbb{N} \cap \mathbb{N}^- =$ ?

$\mathbb{N} \setminus \mathbb{N}^- =$ ?

$\mathbb{N}^- \setminus \mathbb{N} =$ ?

**Def 0.1.4.** The *union* of $S$ and $T$ $(S \cup T)$ is the set $\{x : x \in S \vee x \in T\}$.

**Def 0.1.5.** The *intersection* of $S$ and $T$ $(S \cap T)$ is the set $\{x : x \in S \wedge x \in T\}$.

**Def 0.1.6.** The *set-theoretic difference* of $S$ with $T$ $(S \setminus T)$ is the set $\{x : x \in S \wedge x \notin T\}$.

**Examples:**

$\mathbb{N} = \{0, 1, 2, 3, ...\}$.

$\mathbb{N}^- = \{0, -1, -2, -3, ...\}$.

$\mathbb{N} \cup \mathbb{N}^- = \{..., -3, -2, -1, 0, 1, 2, 3, ...\} = \mathbb{Z}$.

$\mathbb{N} \cap \mathbb{N}^- = \{0\}$.

$\mathbb{N} \setminus \mathbb{N}^- = \{1, 2, 3, ...\}$.

$\mathbb{N}^- \setminus \mathbb{N} = \{-1, -2, -3, ...\}$.

**Def 0.1.7.** The *empty set* $\emptyset$ is the set $\{x : x \neq x\}$.

**Def 0.1.8.** $S$ and $T$ are *disjoint* just in case $S \cap T = \emptyset$.

**Examples:**

?

**Def 0.1.7.** The *empty set* $\emptyset$ is the set $\{x : x \neq x\}$.

**Def 0.1.8.** $S$ and $T$ are *disjoint* just in case $S \cap T = \emptyset$.

**Examples:**

$\{x : x \text{ is a circle}\} \cap \{x : x \text{ is a square}\} = \emptyset$.

$\mathbb{N} \cap \{-5\} = \emptyset$.

**Lem 0.1.1.** $((S \subseteq U) \wedge (T \subseteq U)) \supset (S \cup T \subseteq U)$.

**Proof.** Suppose that $S \subseteq U$ and $T \subseteq U$.

Consider $a \in S \cup T$. By definition of union, $a \in S$ or $a \in T$.

If $a \in S$, then $a \in U$ since $S \subseteq U$.

If $a \in T$, then $a \in U$ since $T \subseteq U$.

But $a$ is arbitrary, so $\forall x(x \in S \cup T \supset x \in U)$.

That is, $S \cup T \subseteq U$.

**Lem 0.1.2.** $(S \subseteq T) \supset (T \setminus (T \setminus S) = S)$.

**Proof.** Suppose that $S \subseteq T$. It suffices to show that $T \setminus (T \setminus S) \subseteq S$ and $S \subseteq T \setminus (T \setminus S)$.

To show that $T \setminus (T \setminus S) \subseteq S$, consider $a \in T \setminus (T \setminus S)$.

By definition of difference, $a \in T$ and $a \notin T \setminus S$.

$a \notin T \setminus S$ only if either $a \notin T$ or $a \in S$, so $a \in S$.

But $a$ is arbitrary, so $\forall x(x \in T \setminus (T \setminus S) \supset x \in S)$.

To show that $S \subseteq T \setminus (T \setminus S)$, consider $b \in S$.

Since $S \subseteq T$, $b \in T$.

$b \in S \cap T$ only if $b \notin T \setminus S$, so $b \in T \setminus (T \setminus S)$.

But $b$ is arbitrary, so $\forall x(x \in S \supset x \in T \setminus (T \setminus S))$.

Recall that sets are determined by their members. Order does not matter.

$\{1,2,3\} = \{1,3,2\} = \{2,1,3\} = \{2,3,1\} = \{3,1,2\} = \{3,2,1\}$.

But we sometimes care about order.

**Def 0.1.9.** The *ordered pair* of $x$ and $y$ ($\langle x, y\rangle$) is the set $\{\{x\},\{x,y\}\}$.

Note that $\langle x, y\rangle = \{\{x\},\{x,y\}\} \neq \{\{y\},\{x,y\}\} = \langle y, x\rangle$.

$\langle x_1, ..., x_n\rangle = \langle\langle x_1, ..., x_{n-1}\rangle, x_n\rangle$.

**Example:** $\langle 1,2,3\rangle = \langle\langle 1,2\rangle,3\rangle = \{\{\langle 1,2\rangle\},\{\langle 1,2\rangle,3\}\}$.

Note that $\langle x, y \rangle$ is a set of sets. Here is another important kind of set of sets:

**Def 0.1.10.** The *power set* of $S$ ($2^S$) is the set $\{x : x \subseteq S\}$.

That is, the power set of $S$ is the collection of all subsets of $S$.

**Example:** $2^{\{1,2,3\}} = ?$

Note that $\langle x, y \rangle$ is a set of sets. Here is another important kind of set of sets:

**Def 0.1.10.** The *power set* of $S$ ($2^S$) is the set $\{x : x \subseteq S\}$.

That is, the power set of $S$ is the collection of all subsets of $S$.

**Example:** $2^{\{1,2,3\}} = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$.

Taking the power set pushes us up the set-theoretic hierarchy.

The following operations push us down the hierarchy:

**Def 0.1.11.** The *union* of $S$ ($\bigcup S$) is $\{x : \exists y(y \in S \wedge x \in y)\}$.

**Def 0.1.12.** The *intersection* of $S$ ($\bigcap S$) is $\{x : \forall y(y \in S \supset x \in y)\}$.

**Examples:**

$\bigcup\{\{1,2,3\},\{2,3,4\},\{3,4,5\}\} = ?$

$\bigcap\{\{1,2,3\},\{2,3,4\},\{3,4,5\}\} = ?$

Taking the power set pushes us up the set-theoretic hierarchy.

The following operations push us down the hierarchy:

**Def 0.1.11.** The *union* of $S$ ($\bigcup S$) is $\{x : \exists y(y \in S \wedge x \in y)\}$.

**Def 0.1.12.** The *intersection* of $S$ ($\bigcap S$) is $\{x : \forall y(y \in S \supset x \in y)\}$.

**Examples:**

$\bigcup\{\{1,2,3\},\{2,3,4\},\{3,4,5\}\} = \{1,2,3,4,5\}$.

$\bigcap\{\{1,2,3\},\{2,3,4\},\{3,4,5\}\} = \{3\}$.

$S \cup T = \bigcup\{S, T\}$ and $S \cap T = \bigcap\{S, T\}$.

**Def 0.1.13.** The *Cartesian product* of $S$ and $T$ ($S \times T$) is the set $\{\langle x, y\rangle : x \in S \wedge y \in T\}$.

More generally, the Cartesian product of $S_1, ..., S_n$ ($S_1 \times ... \times S_n$) is the set $\{\langle x_1, ..., x_n\rangle : x_1 \in S_1 \wedge ... \wedge x_n \in S_n\}$.

**Def 0.1.14.** $\mathcal{R}$ is a *unary relation* on $S$ just in case $\mathcal{R} \subseteq S$.

**Def 0.1.15.** $\mathcal{R}$ is a *binary relation* on $S$ just in case $\mathcal{R} \subseteq S \times S$.

**Def 0.1.16.** $\mathcal{R}$ is an *n-ary relation* on $S$ just in case $\mathcal{R}$ is a subset of the $n$-fold Cartesian product of $S$.

**Examples:**

$x > 5$ is a unary relation on $\mathbb{N}$: $\{6, 7, 8, ...\} \subseteq \mathbb{N}$.

$x \geq y$ is a binary relation on $\mathbb{N}$: $\{\langle 1, 1\rangle, \langle 3, 2\rangle, \langle 12, 4\rangle, ...\} \subseteq \mathbb{N} \times \mathbb{N}$.

$x + y = z$ is a tertiary relation on $\mathbb{N}$: $\{\langle 1, 2, 3\rangle, \langle 5, 2, 7\rangle, ...\} \subseteq \mathbb{N} \times \mathbb{N} \times \mathbb{N}$.

Consider $\mathcal{R} \subseteq S \times S$.

**Def 0.1.17.** $\mathcal{R}$ is *reflexive* just in case $\forall x(\mathcal{R}xx)$.

**Def 0.1.18.** $\mathcal{R}$ is *symmetric* just in case $\forall xy(\mathcal{R}xy \supset \mathcal{R}yx)$.

**Def 0.1.19.** $\mathcal{R}$ is *antisymmetric* just in case $\forall xy((\mathcal{R}xy \wedge x \neq y) \supset \neg\mathcal{R}yx)$.

**Def 0.1.20.** $\mathcal{R}$ is *connected* just in case $\forall xy(x \neq y \supset (\mathcal{R}xy \vee \mathcal{R}yx))$.

**Def 0.1.21.** $\mathcal{R}$ is *serial* just in case $\forall x\exists y(\mathcal{R}xy)$.

**Def 0.1.22.** $\mathcal{R}$ is *transitive* just in case $\forall xyz((\mathcal{R}xy \wedge \mathcal{R}yz) \supset \mathcal{R}xz)$.

**Examples:**

Which of these properties is satisfied by...

$\leq$ over $\mathbb{N}$?

$<$ over $\mathbb{N}$?

*taller than* over the set of all Hopkins students?

$\in$ over a set $S$?

Consider $\mathcal{R} \subseteq S \times S$.

**Def 0.1.17.** $\mathcal{R}$ is *reflexive* just in case $\forall x(\mathcal{R}xx)$.

**Def 0.1.18.** $\mathcal{R}$ is *symmetric* just in case $\forall xy(\mathcal{R}xy \supset \mathcal{R}yx)$.

**Def 0.1.19.** $\mathcal{R}$ is *antisymmetric* just in case $\forall xy((\mathcal{R}xy \wedge x \neq y) \supset \neg\mathcal{R}yx)$.

**Def 0.1.20.** $\mathcal{R}$ is *connected* just in case $\forall xy(x \neq y \supset (\mathcal{R}xy \vee \mathcal{R}yx))$.

**Def 0.1.21.** $\mathcal{R}$ is *serial* just in case $\forall x\exists y(\mathcal{R}xy)$.

**Def 0.1.22.** $\mathcal{R}$ is *transitive* just in case $\forall xyz((\mathcal{R}xy \wedge \mathcal{R}yz) \supset \mathcal{R}xz)$.

**Examples:**

Which of these properties is satisfied by...

$\leq$ over $\mathbb{N}$: reflexivity, antisymmetry, connectedness, seriality, transitivity.

$<$ over $\mathbb{N}$: antisymmetry, connectedness, seriality, transitivity.

*taller than* over the set of all Hopkins students: antisymmetry, transitivity.

$\in$ over a set $S$: antisymmetry.

**Def 0.1.23.** $\mathcal{R}$ is an *equivalence relation* on $S$ just in case $\mathcal{R}$ is reflexive, symmetric, and transitive.

An equivalence relation $\mathcal{R}$ partitions $S$ into a collection of disjoint *equivalence classes* of objects related by $\mathcal{R}$.

**Examples:** $=$ over $\mathbb{N}$, *same height as* over the set of Hopkins students.

**Def 0.1.24.** $\mathcal{R}$ is a *partial ordering* of $S$ just in case $\mathcal{R}$ is reflexive, antisymmetric, and transitive.

**Examples:** $\leq$ over $\mathbb{N}$, $\subseteq$ over $\{S_1, ...S_n\}$.

**Def 0.1.25.** $\mathcal{R}$ is a *total* or *linear ordering* of $S$ just in case $\mathcal{R}$ is a connected partial ordering.

**Example:** $\leq$ over $\mathbb{N}$.

Consider an $n+1$-ary relation $\mathcal{R}$ on $S$.

**Def 0.1.26.** The *domain* of $\mathcal{R}$ (*dom*($\mathcal{R}$)) is the set $\{x : \exists y(\langle x, y\rangle \in \mathcal{R})\}$.

Note that *dom*($\mathcal{R}$) is a set of $n$-tuples of elements from $S$.

**Def 0.1.27.** The *range* of $\mathcal{R}$ (*ran*($\mathcal{R}$)) is the set $\{x : \exists y(\langle y, x\rangle \in \mathcal{R})\}$.

Note that *ran*($\mathcal{R}$) is a set of elements from $S$.

**Def 0.1.28.** $f$ is an *n-ary function* just in case $f$ is an $n+1$-ary relation such that for each $x \in dom(f)$, there is exactly one $y \in ran(f)$ where $\langle x, y\rangle \in f$.

**Notation:**

$f : S \to T$ designates that $f$ is a function with $dom(f) = S$ and $ran(f) \subseteq T$.

**Examples:**

$+ : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, $\times : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$

*father of*: $H \to H$ (where $H$ is the set of past and present humans)

**Exercise 0.1.1**

Prove the following lemma:

**Lem.** $S \cap (T \cup U) = (S \cap T) \cup (S \cap U)$.

**Exercise 0.1.2**

Prove the following lemma:

**Lem.** $(S \subseteq T) \supset (S \cup (T \setminus S) = T)$.

**Exercise 0.1.3**

What are $\bigcup \emptyset$ and $\bigcap \emptyset$?

**Exercise 0.1.4**

Prove that a binary relation $\mathcal{R}$ that is symmetric, serial, and transitive is also reflexive.

# Philosophical Logic

## 0.2 Logic Review

Johns Hopkins University, Spring 2015

What is logic?

Largely pre-theoretic starting point:

Some arguments are *deductively good arguments* that we can appropriately make, by virtue of logical form, in both categorical deliberative contexts where the premises of the argument are known and hypothetical contexts where the premises are only supposed—at least, we can appropriately make these arguments in any deliberative context where doing so is not simply a waste or misuse of scarce cognitive resources.

For example, these arguments are deductively good:

(P1) Gomez is married to Morticia.
(C) Gomez or Fester is married to Morticia.

(P1) Every member of the Addams Family delights in the macabre.
(P2) Pugsley is an Addams.
(C) Pugsley delights in the macabre.

These arguments are not:

(P1) Gomez or Fester is married to Morticia.
(C) Gomez is married to Morticia.

(P1) No member of the Addams Family delights in the macabre.
(P2) Pugsley is an Addams.
(C) Pugsley delights in the macabre.

Once we start to talk about *logical validity*, we have started to theorize about what makes some arguments deductively good ones.

The orthodox *truth preservation view* is a cluster of widespread intuitions about the *informal* concept of logical validity:

- Core intuition: A logically valid argument with true premises has a true conclusion.
- Modal strengthening: It is *impossible* for each of the premises of a logically valid argument to be true and for the conclusion to be false.
- A logically valid argument preserves truth *by virtue of its logical form* and not due to the meaning of any non-logical symbols.

These intuitions all come together in the following definition:

**Def 0.2.1.** The argument from $\varphi_1, ..., \varphi_n$ to $\psi$ is *logically valid* just in case it is impossible for each of $\varphi_1, ..., \varphi_n$ to be true and for $\psi$ to be false by virtue of logical form.

This informal characterization of validity leaves room for disagreement:

- What sense of ‘impossible’ is relevant here? Is the modality alethic, metaphysical, or epistemic?
- What counts as ‘logical form’? Are higher-order quantifiers, say, logical constants?

Later on in our course, we will call the truth preservation view itself into question. But for the time being, we will maintain this orthodoxy.

In mathematical logic, this informal target notion is typically analyzed in terms of *truth in a model*. Open up just about any logic textbook and you will see something like this:

**Def 0.2.2.** The argument from $\varphi_1, ..., \varphi_n$ to $\psi$ is *logically valid* if and only if there is no model $\mathcal{M}$ for formal language $\mathcal{L}$ such that the translations of $\varphi_1, ..., \varphi_n$ into $\mathcal{L}$ are all true in $\mathcal{M}$ but the translation of $\psi$ into $\mathcal{L}$ is false in $\mathcal{M}$

where $\mathcal{L}$ purportedly makes the logical form of $\varphi_1, ..., \varphi_n, \psi$ explicit, and a model $\mathcal{M}$ for $\mathcal{L}$ is, roughly, something that provides enough information to determine the extensions of all well-formed sentences $S_{\mathcal{L}}$ of this formal language.

In this Tarskian model-theoretic paradigm, the necessity modal in the informal truth preservation characterization of validity is cashed out by quantifying over all models.

By varying our formal languages and models, logicians and philosophers have generated a large family of formal notions of logical validity.

They agree on which arguments count as valid on this or that formal characterization.

They disagree on which of the formal notions extensionally coincide with the *genuine* informal notion of logical validity.

# Review of Sentential Logic

**Def 0.2.3.** A *language of sentential logic* $\mathcal{L}_{SENT}$ has this syntax:

$p \mid \perp \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi)$

Note that this language is slightly redundant since $\perp$ can be regarded as an abbreviation for $(A \wedge \neg A)$ and $\vee/\wedge$ can be defined in terms of $\neg$ and $\wedge/\vee$ (on the standard semantics).

The material conditional $\supset$ and biconditional $\equiv$ can also be defined in the usual fashion.

$At_{\mathcal{L}_{SENT}} = \{A, B, ...\}$ is the set of atoms in $\mathcal{L}_{SENT}$.
$S_{\mathcal{L}_{SENT}}$ is the set of sentences in $\mathcal{L}_{SENT}$.

# Review of Sentential Logic

**Def 0.2.4.** A *model* $\mathcal{M} = \langle \mathcal{V} \rangle$ for $\mathcal{L}_{SENT}$ consists of a valuation function $\mathcal{V} : At_{\mathcal{L}_{SENT}} \rightarrow \{T, F\}$ mapping each sentence letter $p \in At_{\mathcal{L}_{SENT}}$ to a truth value.

Each row of a truth table corresponds to an equivalence class of models where the range of the valuation function agrees on all of the atomic sentence letters that appear in the reference columns of the truth table.

# Review of Sentential Logic

**Def 0.2.5.** A recursive specification of *truth in a model* lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT}} \to \{T, F\}$ for $\mathcal{L}_{SENT}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{SENT}}$ to a truth value:

$$\begin{array}{rcl}
[\![p]\!]_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p) = T \\
[\![\bot]\!]_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}} = T \text{ and } [\![\psi]\!]_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}} = T \text{ or } [\![\psi]\!]_{\mathcal{M}} = T
\end{array}$$

This compositional semantics is often presented using truth tables.

The sentential constants $\neg$, $\wedge$, and $\vee$ are *truth functional*: the truth value of a complex sentence generated with these constants depends only on the truth value of the constituent subsentence(s).

# Review of Sentential Logic

**Example:**

Consider a model $\mathcal{M} = \langle \mathcal{V} \rangle$ where $\mathcal{V}(A) = T$, $\mathcal{V}(B) = T$, and $\mathcal{V}(C) = F$.

$[\![(A \vee B) \wedge (B \vee C)]\!]_{\mathcal{M}} = ?$

# Review of Sentential Logic

**Example:**

Consider a model $\mathcal{M} = \langle \mathcal{V} \rangle$ where $\mathcal{V}(A) = T$, $\mathcal{V}(B) = T$, and $\mathcal{V}(C) = F$.

$[\![(A \vee B) \wedge (B \vee C)]\!]_{\mathcal{M}} = \mathrm{T}$. Why?

By the semantic clause for sentence letters, $[\![A]\!]_{\mathcal{M}} = T$, $[\![B]\!]_{\mathcal{M}} = T$, and $[\![C]\!]_{\mathcal{M}} = F$.

By the semantic clause for $\vee$, $[\![A \vee B]\!]_{\mathcal{M}} = T$ and $[\![B \vee C]\!]_{\mathcal{M}} = T$.

By the semantic clause for $\wedge$, $[\![(A \vee B) \wedge (B \vee C)]\!]_{\mathcal{M}} = T$.

# Review of Sentential Logic

**Def 0.2.6.** The argument from $\varphi_1, ..., \varphi_n$ to $\psi$ is *tautologically valid*, $\{\varphi_1, ..., \varphi_n\} \models_{SENT} \psi$, just in case there is no model $\mathcal{M}$ for $\mathcal{L}_{SENT}$ where $[\![\varphi_1]\!]_{\mathcal{M}} = ... = [\![\varphi_n]\!]_{\mathcal{M}} = T$ but $[\![\psi]\!]_{\mathcal{M}} = F$.

The expression 'tautologically valid' is used here because nobody thinks that $\mathcal{L}_{SENT}$ fully captures logical form. That is, nobody thinks that $\models_{SENT}$ explicates our informal target notion of logical validity.

A constellation of related notions are definable in terms of tautological validity:

**Def 0.2.7.** The sentence $\varphi$ is a *tautology* just in case $\emptyset \models_{SENT} \varphi$.

**Def 0.2.8.** The sentences $\varphi$ and $\psi$ are *tautologically equivalent* just in case both $\{\varphi\} \models_{SENT} \psi$ and $\{\psi\} \models_{SENT} \varphi$.

And so forth.

# Review of Sentential Logic

(P1) Gomez is married to Morticia.
(C) Gomez or Fester is married to Morticia.

(P1) $G$
(C) $G \vee F$

This argument is tautologically valid.

For any model $\mathcal{M}$, $[\![G]\!]_{\mathcal{M}} = T$ implies $[\![G \vee F]\!]_{\mathcal{M}} = T$ by the semantic clause for $\vee$.

Thus, $\{G\} \models_{SENT} G \vee F$.

# Review of Sentential Logic

(P1) Gomez or Fester is married to Morticia.
(C) Gomez is married to Morticia.

(P1) $G \vee F$
(C) $G$

This argument is tautologically invalid.

Countermodel: $\mathcal{V}(G) = F$ and $\mathcal{V}(F) = T$.

Thus, $\{G \vee F\} \not\models_{SENT} G$.

# Review of Sentential Logic

(P1) Every member of the Addams Family delights in the macabre.
(P2) Pugsley is an Addams.
(C) Pugsley delights in the macabre.

(P1) A
(P2) B
(C) C

Although the English argument from (P1) and (P2) to (C) is logically valid, it comes out tautologically invalid.

We need a more fine-grained language than $\mathcal{L}_{SENT}$.

# Review of First-Order Logic

A *language of first-order logic* $\mathcal{L}_{FOL}$ is built up from:

*variables* $\{x_1, x_2, x_3, ...\}$

*constants* $\{c_1, c_2, c_3, ...\}$

*n-ary function symbols* $\{f_1^n, f_2^n, f_3^n, ...\}$ for various $n \in \mathbb{N}$

*n-ary predicate symbols* $\{P_1^n, P_2^n, P_3^n, ...\}$ for various $n \in \mathbb{N}$

**Def 0.2.9.** The *terms* $\{t_1, t_2, ...\}$ of $\mathcal{L}_{FOL}$ are the variables, constants, and $n$-ary functions applied to $n$ terms. The *closed terms* $\{t_1', t_2', ...\}$ of $\mathcal{L}_{FOL}$ are terms without variables.

**Def 0.2.10.** $\mathcal{L}_{FOL}$ has the following syntax for formulae:

$P_i^n t_{j_1} ... t_{j_n} \mid t_i = t_j \mid \perp \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \forall x_i \varphi \mid \exists x_i \varphi$

# Review of First-Order Logic

**Running Example:**

Consider the first-order Addams Family language with these symbols:

$x, y, z$ (for simplicity, assume that there are only three variables)

$g$, $m$, $f$, $p$ (names for Gomez, Morticia, Fester, and Pugsley respectively)

$M$ (1-ary *is married* predicate)

$D$ (1-ary *delights in the macabre* predicate)

The terms are the variables $x, y, z$ and the constants $g$, $m$, $f$, $p$.

The closed terms are just the constants.

Well-formed formulae of this language include $Mg$, $Dx$, $y = f$, $Mm \vee Mf$, $\forall z Dz$, $\exists x \neg Mx$, and so forth.

# Review of First-Order Logic

**Def 0.2.11.** The *scope* of a quantifier is the formula directly following it.

**Example:** The scope of the existential quantifier $\exists y$ in $\forall x \exists y \forall x (Mx \vee y = z)$ is $\forall x (Mx \vee y = z)$.

**Def 0.2.12.** The quantifiers $\forall x_i$ and $\exists x_i$ *bind* all occurrences of $x_i$ in their scope that are not already bound by some other quantifier. A variable is either *bound* by some quantifier or *free*.

**Example:** In the embedded formula $(Mx \vee y = z)$ in the previous example, $x$ is bound by the rightmost universal quantifier, $y$ is bound by the existential quantifier, and $z$ is free.

**Def 0.2.13.** A formula containing free variables is *open*. A formula sans free variables is a *closed* formula or *sentence*.

$At_{\mathcal{L}_{FOL}}$ is the set of atomic sentences in $\mathcal{L}_{FOL}$.
$S_{\mathcal{L}_{FOL}}$ is the set of sentences in $\mathcal{L}_{FOL}$.

# Review of First-Order Logic

**Def 0.2.14.** A *model* $\mathcal{M} = \langle \mathcal{D}, \mathcal{I} \rangle$ for $\mathcal{L}_{FOL}$ consists of a nonempty domain of objects $\mathcal{D}$ and an interpretation function $\mathcal{I}$ mapping each constant $c_i$ to a single object in $\mathcal{D}$, each $n$-ary function symbol $f_i^n$ to a function from ordered $n$-tuples of objects in $\mathcal{D}$ to a single object in $\mathcal{D}$, and each $n$-ary predicate $P_i^n$ to a set of ordered $n$-tuples of objects in $\mathcal{D}$.

**Running Example:**

$\mathcal{D} = \{$Gomez, Morticia, Fester, Pugsley$\}$.

$\mathcal{I}(g) =$ Gomez, $\mathcal{I}(m) =$ Morticia, $\mathcal{I}(f) =$ Fester, $\mathcal{I}(p) =$ Pugsley.

$\mathcal{I}(M) = \{$Gomez, Morticia$\}$, $\mathcal{I}(D) = \{$Gomez, Morticia, Fester, Pugsley$\}$.

# Review of First-Order Logic

**Def 0.2.15.** A *variable assignment* $g : \{x_1, x_2, ...\} \rightarrow \mathcal{D}$ is a function mapping each variable to a member of $\mathcal{D}$.

**Running Example:**

Here are three possible variable assignments:

$g_1(x) =$ Gomez, $g_1(y) =$ Fester, $g_1(z) =$ Gomez.

$g_2(x) =$ Pugsley, $g_2(y) =$ Morticia, $g_2(z) =$ Fester.

$g_3(x) =$ Morticia, $g_3(y) =$ Morticia, $g_3(z) =$ Morticia.

# Review of First-Order Logic

**Def 0.2.16.** The *extension* $[\![t_i]\!]^g_{\mathcal{M}} \in \mathcal{D}$ of each term (open and closed) of $\mathcal{L}_{FOL}$ in $\mathcal{M}$ under variable assignment $g$ is determined recursively as follows:

$[\![x_i]\!]^g_{\mathcal{M}} = g(x_i)$
$[\![c_i]\!]^g_{\mathcal{M}} = \mathcal{I}(c_i)$
$[\![f^n_i(t_{j_1}, ..., t_{j_n})]\!]^g_{\mathcal{M}} = \mathcal{I}(f^n_i)([\![t_{j_1}]\!]^g_{\mathcal{M}}, ..., [\![t_{j_n}]\!]^g_{\mathcal{M}})$

**Running Example:**

$[\![x]\!]^{g_2}_{\mathcal{M}} =$ Pugsley, $[\![y]\!]^{g_2}_{\mathcal{M}} =$ Morticia, $[\![z]\!]^{g_2}_{\mathcal{M}} =$ Fester.

$[\![g]\!]^{g_2}_{\mathcal{M}} =$ Gomez, $[\![m]\!]^{g_2}_{\mathcal{M}} =$ Morticia, $[\![f]\!]^{g_2}_{\mathcal{M}} =$ Fester, $[\![p]\!]^{g_2}_{\mathcal{M}} =$ Pugsley.

# Review of First-Order Logic

To define truth in a model for sentences, we first define an intermediate assignment-relative notion of truth for arbitrary open and closed formulae.

**Def 0.2.17.** *Truth in a model under a variable assignment* is determined by the following recursive clauses:

$$
\begin{array}{rcl}
[\![P_i^n t_{j_1} ... t_{j_n}]\!]_{\mathcal{M}}^g = T & \text{iff} & \langle [\![t_{j_1}]\!]_{\mathcal{M}}^g, ..., [\![t_{j_n}]\!]_{\mathcal{M}}^g \rangle \in \mathcal{I}(P_i^n) \\
[\![t_i = t_j]\!]_{\mathcal{M}}^g = T & \text{iff} & [\![t_i]\!]_{\mathcal{M}}^g = [\![t_j]\!]_{\mathcal{M}}^g \\
[\![\bot]\!]_{\mathcal{M}}^g = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]_{\mathcal{M}}^g = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}}^g = F \\
[\![\varphi \wedge \psi]\!]_{\mathcal{M}}^g = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}}^g = T \text{ and } [\![\psi]\!]_{\mathcal{M}}^g = T \\
[\![\varphi \vee \psi]\!]_{\mathcal{M}}^g = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}}^g = T \text{ or } [\![\psi]\!]_{\mathcal{M}}^g = T \\
[\![\forall x_i \varphi]\!]_{\mathcal{M}}^g = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}}^{g[x_i \to d]} = T \text{ for every } d \in \mathcal{D} \\
[\![\exists x_i \varphi]\!]_{\mathcal{M}}^g = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}}^{g[x_i \to d]} = T \text{ for some } d \in \mathcal{D}
\end{array}
$$

where $g[x_i \to d](x_i) = d$ and $g(x_j) = g[x_i \to d](x_j)$ for $j \neq i$; that is, $g[x_i \to d]$ is a new variable assignment that is exactly like $g$ except that it maps $x_i$ to $d \in \mathcal{D}$.

# Review of First-Order Logic

**Running Example:**

$[\![ Mg ]\!]^{g_2}_{\mathcal{M}} = ?$

# Review of First-Order Logic

**Running Example:**

$[\![Mg]\!]^{g_2}_{\mathcal{M}} = T$. Why?

$[\![g]\!]^{g_2}_{\mathcal{M}} =$ Gomez.

$\mathcal{I}(M) = \{$Gomez, Morticia$\}$.

$[\![g]\!]^{g_2}_{\mathcal{M}} \in \mathcal{I}(M)$.

So by the semantic clause for predicate symbols, $[\![Mg]\!]^{g_2}_{\mathcal{M}} = T$.

# Review of First-Order Logic

**Running Example:**

$[\![ y = f ]\!]^{g_2}_{\mathcal{M}} = ?$

# Review of First-Order Logic

**Running Example:**

$[\![ y = f ]\!]^{g_2}_{\mathcal{M}} = F$. Why?

$[\![ y ]\!]^{g_2}_{\mathcal{M}} =$ Morticia.

$[\![ f ]\!]^{g_2}_{\mathcal{M}} =$ Fester.

$[\![ y ]\!]^{g_2}_{\mathcal{M}} \neq [\![ f ]\!]^{g_2}_{\mathcal{M}}$.

So by the semantic clause for $=$, $[\![ y = f ]\!]^{g_2}_{\mathcal{M}} = F$.

# Review of First-Order Logic

**Running Example:**

$[\![\forall z D z]\!]^{g_2}_{\mathcal{M}} = ?$

# Review of First-Order Logic

**Running Example:**

$[\![\forall z D z]\!]^{g_2}_{\mathcal{M}} = T$. Why?

# Review of First-Order Logic

$g_2[z \to \text{Gomez}](x) = \text{Pugsley}$.
$g_2[z \to \text{Gomez}](y) = \text{Morticia}$.
$g_2[z \to \text{Gomez}](z) = \text{Gomez}$.

$[\![z]\!]_{\mathcal{M}}^{g_2[z \to \text{Gomez}]} = \text{Gomez}$.

$\mathcal{I}(D) = \{\text{Gomez, Morticia, Fester, Pugsley}\}$.

$[\![z]\!]_{\mathcal{M}}^{g_2[z \to \text{Gomez}]} \in \mathcal{I}(D)$.

So by the semantic clause for predicate symbols, $[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Gomez}]} = T$.

# Review of First-Order Logic

$g_2[z \to \text{Morticia}](x) = \text{Pugsley}.$
$g_2[z \to \text{Morticia}](y) = \text{Morticia}.$
$g_2[z \to \text{Morticia}](z) = \text{Morticia}.$

$[\![z]\!]_{\mathcal{M}}^{g_2[z \to \text{Morticia}]} = \text{Morticia}.$

$\mathcal{I}(D) = \{\text{Gomez, Morticia, Fester, Pugsley}\}.$

$[\![z]\!]_{\mathcal{M}}^{g_2[z \to \text{Morticia}]} \in \mathcal{I}(D).$

So by the semantic clause for predicate symbols, $[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Morticia}]} = T.$

# Review of First-Order Logic

$g_2[z \to \text{Fester}](x) = \text{Pugsley}.$
$g_2[z \to \text{Fester}](y) = \text{Morticia}.$
$g_2[z \to \text{Fester}](z) = \text{Fester}.$

$[\![z]\!]_{\mathcal{M}}^{g_2[z \to \text{Fester}]} = \text{Fester}.$

$\mathcal{I}(D) = \{\text{Gomez, Morticia, Fester, Pugsley}\}.$

$[\![z]\!]_{\mathcal{M}}^{g_2[z \to \text{Fester}]} \in \mathcal{I}(D).$

So by the semantic clause for predicate symbols, $[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Fester}]} = T$.

# Review of First-Order Logic

$g_2[z \rightarrow \text{Pugsley}](x) = \text{Pugsley}.$
$g_2[z \rightarrow \text{Pugsley}](y) = \text{Morticia}.$
$g_2[z \rightarrow \text{Pugsley}](z) = \text{Pugsley}.$

$[\![z]\!]_{\mathcal{M}}^{g_2[z \rightarrow \text{Pugsley}]} = \text{Pugsley}.$

$\mathcal{I}(D) = \{\text{Gomez, Morticia, Fester, Pugsley}\}.$

$[\![z]\!]_{\mathcal{M}}^{g_2[z \rightarrow \text{Pugsley}]} \in \mathcal{I}(D).$

So by the semantic clause for predicate symbols, $[\![Dz]\!]_{\mathcal{M}}^{g_2[z \rightarrow \text{Pugsley}]} = T.$

# Review of First-Order Logic

$[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Gomez}]} = T$.

$[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Morticia}]} = T$.

$[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Fester}]} = T$.

$[\![Dz]\!]_{\mathcal{M}}^{g_2[z \to \text{Pugsley}]} = T$.

So by the semantic clause for $\forall$, $[\![\forall z Dz]\!]_{\mathcal{M}}^{g_2} = T$.

# Review of First-Order Logic

*Truth in a model* for each sentence $\varphi \in S_{\mathcal{L}_{FOL}}$ is now easily defined:

**Def 0.2.18.** $[\![\varphi]\!]_{\mathcal{M}} = T$ just in case $[\![\varphi]\!]^{g}_{\mathcal{M}} = T$ for any $g$.

The truth of a sentence $\varphi$ in $\mathcal{M}$ is variable assignment independent.

Recall that we can formally define validity in terms of truth in a model:

**Def 0.2.19.** The argument from $\varphi_1, ..., \varphi_n$ to $\psi$ is *logically valid*, $\{\varphi_1, ..., \varphi_n\} \models_{FOL} \psi$, just in case there is no model $\mathcal{M}$ for $\mathcal{L}_{FOL}$ where $[\![\varphi_1]\!]_{\mathcal{M}} = ... = [\![\varphi_n]\!]_{\mathcal{M}} = T$ and $[\![\psi]\!]_{\mathcal{M}} = F$.

The hope, again, is that this formal notion of logical validity extensionally captures the informal target notion.

# Review of First-Order Logic

(P1) Every member of the Addams Family delights in the macabre.
(P2) Pugsley is an Addams.
(C) Pugsley delights in the macabre.

(P1) $\forall x(Ax \supset Dx)$
(P2) $Ap$
(C) $Dp$

This argument is logically valid.

For any model $\mathcal{M}$, $[\![\forall x(Ax \supset Dx)]\!]_{\mathcal{M}} = T$ only if $\mathcal{I}(p) \notin \mathcal{I}(A)$ or $\mathcal{I}(p) \in \mathcal{I}(D)$.

$[\![Ap]\!]_{\mathcal{M}} = T$ only if $\mathcal{I}(p) \in \mathcal{I}(A)$.

So $\mathcal{I}(p) \in \mathcal{I}(D)$. That is, $[\![Dp]\!]_{\mathcal{M}} = T$.

Thus, $\{\forall x(Ax \supset Dx), Ap\} \models_{FOL} Dp$.

# Review of First-Order Logic

(P1) No member of the Addams Family delights in the macabre.
(P2) Pugsley is an Addams.
(C) Pugsley delights in the macabre.

(P1) $\neg\exists x(Ax \wedge Dx)$
(P2) $Ap$
(C) $Dp$

This argument is logically invalid.

Countermodel: $\mathcal{D} = \{\text{Gomez, Pugsley}\}$, $\mathcal{I}(p) = \text{Pugsley}$, $\mathcal{I}(A) = \{\text{Gomez, Pugsley}\}$, $\mathcal{I}(D) = \emptyset$.

Thus, $\{\neg\exists x(Ax \wedge Dx), Ap\} \not\models Dp$.

# Review of First-Order Logic

Logicians have also developed proof systems for establishing validity.

In the Fitch-style system $\mathcal{F}$ in Barwise and Etchemendy [1999], we can prove $Dp$ from $\{\forall x(Ax \supset Dx), Ap\}$ as follows:

| | | |
|---|---|---|
| 1 | $\forall x(Ax \supset Dx)$ | Premise |
| 2 | $Ap$ | Premise |
| 3 | $Ap \supset Dp$ | $\forall$ Elim:1 |
| 4 | $Dp$ | $\supset$ Elim: 3,2 |

**Def 0.2.20.** $\{\varphi_1, ..., \varphi_n\} \vdash_{\mathcal{F}} \psi$ just in case there is a proof in $\mathcal{F}$ beginning with premises $\varphi_1, ..., \varphi_n$ that concludes with $\psi$.

As shown, $\{\forall x(Ax \supset Dx), Ap\} \vdash_{\mathcal{F}} Dp$.

# Review of First-Order Logic

Barwise and Etchemendy prove that $\mathcal{F}$ is *sound* and *complete* with respect to $\models_{FOL}$.

**Thm 0.2.1 (Soundness Theorem for $\mathcal{F}$).** $\{\varphi_1, ..., \varphi_n\} \vdash_{\mathcal{F}} \psi$ only if $\{\varphi_1, ..., \varphi_n\} \models_{FOL} \psi$.

**Thm 0.2.2 (Completeness Theorem for $\mathcal{F}$).** $\{\varphi_1, ..., \varphi_n\} \models_{FOL} \psi$ only if $\{\varphi_1, ..., \varphi_n\} \vdash_{\mathcal{F}} \psi$.

By Soundness, $\{\forall x(Ax \supset Dx), Ap\} \models_{FOL} Dp$.

# Review of First-Order Logic

First-order logic clearly improves on sentential logic.

Intuitively valid arguments involving quantificational structure that are tautologically invalid come out logically valid in FOL.

But is the formal relation $\models_{FOL}$ extensionally adequate?

Many think not. As we will see, $\mathcal{L}_{FOL}$ also has expressive limitations.

**Exercise 0.2.1**

Consider a model $\mathcal{M} = \langle \mathcal{V} \rangle$ for $\mathcal{L}_{SENT}$ where $\mathcal{V}(A) = F$, $\mathcal{V}(B) = T$, and $\mathcal{V}(C) = F$.

Are the following sentences true or false in $\mathcal{M}$? Justify your answers by appealing to the compositional semantics in Def 0.2.5.

$(\neg A \wedge C) \vee (\neg B \wedge C)$

$(A \vee \bot) \vee (B \vee \bot)$

$(C \vee \neg C) \wedge (C \wedge \neg C)$

**Exercise 0.2.2**

Which (if any) of the following sentences are tautologically equivalent?

$A \vee B$

$\neg(A \wedge \neg B)$

$A \vee \neg A$

$\neg(\neg A \wedge \neg B)$

$A \vee \neg B$

$\neg A \vee B$

$\neg(\neg A \vee (A \wedge \neg A)) \vee \neg A$

**Exercise 0.2.3**

Consider the following model $\mathcal{M}$ for the Addams Family language:

$\mathcal{D} = \{\text{Gomez, Morticia, Fester, Pugsley}\}$.

$\mathcal{I}(g) = \text{Gomez}$, $\mathcal{I}(m) = \text{Morticia}$, $\mathcal{I}(f) = \text{Fester}$, $\mathcal{I}(p) = \text{Pugsley}$.

$\mathcal{I}(M) = \{\text{Gomez, Morticia}\}$, $\mathcal{I}(D) = \{\text{Fester, Pugsley}\}$.

Are the following sentences true or false in $\mathcal{M}$? Justify your answers by appealing to the semantics in Defs 0.2.17 and 0.2.18.

$M(g) \vee D(g)$

$\exists x(Mx \wedge Dx)$

$\forall x \exists y(Mx \vee Dy)$

**Exercise 0.2.4**

Circle the free variables in each of the following sentences and draw an arrow from each quantifier to the variable(s) that it binds.

$\neg\exists x(Ax \wedge \forall y(By \vee Cxy))$

$\forall x(Ay \wedge \forall x(By \wedge Cx) \wedge \exists y(Ey \wedge Fx))$

**Exercise 0.2.5**

Provide a formula in $\mathcal{L}_{FOL}$ involving $\forall$ and $=$ that is true in every model whose domain $\mathcal{D}$ has only one element but false in every model whose domain $\mathcal{D}$ has multiple elements.

Provide a formula in $\mathcal{L}_{FOL}$ *sans* quantifiers and identity that is true in every model whose domain $\mathcal{D}$ has only one element but false in some model whose domain $\mathcal{D}$ has multiple elements.

# Philosophical Logic

## 1.1 Generalized Quantifiers

Johns Hopkins University, Spring 2015

Barwise and Cooper [1981] think that the first-order quantifiers $\forall$ and $\exists$ are inadequate to treat the quantificational structures found in natural languages in two respects.

First, the syntactic structure of quantified sentences in natural language is completely different from the syntactic structure of quantified sentences in $\mathcal{L}_{FOL}$.

Second, many natural language quantifiers cannot be captured in $\mathcal{L}_{FOL}$.

In $\mathcal{L}_{FOL}$, a quantifier $\mathrm{Q}x_i$ can be attached to a single formula to generate another formula. If the quantifier binds all of the free occurrences of $x_i$ in its scope and no occurrence of any other variable is free, then the result is a sentence.

(1) Every farmer sweats.

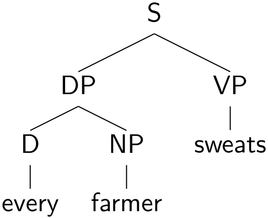
**Translation:** $\forall x(Fx \supset Sx)$.

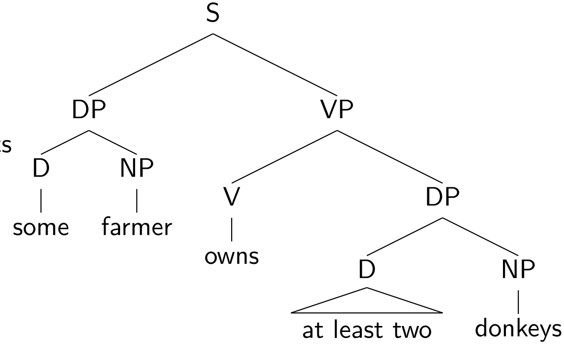
(2) Some farmer owns at least two donkeys.

**Translation:** $\exists x(Fx \wedge \exists y \exists z(y \neq z \wedge Dy \wedge Dz \wedge Oxy \wedge Oxz))$.

Natural language suggests a different approach.

Grammatically, expressions like ‘every’, ‘some’, ‘most’, ‘five’, ‘at least two’, ‘finitely many’ are *determiners* that modify a noun phrase. The resulting *determiner phrase* can combine with a verb phrase to form a sentence:



This phrase structure suggests that a quantifier $\mathrm{Q}x_i$ should operate on *two* formulae.

(1) Every farmer sweats.

**Translation:** $\mathit{every}_x(Fx, Sx)$.

(2) Some farmer owns at least two donkeys.

**Translation:** $\mathit{some}_x(Fx, \mathit{at\text{-}least\text{-}two}_y(Dy, Oxy))$.

(3) Everyone likes some donkey.

**Translation:** $\mathit{every}_x(x = x, \mathit{some}_y(Dy, Lxy))$.

(4) Most farmers are poor.

**Translation:** $\mathit{most}_x(Fx, Px)$.

Aside:

I am adopting the modern view that expressions like ‘some’ and ‘most’ are *binary* quantifiers.

Barwise and Cooper themselves take determiner (noun) phrases like ‘every farmer’ and ‘at least two donkeys’ to be quantifiers. For them, quantifiers are *unary*.

N. B. This unary view implies that the meaning of quantifiers is not fixed across models. Since the meaning of ‘every farmer’ involves the meaning of ‘farmer’ and this can vary from model to model, the meaning of the full quantifier is not model-independent.

N. B. Barwise and Cooper think that even the meaning of determiners like ‘most’ is model-dependent (so they distinguish between logical and non-logical determiners).

Binary quantifiers can be interpreted by giving the kind of compositional semantics introduced in the logic review.

**Def 1.1.1.** *Truth in a model under a variable assignment* for formulae involving binary quantifiers is determined with recursive clauses like these:

$$
\begin{array}{rcl}
[\![ \mathit{every}_{x_i}(\varphi,\psi) ]\!]^{g}_{\mathcal{M}} = T & \text{iff} & \text{for every } d \text{ s.t. } [\![\varphi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T, \\
& & [\![\psi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T \\
[\![ \mathit{some}_{x_i}(\varphi,\psi) ]\!]^{g}_{\mathcal{M}} = T & \text{iff} & \text{for some } d \text{ s.t. } [\![\varphi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T, \\
& & [\![\psi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T \\
[\![ \mathit{at\text{-}least\text{-}two}_{x_i}(\varphi,\psi) ]\!]^{g}_{\mathcal{M}} = T & \text{iff} & \text{for at least two } d \text{ s.t. } [\![\varphi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T, \\
& & [\![\psi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T \\
[\![ \mathit{most}_{x_i}(\varphi,\psi) ]\!]^{g}_{\mathcal{M}} = T & \text{iff} & \text{for most } d \text{ s.t. } [\![\varphi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T, \\
& & [\![\psi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T
\end{array}
$$

Barwise and Cooper themselves provide an alternative semantics on which determiner phrases (their 'quantifiers') denote sets of subsets of $\mathcal{D}$.

(1) Every farmer sweats.

$[\![Every]\!]_{\mathcal{M}}(x) = \{y : y \subseteq \mathcal{D} \wedge x \subseteq y\}$ for each $x \subseteq \mathcal{D}$.

$[\![Every\ farmer]\!]_{\mathcal{M}} = [\![Every]\!]_{\mathcal{M}}(\mathcal{I}(F)) = \{y : y \subseteq \mathcal{D} \wedge \mathcal{I}(F) \subseteq y\}$.

$[\![Every\ farmer\ sweats]\!]_{\mathcal{M}} = T$ iff $\mathcal{I}(S) \in [\![Every\ farmer]\!]_{\mathcal{M}}$.

(4) Most farmers are poor.

$[\![Most]\!]_{\mathcal{M}}(x) = \{y : y \subseteq \mathcal{D} \wedge |x \cap y| > \frac{|x|}{2}\}$ for each $x \subseteq \mathcal{D}$.

($|x|$ is the size or *cardinality* of $x$.)

$[\![Most\ farmers]\!]_{\mathcal{M}} = [\![Most]\!]_{\mathcal{M}}(\mathcal{I}(F)) = \{y : y \subseteq \mathcal{D} \wedge |\mathcal{I}(F) \cap y| > \frac{|\mathcal{I}(F)|}{2}\}$.

$[\![Most\ farmers\ are\ poor]\!]_{\mathcal{M}} = T$ iff $\mathcal{I}(P) \in [\![Most\ farmers]\!]_{\mathcal{M}}$.

Going binary might strike you as unnecessary since many determiners can be captured with $\forall$ and $\exists$:

$some_x(Px, Qx) \equiv \exists x(Px \wedge Qx)$.

$every_x(Px, Qx) \equiv \forall x(Px \supset Qx)$.

$at\text{-}least\text{-}two_x(Px, Qx) \equiv \exists x \exists y(x \neq y \wedge Px \wedge Py \wedge Qx \wedge Qy)$.

But this brings us to Barwise and Cooper's second point: many natural language quantified sentences are nonfirstorderizable.

**Examples:**

(5) More than half of Cupid's arrows hit the target.

(6) Most of Cupid's arrows hit the target.

(7) Finitely many of Cupid's arrows hit the target.

(8) Infinitely many of Cupid's arrows hit the target.

Barwise and Cooper prove that no sentence in $S_{\mathcal{L}_{FOL}}$ is equivalent to $most_x(Px, Qx)$.

In fact, they prove something stronger: even if we help ourselves to the unary quantifier $M$ where

$$[\![Mx_i\varphi]\!]^g_{\mathcal{M}} = T \quad \text{iff} \quad [\![\varphi]\!]^{g[x_i \to d]}_{\mathcal{M}} = T \text{ for most } d \in \mathcal{D}$$

$most_x(Px, Qx)$ is still not definable in the expanded language.

Nowadays, it is commonly thought that 'Most' is *essentially binary*. But note that Barwise and Cooper show only that $most_x(Px, Qx)$ cannot be captured by this particular unary quantifier $M$.

Why should logicians care about the nonfirstorderizability of natural language quantified sentences involving determiners like 'most'?

Well, consider the following argument:

(P1) Most farmers are poor.
(P2) All poor people live in small houses.
(C) Most farmers live in small houses.

This argument is deductively good. So, if 'most' is a logical constant, arguments like this reveal that $\mathcal{L}_{FOL}$ is inadequate.

Barwise and Cooper end their classic paper with this passage on the relationship between logic and natural language semantics:

"The traditional logical notions of validity and inference are a part of linguistics—a conclusion not likely to comfort many logicians or linguists. The phenomenal success of first-order logic with mathematics has obscured, indeed, nearly severed, its tied with its origins in language. Except for tense and modal logic, research in model theory in the past twenty five years has taken its problems almost entirely from pure mathematics, becoming ever more specialized and remote from language. Even the work in generalized quantifiers mentioned in the introduction is devoted almost exclusively to mathematical quantifiers, going out of its way to avoid mentioning possible applications to natural language. This same success of first-order logic within mathematics also fostered the mistaken idea...that the 'laws of logic' are autonomous, perhaps part of mathematics, but not a property of language and language use.

It is here that Montague made his biggest contribution. To most logicians (like the first author) trained in model-theoretic semantics, natural language was an anathema, impossibly vague and incoherent. To us, *the* revolutionary idea in Montague's paper PTQ (and earlier papers) is the claim that natural language is not impossibly incoherent, as his teacher Tarski had led us to believe, but that large portions of its semantics can be treated by combining known tools from logic, tools like functions of finite type, the $\lambda$-calculus, generalized quantifiers, tense and modal logic, and all the rest." (p. 203-4)

**Exercise 1.1.1**

Translate the following sentences using binary quantifiers:

At least one hundred farmers are married to teachers.

Most people like most dogs.

Someone is Barwise.

**Exercise 1.1.2**

Consider the following model $\mathcal{M}$ for the Addams Family language:

$\mathcal{D} = \{$Gomez, Morticia, Fester, Pugsley$\}$.

$\mathcal{I}(g) =$ Gomez, $\mathcal{I}(m) =$ Morticia, $\mathcal{I}(f) =$ Fester, $\mathcal{I}(p) =$ Pugsley.

$\mathcal{I}(M) = \{$Gomez, Morticia$\}$, $\mathcal{I}(D) = \{$Gomez, Morticia, Fester, Pugsley$\}$.

Are the following sentences true or false in $\mathcal{M}$? Justify your answers by appealing to the semantics in Defs 1.1.1 and 0.2.18.

$\mathit{all}_x(Dx, Mx)$

$\mathit{at\text{-}least\text{-}two}_x(x = x, Dx)$

**Exercise 1.1.3**

On the Barwise and Cooper semantics for quantifiers, what are $[\![\mathit{Some}]\!]_{\mathcal{M}}(x)$ and $[\![\text{At least two}]\!]_{\mathcal{M}}(x)$?

# Philosophical Logic

## 1.2 Substitutional Quantifiers

Johns Hopkins University, Spring 2015

Linsky [1972] distinguishes between two kinds of quantification on the logical scene:

$\forall x_i P x_i$ and $\exists x_i P x_i$ involve *objectual quantification*. To evaluate these sentences for truth, we evaluate the embedded open formula $Px_i$ for truth under different assignments of objects in the domain $\mathcal{D}$ to $x_i$.

$\Pi x_i P x_i$ and $\Sigma x_i P x_i$ involve *substitutional quantification*. To evaluate these sentences for truth, we evaluate the embedded open formula $Px_i$ for truth after replacing $x_i$ with closed terms.

Here, the substitution class for $\Pi$ and $\Sigma$ is the class of closed singular terms. But nothing prevents us from using other substitution classes such as the class whose sole member is the left-hand parenthesis (this example is from Lesniewski via Quine).

Substitutional quantification is often credited to Ruth Barcan Marcus.

There are different ways to formulate a semantics for a first-order language $\mathcal{L}$ that includes substitutional but not objectual quantifiers.

First, we might use first-order models of the form $\mathcal{M} = \langle \mathcal{D}, \mathcal{I} \rangle$.

**Def 1.2.1.** *Truth in a model* is determined for each $\varphi \in S_{\mathcal{L}}$ as follows:

$$
\begin{aligned}
[\![P^n_i t'_{j_1} ... t'_{j_n}]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad \langle [\![t'_{j_1}]\!]_{\mathcal{M}}, ..., [\![t'_{j_n}]\!]_{\mathcal{M}} \rangle \in \mathcal{I}(P^n_i) \\
[\![t'_i = t'_j]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad [\![t'_i]\!]_{\mathcal{M}} = [\![t'_j]\!]_{\mathcal{M}} \\
[\![\bot]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad 0 = 1 \\
[\![\neg\varphi]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad [\![\varphi]\!]_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad [\![\varphi]\!]_{\mathcal{M}} = T \text{ and } [\![\psi]\!]_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad [\![\varphi]\!]_{\mathcal{M}} = T \text{ or } [\![\psi]\!]_{\mathcal{M}} = T \\
[\![\Pi x_i \varphi]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad [\![\varphi[x_i \to t']]\!]_{\mathcal{M}} = T \text{ for every closed term } t' \text{ in } \mathcal{L} \\
[\![\Sigma x_i \varphi]\!]_{\mathcal{M}} = T \quad &\text{iff} \quad [\![\varphi[x_i \to t']]\!]_{\mathcal{M}} = T \text{ for some closed term } t' \text{ in } \mathcal{L}
\end{aligned}
$$

where $\varphi[x_i \to c]$ is the sentence obtained by substituting each free occurrence of $x_i$ in $\varphi$ with the closed term $t'$.

Note that this semantics does not involve variable assignments. To give a combined semantics for a language with both objectual and substitutional quantifiers, we would still need variable assignments to handle $\forall$ and $\exists$.

Second, we might use sentential models of the form $\mathcal{M} = \langle \mathcal{V} \rangle$ where $\mathcal{V} : At_{\mathcal{L}} \to \{T, F\}$ maps each atomic sentence $\varphi \in At_{\mathcal{L}}$ to a truth value.

**Def 1.2.2.** *Truth in a model* is determined for each $\varphi \in S_{\mathcal{L}}$ as follows:

$$\begin{array}{rcl}
[\![P^n_i t'_{j_1} ... t'_{j_n}]\!]_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(P^n_i t'_{j_1} ... t'_{j_n}) = T \\
[\![t'_i = t'_j]\!]_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(t'_i = t'_j) = T \\
[\![\bot]\!]_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}} = T \text{ and } [\![\psi]\!]_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]_{\mathcal{M}} = T \text{ or } [\![\psi]\!]_{\mathcal{M}} = T \\
[\![\Pi x_i \varphi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi[x_i \to t']]\!]_{\mathcal{M}} = T \text{ for every } t' \text{ in } \mathcal{L} \\
[\![\Sigma x_i \varphi]\!]_{\mathcal{M}} = T & \text{iff} & [\![\varphi[x_i \to t']]\!]_{\mathcal{M}} = T \text{ for some } t' \text{ in } \mathcal{L}
\end{array}$$

N. B. Linsky presents this kind of semantics in the appendix to his paper.

**Example:**

Consider the first-order language with these symbols:

$g$, $m$ (names for Gomez and Morticia)

$M$ (1-ary *is married* predicate)

Here is a first-order model for this language:

$\mathcal{D} = \{$Gomez, Morticia, Fester, Pugsley$\}$.

$\mathcal{I}(g) =$ Gomez, $\mathcal{I}(m) =$ Morticia.

$\mathcal{I}(M) = \{$Gomez, Morticia$\}$.

Here is the corresponding sentential model for this language:

$\mathcal{V}(Mg) = \mathcal{V}(Mm) = T$.

$\mathcal{V}(g = g) = \mathcal{V}(m = m) = T$.

$\mathcal{V}(g = m) = \mathcal{V}(m = g) = F$.

$[\![\Pi x M x]\!]_{\mathcal{M}} = ?$ (where $\mathcal{M}$ is either of the above models)

$[\![\Pi x Mx]\!]_{\mathcal{M}} = T$. Why?

$[\![Mg]\!]_{\mathcal{M}} = T$.

$[\![Mm]\!]_{\mathcal{M}} = T$.

Since $g$ and $m$ are the only closed terms, $[\![\Pi x Mx]\!]_{\mathcal{M}} = T$.

Note that $[\![\forall x Mx]\!]_{\mathcal{M}} = F$. Moral: When some objects in the domain are unnamed, the truth values of substitutional and objectual quantified sentences can come apart.

Why introduce substitutional quantifiers into the language?

**First Motivation:** Nonexistence claims

(P1) Pegasus does not exist.
(C) There is something that does not exist.

This argument is puzzling since it seems to be valid yet have a true premise and a false conclusion.

**Response:** Accept that (C) is true by allowing the domain $\mathcal{D}$ to include nonexistent objects.

**Response:** Deny that (P1) is true. Since 'Pegasus' does not refer, (P1) is not even truth-apt.

**Response:** Deny that the argument is logically valid. The logical form of (P1) is actually $\neg\exists x Px$ for some predicate $P$ (Quine: "pegasizes") and this does not entail $\exists x(\neg(x \text{ exists}))$.

**Response:** Accept that (C) is true by appealing to substitutional quantifiers instead of nonexistent objects in $\mathcal{D}$:

(P1) $\neg(p \text{ exists})$
(C) $\Sigma x \neg(x \text{ exists})$

**Second Motivation:** Quantifying into attitude constructions

(P1) James thinks Mark Twain is a great writer.
(C) There is someone who James thinks is a great writer.

But what if James does not think Samuel Clemens is a great writer? How should we regard the conclusion (C)? If the quantifier here is objectual, then there seems to be an object in $\mathcal{D}$ that James thinks is a great writer *and* James does not think is a great writer.

**Response:** Take (C) to be $\Sigma x$(James thinks $x$ is a great writer).

**Third Motivation:** Quantifying into quotes

$\Pi$ and $\Sigma$ can bind variables occurring inside of quotation marks.

**Example:** (MacFarlane)

$\Sigma x$(Bush said '$x$ is the worst threat to our security' when $x$ was dead).

Let us close with a seemingly powerful application of substitutional quantifiers.

In a language containing both objectual quantifiers and substitutional quantifiers that can quantify into quotation, we can give this simple Tarskian definition of *truth*:

$\forall x(x \text{ is true} \equiv \Sigma p(x = `p\text{'} \wedge p))$.

But there are two serious worries with this definition.

First, as Linsky outlines at the end of his paper, unrestricted quantification into quotation leads to paradox.

Second, if substitutional quantified sentences are about the *truth* of sentences generated through substitution, then defining truth using substitutional quantifiers is circular.

There is plenty more to say about all of this.

**Exercise 1.2.1**

Translate the following sentence using a substitutional quantifier whose substitution class is sentences:

If the Pope utters a sentence, then it is true.

**Exercise 1.2.2**

Provide a first-order language involving at least the 1-ary predicate $P$ and a model $\mathcal{M}$ for this language such that $[\![\exists x P x]\!]_{\mathcal{M}} = T$ but $[\![\Sigma x P x]\!]_{\mathcal{M}} = F$.

**Exercise 1.2.3**

How would you respond to the puzzle of nonexistence claims? Argue for your preferred approach in a couple of paragraphs.

# Philosophical Logic

## 1.3 Plural Quantifiers

Johns Hopkins University, Spring 2015

Boolos [1984a] questions whether FOL provides an adequate theory of quantification and cross reference in English.

First, recall that sentences involving numerical quantifiers (and sentences synonymous with such sentences) present a challenge:

(1) Most Democrats are left-of-center.

(2) More Democrats are left-of-center than right-of-center.

(3) For every Democrat, there is a Republican.

N. B. Boolos [1981] provides a short elegant proof of the nonfirstorderizability of (3) using the Compactness and Löwenheim-Skolem Theorems.

There are many more problematic sentences besides.

Here is the famous *Geach-Kaplan* sentence:

(4) Some critics admire only one another.

This sentence can be formalized using second-order quantifiers:

$$\exists X(\exists x Xx \wedge \forall x(Xx \supset Cx) \wedge \forall x \forall y((Xx \wedge Axy) \supset (x \neq y \wedge Xy)))$$

However, this sentence is nonfirstorderizable.

N. B. Boolos provides an elegant proof due to Kaplan that appeals to Gödel's First Incompleteness Theorem.

Here are some other examples discussed by Boolos:

(5) There are some gunslingers each of whom has shot the right foot of at least one of the others.

$$\exists X(\exists x Xx \wedge \forall x(Xx \supset Gx) \wedge \forall x(Xx \supset \exists y(Xy \wedge y \neq x \wedge Sxy)))$$

(6) Some of Fiorecchio's men entered the building unaccompanied by anyone else.

$$\exists X(\exists x Xx \wedge \forall x(Xx \supset Fx) \wedge \forall x(Xx \supset Ex) \wedge \forall x \forall y((Xx \wedge Axy) \supset Xy))$$

Since quantified sentences like these can enter into deductive arguments, they motivate moving beyond FOL.

There is a fine line between the firstorderizable and nonfirstorderizable.

(7) There is a horse that is faster than Zev and also faster than the sire of any horse that is slower than it.

$$\exists x(Fxz \wedge \forall y(Fxy \supset Fxs(y)))$$

(8) There are some horses that are all faster than Zev and also faster than the sire of any horse that is slower than all of them.

$$\exists X(\exists x Xx \wedge \forall x(Xx \supset Fxz) \wedge \forall y(\forall x(Xx \supset Fxy) \supset \forall x(Xx \supset Fxs(y))))$$

Boolos [1984b] attributes the nonfirstorderizability of (8) to the plural forms in this sentence.

The above nonfirstorderizable quantified sentences (along with sentences whose most natural formalization involves second-order quantifiers) suggest that we should add second-order quantifiers to our language.

But how should we interpret $\exists XXa$?

A couple of options:

- There is a concept under whose extension $a$ falls.
- There is a set of which $a$ is a member.

$\exists X$ quantifies over concepts or sets.

But then why not just restate $\exists XXa$ in first-order logic where the domain $\mathcal{D}$ includes concepts and/or sets?

Quine's famous objection: second-order logic is set theory in sheep's clothing.

Boolos resists such interpretations for two reasons.

First, we sometimes want to say things like 'There are some sets of which every set that is not a member of itself is one'. However, such sentences are problematic if they entail the existence of a set of all sets.

Second, when we say 'There are some trucks of which every truck is one', we seem to be talking about trucks, not concepts or sets.

Boolos' alternative: Second-order quantifiers are *plural quantifiers* that range over some of the objects of first-order quantification all at once.

Boolos concludes his paper with this passage:

“The lesson to be drawn from the foregoing reflections on plurals and second-order logic is that neither the use of plurals nor the employment of second-order logic commits us to the existence of extra items beyond those to which we are already committed. We need not construe second-order quantifiers as ranging over anything other than the objects over which our first-order quantifiers range, and, in the absence of other reasons for thinking so, we need not think that there are collections of (say) Cheerios, in addition to the Cheerios.

Ontological commitment is carried by our *first*-order quantifiers; a second-order quantifier needn't be taken to be a kind of first-order quantifier in disguise, having items of a special kind, collections, in its range. It is not as though there were two sorts of things in the world, individuals, and collections of them, which our first- and second- order variables, respectively, range over and which our singular and plural forms, respectively, denote. There are, rather, two (at least) different ways of referring to the same things, among which there may well be many, many collections.

Leibniz once said, 'Whatever is, is one.'

Russell replied, 'And whatever are, are many.'" (p. 449)

**Exercise 1.3.1**

Translate the following formulae using second-order quantifiers:

If there are some numbers of which the successor of any one of them is also one, then if zero is one of them, $x$ is one of them.

If there are some persons of whom each parent of any one of them is also one, then if each parent of $y$ is one of them, $x$ is one of them; and someone is a parent of $y$.

**Exercise 1.3.2**

Are any of the following sentences firstorderizable? If so, provide their first-order equivalents.

There are some romance novels of which no one person has read them all.

Some yentas gossip only with one another.

Some yentas gossip with one another and no one else.

# Philosophical Logic

## 2.1 Sentential Modal Logic

Johns Hopkins University, Spring 2015

A *modal operator* qualifies a statement. These come in a variety of different flavors:

**epistemic**
- According to what McNulty knows, it is possible that...
- McNulty knows that...

**doxastic**
- According to what McNulty believes, it is possible that...
- McNulty believes that...

**temporal**
- Going forward, it will sometime be that...
- Going forward, it will always be that...

**deontic**
- It is permissible that...
- It is obligatory that...

And so forth. One of our primary goals in this unit is to develop formal techniques for systematically determining whether an argument involving modal operators is logically valid.

**Def 2.1.1.** A *language of sentential modal logic* $\mathcal{L}_{SENT_\Diamond}$ has the following syntax:

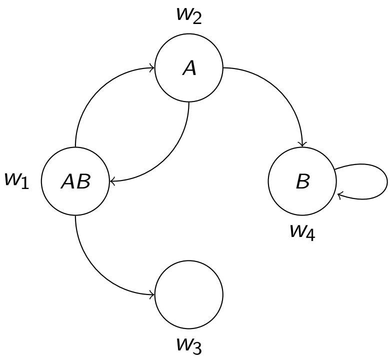
$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \Diamond\varphi \mid \Box\varphi$

More redundancy: $\Diamond$ and $\Box$ are interdefinable using $\neg$.

$At_{\mathcal{L}_{SENT_\Diamond}} = \{A, B, ...\}$ is the set of atoms in $\mathcal{L}_{SENT_\Diamond}$.
$S_{\mathcal{L}_{SENT_\Diamond}}$ is the set of sentences in $\mathcal{L}_{SENT_\Diamond}$.

We will later work with a *polymodal* language with multiple modalities.

**Def 2.1.2.** A *Kripke model* $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ for $\mathcal{L}_{SENT_\Diamond}$ consists of a nonempty set $\mathcal{W}$ of possible worlds, a binary accessibility relation $\mathcal{R} \subseteq \mathcal{W} \times \mathcal{W}$ between possible worlds, and a valuation function $\mathcal{V} : At_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \to \{T, F\}$ mapping each sentence letter $p \in At_{\mathcal{L}_{SENT_\Diamond}}$ and world $w \in \mathcal{W}$ to a truth value.




$\mathcal{W} = \{w_1, w_2, w_3, w_4\}$
$\mathcal{R} = \{\langle w_1, w_2\rangle, \langle w_1, w_3\rangle, \langle w_2, w_1\rangle, \langle w_2, w_4\rangle, \langle w_4, w_4\rangle\}$
$\mathcal{V}(A, w_1) = \mathcal{V}(A, w_2) = T$, $\mathcal{V}(A, w_3) = \mathcal{V}(A, w_4) = F$
$\mathcal{V}(B, w_1) = \mathcal{V}(B, w_4) = T$, $\mathcal{V}(B, w_2) = \mathcal{V}(B, w_3) = F$

**Def 2.1.3.** A recursive specification of *truth in a model* lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \to \{T, F\}$ for $\mathcal{L}_{SENT_\Diamond}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ and world $w \in \mathcal{W}$ to a truth value:

$$
\begin{array}{rcl}
[\![p]\!]^w_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p, w) = T \\
[\![\bot]\!]^w_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\Diamond\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & \exists v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = T) \\
[\![\Box\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & \forall v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)
\end{array}
$$

Note that $\Diamond$ is a kind of restricted existential quantifier and $\Box$ is a kind of restricted universal quantifier.

**Example:**

Given the previous model,

$[\![B]\!]^{w_2}_{\mathcal{M}} = ?$

$[\![\Diamond(A \wedge B)]\!]^{w_2}_{\mathcal{M}} = ?$

$[\![\Box B]\!]^{w_2}_{\mathcal{M}} = ?$

**Def 2.1.3.** A recursive specification of *truth in a model* lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \mapsto \{T, F\}$ for $\mathcal{L}_{SENT_\Diamond}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ and world $w \in \mathcal{W}$ to a truth value:

$$
\begin{array}{rcl}
[\![p]\!]^w_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p, w) = T \\
[\![\bot]\!]^w_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\Diamond\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & \exists v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = T) \\
[\![\Box\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & \forall v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)
\end{array}
$$

Note that $\Diamond$ is a kind of restricted existential quantifier and $\Box$ is a kind of restricted universal quantifier.

**Example:**

Given the previous model,

$[\![B]\!]^{w_2}_{\mathcal{M}} = F$.

$[\![\Diamond(A \wedge B)]\!]^{w_2}_{\mathcal{M}} = T$.

$[\![\Box B]\!]^{w_2}_{\mathcal{M}} = T$.

We can again define a family of formal logical notions in terms of *truth in a pointed model*:

**Def 2.1.4.** The argument from premises $\varphi_1, ..., \varphi_n$ to conclusion $\psi$ is *logically valid*, $\{\varphi_1, ..., \varphi_n\} \models_{SENT_\diamond} \psi$, just in case there is no pointed model $\mathcal{M}, w$ such that $[\![\varphi_1]\!]^w_\mathcal{M} = ... = [\![\varphi_n]\!]^w_\mathcal{M} = T$ but $[\![\psi]\!]^w_\mathcal{M} = F$.

**Def 2.1.5.** The sentence $\varphi$ is a *logical validity*, $\models_{SENT_\diamond} \varphi$, just in case there is no pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_\mathcal{M} = F$.

**Def 2.1.6.** The sentences $\varphi_1, ..., \varphi_n$ are *logically consistent* just in case there is a pointed model $\mathcal{M}, w$ such that $[\![\varphi_1]\!]^w_\mathcal{M} = ... = [\![\varphi_n]\!]^w_\mathcal{M} = T$.

And so forth.

(P1) It must be the case that Avon is in the tower.
(P2) It must be the case that Stringer is in the tower.
(C) It must be the case that both Avon and Stringer are in the tower.

(P1) $\Box A$
(P2) $\Box S$
(C) $\Box(A \wedge S)$

This argument is logically valid.

Suppose that $[\![\Box A]\!]^w_{\mathcal{M}} = T$ and $[\![\Box S]\!]^w_{\mathcal{M}} = T$.

By the semantic clause for $\Box$, $\forall v \in \{v : w\mathcal{R}v\}([\![A]\!]^v_{\mathcal{M}} = [\![S]\!]^v_{\mathcal{M}} = T)$.

By the semantic clause for $\wedge$, $\forall v \in \{v : w\mathcal{R}v\}([\![A \wedge S]\!]^v_{\mathcal{M}} = T)$.

By the semantic clause for $\Box$, $[\![\Box(A \wedge S)]\!]^w_{\mathcal{M}} = T$.

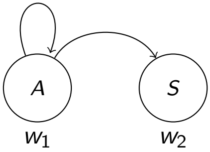
Thus, $\{\Box A, \Box S\} \models_{SENT_\Diamond} \Box(A \wedge S)$.

(P1) It might be the case that Avon is in the tower.
(P2) It might be the case that Stringer is in the tower.
(C) It might be the case that both Avon and Stringer are in the tower.

(P1) $\Diamond A$
(P2) $\Diamond S$
(C) $\Diamond(A \wedge S)$

This argument is logically invalid.

Countermodel:




Thus, $\{\Diamond A, \Diamond S\} \not\models_{SENT_\Diamond} \Diamond(A \wedge S)$.

Now that we have a syntax and semantics for the basic sentential modal language $\mathcal{L}_{SENT_\Diamond}$, let us ask: when are two Kripke models effectively the same with respect to $\mathcal{L}_{SENT_\Diamond}$?

**Def 2.1.7.** Pointed models $\mathcal{M}, w$ and $\mathcal{N}, v$ are *modally equivalent*, $\mathcal{M}, w \leftrightsquigarrow \mathcal{N}, v$, provided that for every $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$, $[\![\varphi]\!]^w_{\mathcal{M}} = [\![\varphi]\!]^v_{\mathcal{N}}$.

That is, $\mathcal{L}_{SENT_\Diamond}$ cannot tell modally equivalent pointed models apart.

A sufficient condition for modal equivalence is that a special kind of relation holds between pointed models:

**Def 2.1.8.** Given $\mathcal{M} = \langle \mathcal{W}^{\mathcal{M}}, \mathcal{R}^{\mathcal{M}}, \mathcal{V}^{\mathcal{M}} \rangle$ and $\mathcal{N} = \langle \mathcal{W}^{\mathcal{N}}, \mathcal{R}^{\mathcal{N}}, \mathcal{V}^{\mathcal{N}} \rangle$, a *bisimulation* between $\mathcal{M}, w$ and $\mathcal{N}, v$ is a binary relation $\mathcal{Z} \subseteq \mathcal{W}^{\mathcal{M}} \times \mathcal{W}^{\mathcal{N}}$ such that $w\mathcal{Z}v$ and for all worlds $x \in \mathcal{W}^{\mathcal{M}}$ and $y \in \mathcal{W}^{\mathcal{N}}$, if $x\mathcal{Z}y$ then:

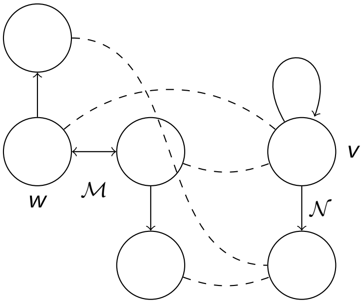
(atomic harmony) For all $p \in At_{\mathcal{L}_{SENT_\Diamond}}$, $\mathcal{V}^{\mathcal{M}}(p, x) = \mathcal{V}^{\mathcal{N}}(p, y)$.

(zig) If $x\mathcal{R}^{\mathcal{M}}z$, then there exists $z' \in \mathcal{W}^{\mathcal{N}}$ such that $y\mathcal{R}^{\mathcal{N}}z'$ and $z\mathcal{Z}z'$.

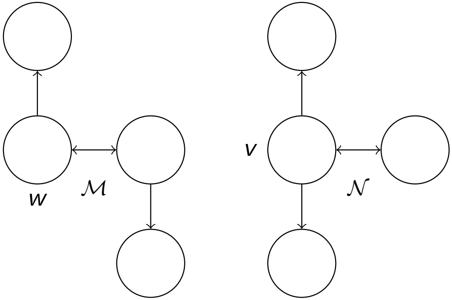
(zag) If $y\mathcal{R}^{\mathcal{N}}z'$, then there exists $z \in \mathcal{W}^{\mathcal{M}}$ such that $x\mathcal{R}^{\mathcal{M}}z$ and $z'\mathcal{Z}z$.

We say that $\mathcal{M}, w$ and $\mathcal{N}, v$ are *bisimilar*: $\mathcal{M}, w \leftrightarrows \mathcal{N}, v$.

These pointed models are bisimilar:



These pointed models are not:



**Lem 2.1.1 (Invariance Lemma).** $\mathcal{M}, w \leftrightarroweq \mathcal{N}, v$ only if $\mathcal{M}, w \leftrightsquigarrow \mathcal{N}, v$.

The proof is a straightforward induction on the complexity of sentences in $S_{\mathcal{L}_{SENT_\Diamond}}$.

To establish that $\mathcal{M}, w \not\leftrightarroweq \mathcal{N}, v$, it suffices to find some sentence $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ such that $[\![\varphi]\!]^w_{\mathcal{M}} \neq [\![\varphi]\!]^v_{\mathcal{N}}$.

**Example:**

What about the previous non-bisimilar pointed models?

$[\![\Box(\Box\bot \vee \Diamond\Box\bot)]\!]^w_{\mathcal{M}} = T$ but $[\![\Box(\Box\bot \vee \Diamond\Box\bot)]\!]^v_{\mathcal{N}} = F$.

The converse of the Invariance Lemma also holds when the pointed models are *finite*—that is, when $|\mathcal{W}^{\mathcal{M}}| = m_1$ and $|\mathcal{W}^{\mathcal{N}}| = m_2$ for $m_1, m_2 \in \mathbb{N}$.

**Lem 2.1.2.** For finite pointed models $\mathcal{M}, w$ and $\mathcal{N}, v$, $\mathcal{M}, w \leftrightsquigarrow \mathcal{N}, v$ only if $\mathcal{M}, w \leftrightarroweq \mathcal{N}, v$.

For the proof, let $x \mathcal{Z} y$ when $\mathcal{M}, x \leftrightsquigarrow \mathcal{N}, y$.

Since $\mathcal{L}_{SENT_\Diamond}$ cannot tell bisimilar pointed models apart, it is sometimes useful to thin a pointed model by finding a bisimilar pointed submodel. Let us consider two ways to do this.

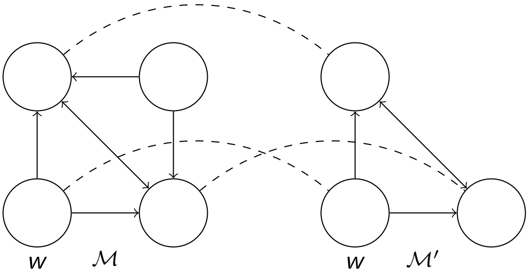
**Def 2.1.9.** Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$, $\mathcal{M}' = \langle \mathcal{W}', \mathcal{R}', \mathcal{V}' \rangle$ is a *submodel* of $\mathcal{M}$ just in case

$\mathcal{W}' \subseteq \mathcal{W}$
$\mathcal{R}'$ is the restriction of $\mathcal{R}$ to $\mathcal{W}'$
$\mathcal{V}'$ is the restriction of $\mathcal{V}$ to $\mathcal{W}'$

**Def 2.1.10.** Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ and $w \in \mathcal{W}$, the *submodel of $\mathcal{M}$ generated from $w$* is the submodel $\mathcal{M}'$ where $\mathcal{W}'$ is the set of worlds reachable from $w$ in 0 or more steps along $\mathcal{R}$.

**Lem 2.1.3.** If $\mathcal{M}'$ is the submodel of $\mathcal{M}$ generated from $w$, $\mathcal{M}, w \leftrightarrows \mathcal{M}', w$.

The identity relation is a bisimulation.



w
$\mathcal{M}$
w
$\mathcal{M}'$

For the second kind of submodel, consider the set $\mathcal{Z}_{\mathcal{M}}$ of bisimulations between a model $\mathcal{M}$ and any of its worlds and this same model $\mathcal{M}$ and any of its worlds (the *autobisimulations* of $\mathcal{M}$).

$\mathcal{Z}_{\mathcal{M}} \neq \emptyset$ since it includes the identity relation.

Consider the union $\cup\mathcal{Z}_{\mathcal{M}}$ of all the bisimulations in $\mathcal{Z}_{\mathcal{M}}$.

$\cup\mathcal{Z}_{\mathcal{M}}$ is both a bisimulation and an equivalence relation on $\mathcal{W}$.

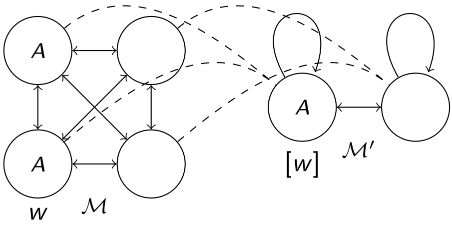
Let $[w] = \{v \in \mathcal{W} : w \cup \mathcal{Z}_{\mathcal{M}} v\}$.

**Def 2.1.11.** Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$, the *bisimulation contraction* of $\mathcal{M}$ is the model $\mathcal{M}' = \langle \mathcal{W}', \mathcal{R}', \mathcal{V}' \rangle$ where

$\mathcal{W}' = \{[w] : w \in \mathcal{W}\}$
$\mathcal{R}' = \{\langle [w], [v] \rangle$ : there is $x \in [w]$ and $y \in [v]$ such that $x\mathcal{R}y\}$
$\mathcal{V}'(p, [w]) = \mathcal{V}(p, w)$

**Lem 2.1.4.** If $\mathcal{M}'$ is the bisimulation contraction of $\mathcal{M}$, $\mathcal{M}, w \leftrightarrows \mathcal{M}', [w]$.

The relation sending $x \in \mathcal{W}$ to $[x] \in \mathcal{W}'$ is a bisimulation.




In this example, the submodel of $\mathcal{M}$ generated from $w$ is just $\mathcal{M}$ itself. So the two kinds of bisimilar submodels under consideration differ.

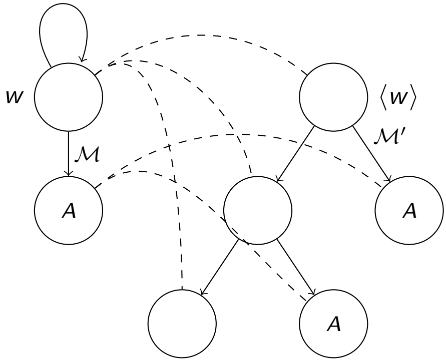
It is also sometimes useful to transform a pointed model into a larger bisimilar pointed model.

**Def 2.1.12.** Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ that is generated from $w$, the *tree unraveling of* $\mathcal{M}$ *around* $w$ is the model $\mathcal{M}' = \langle \mathcal{W}', \mathcal{R}', \mathcal{V}' \rangle$ where

$$\mathcal{W}' = \{\langle w, ..., w_n \rangle : w, ..., w_n \in \mathcal{W} \text{ and } w\mathcal{R}w_2...w_{n-1}\mathcal{R}w_n\}$$
$$\mathcal{R}' = \{\langle\langle x_1, ..., x_n \rangle, \langle y_1, ..., y_m \rangle\rangle : \langle y_1, ..., y_m \rangle = \langle x_1, ..., x_n, z \rangle \text{ for } z \in \mathcal{W}\}$$
$$\mathcal{V}'(p, \langle w, ..., w_n \rangle) = \mathcal{V}(p, w_n)$$

**Lem 2.1.5.** If $\mathcal{M}'$ is the tree unraveling of $\mathcal{M}$ around $w$, $\mathcal{M}, w \leftrightarrows \mathcal{M}', \langle w \rangle$.

The relation sending $x \in \mathcal{W}$ to all worlds $\langle w, ..., w_n \rangle \in \mathcal{W}'$ where $w_n = x$ is a bisimulation.



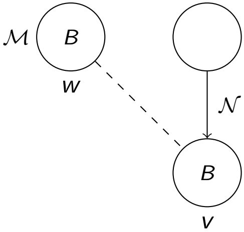
$w$
$\mathcal{M}$
$A$
$\langle w \rangle$
$\mathcal{M}'$
$A$
$A$

A nice application of the Invariance Lemma is that we can use it to prove that certain operators are undefinable in $\mathcal{L}_{SENT_\Diamond}$.

**Lem 2.1.6.** The following universal operator $\mathcal{A}$ is undefinable in $\mathcal{L}_{SENT_\Diamond}$:

$$[\![\mathcal{A}\varphi]\!]^w_\mathcal{M} = T \quad \text{iff} \quad \forall v \in \mathcal{W}([\![\varphi]\!]^v_\mathcal{M} = T)$$

**Proof of Lem 2.1.6.** Suppose that $\mathcal{A}$ is definable in $\mathcal{L}_{SENT_\Diamond}$; that is, suppose that there is a basic modal formula $\alpha(\cdot)$ in $\mathcal{L}_{SENT_\Diamond}$ such that $[\![\mathcal{A}\varphi]\!]^w_{\mathcal{M}} = [\![\alpha(\varphi)]\!]^w_{\mathcal{M}}$.




$[\![\mathcal{A}B]\!]^w_{\mathcal{M}} = T$ so $[\![\alpha(B)]\!]^w_{\mathcal{M}} = T$.

Since $\mathcal{M}, w \leftrightarroweq \mathcal{N}, v$, $\mathcal{M}, w \leftrightsquigarrow \mathcal{N}, v$, so $[\![\alpha(B)]\!]^v_{\mathcal{N}} = [\![\mathcal{A}B]\!]^v_{\mathcal{N}} = T$.

But $[\![\mathcal{A}B]\!]^v_{\mathcal{N}} = F$.

Recall the definition of validity for sentences in $S_{\mathcal{L}_{SENT_\Diamond}}$:

**Def 2.1.5.** The sentence $\varphi$ is a *logical validity*, $\models_{SENT_\Diamond} \varphi$, just in case there is no pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = F$.

We can also define this related notion:

**Def 2.1.13.** The sentence $\varphi$ is *satisfiable* just in case there is a pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = T$.

$\varphi$ is valid if and only if $\neg\varphi$ is not satisfiable.

Like validity in sentential logic and monadic predicate logic but unlike validity in classical first-order logic, logical validity in $S_{\mathcal{L}_{SENT_\Diamond}}$ is *decidable*. There is an algorithmic procedure that for each $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ decides after a finite number of operations whether $\varphi$ is valid.

There are a number of ways to see this. Let us consider only one of them involving *filtration*.

To prove decidability, it suffices to show that basic modal logic has the *effective finite model property*.

The finite model property is this:

**Thm 2.1.1.** The sentence $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ is satisfiable just in case there is a *finite* pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = T$.

The effective finite model property is stronger:

**Thm 2.1.2.** The sentence $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ is satisfiable just in case there is a pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = T$ and $|\mathcal{W}^{\mathcal{M}}| \leq f(|\varphi|)$ where $f$ is a computable function and $|\varphi|$ is the length of $\varphi$.

To decide the validity of $\varphi \in S_{\mathcal{L}}$, we can first compute the effective bound $f(|\neg\varphi|)$ on the size of a verifying model for $\neg\varphi$. Since there are only finitely many models of a fixed size (up to isomorphism) when the valuation function is restricted to the finitely many sentence letters occurring in $\neg\varphi$, we can then check in a finite steps whether $\neg\varphi$ is true in any model of size $\leq f(|\neg\varphi|)$. If so, $\varphi$ is invalid. If not, $\varphi$ is valid.

**Def 2.1.14.** The set $sub(\varphi)$ of subsentences of $\varphi \in S_{\mathcal{L}}$ is the smallest set such that:

$\varphi \in sub(\varphi)$
$\neg\psi \in sub(\varphi)$ only if $\psi \in sub(\varphi)$
$(\psi \wedge \xi) \in sub(\varphi)$ only if $\psi, \xi \in sub(\varphi)$
$\Diamond\psi \in sub(\varphi)$ only if $\psi \in sub(\varphi)$
$\Box\psi \in sub(\varphi)$ only if $\psi \in sub(\varphi)$

Given model $\mathcal{M}$, we can define the following equivalence relation on $\mathcal{W}$:
$w \sim_{\varphi} v$ iff for all $\psi \in sub(\varphi)$, $[\![\psi]\!]^{w}_{\mathcal{M}} = [\![\psi]\!]^{v}_{\mathcal{M}}$.

Let $|w|^{\sim_{\varphi}} = \{v \in \mathcal{W} : w \sim_{\varphi} v\}$.

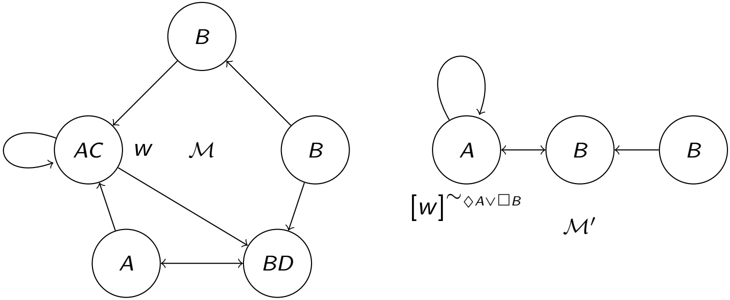
This facilitates our next transformation of $\mathcal{M}$:

**Def 2.1.15.** Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$, the *filtration of $\mathcal{M}$ through $\varphi$* is the model $\mathcal{M}' = \langle \mathcal{W}', \mathcal{R}', \mathcal{V}' \rangle$ where

$\mathcal{W}' = \{[w]^{\sim_\varphi} : w \in \mathcal{W}\}$
$\mathcal{R}' = \{\langle [w]^{\sim_\varphi}, [v]^{\sim_\varphi} \rangle : \text{there is } x \in [w]^{\sim_\varphi} \text{ and } y \in [v]^{\sim_\varphi} \text{ such that } x\mathcal{R}y\}$

$$\mathcal{V}'(p, [w]^{\sim_\varphi}) = \begin{cases} \mathcal{V}(p, w) \text{ if } p \in sub(\varphi) \\ F \text{ otherwise} \end{cases}$$

The filtration of $\mathcal{M}$ through $\Diamond A \vee \Box B$ is this:



**Thm 2.1.3.** If $\mathcal{M}'$ is the filtration of $\mathcal{M}$ through $\varphi$, then for each $w \in \mathcal{W}$ and $\psi \in \mathit{sub}(\varphi)$, $[\![\psi]\!]^{w}_{\mathcal{M}} = [\![\psi]\!]^{[w]^{\sim\varphi}}_{\mathcal{M}'}$.

The proof is a straightforward induction on the complexity of sentences in $S_{\mathcal{L}_{SENT_\Diamond}}$.

Note that $|\mathcal{W}'| \leq 2^{|\mathit{sub}(\varphi)|} \leq 2^{|\varphi|}$. So we have shown that basic modal logic has the effective finite model property.

**Exercise 2.1.1**

Consider a polymodal language with sentence letters $E$ and $M$ that designate 'There is intelligent life on Earth' and 'There is intelligent life on Mars' respectively, the Boolean constants $\neg$ and $\wedge$, and the following modal operators:

G: Going forward in time, it will always be that...
F: Going forward in time, it will sometime be that...
H: Going backward in time, it will always be that...
P: Going backward in time, it will sometime be that...

Translate the following English sentences into this formal language:

There will be intelligent life on Mars.

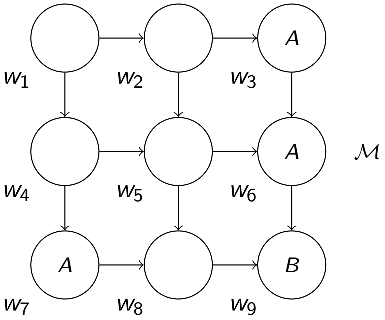
There was never intelligent life on Mars but there was on Earth.

There will never be intelligent life simultaneously on both Earth and Mars.

There will have been intelligent life on Earth.

**Exercise 2.1.2** (Van Benthem [2010])

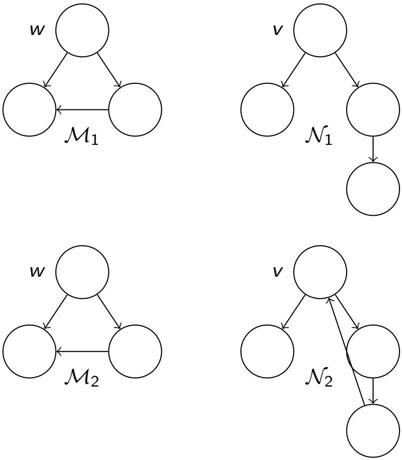
Consider the following Kripke model:




For each world in $\mathcal{M}$, find a sentence that is true only at this world.

## Exercise 2.1.3

For each pair of pointed models, determine whether these models are bisimilar. Justify each of your answers by providing a bisimulation between the pointed models or by appealing to the Invariance Lemma.



**Exercise 2.1.4**

Prove the following lemma:

**Lem.** The following 'difference' operator $\mathcal{D}$ is undefinable in $\mathcal{L}_{SENT_\Diamond}$:

$$[\![\mathcal{D}\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \exists v(w \neq v \wedge [\![\varphi]\!]^v_{\mathcal{M}} = T)$$

# Philosophical Logic

## 2.2 The Modal Zoo

Johns Hopkins University, Spring 2015

**Def 2.2.1.** The *minimal modal logic* **K** has the following rules and axioms:

| | |
|---|---|
| (PL) | All sentences with tautological form are axioms |
| (MP) | From $\varphi$ and $\varphi \supset \psi$ infer $\psi$ |
| (Nec) | From $\varphi$ infer $\Box\varphi$ |
| (K) | $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$ is an axiom schema |
| (Duality) | Expressions involving $\Box$ and $\Diamond$ are interchangeable according to the dualities $\Box \equiv \neg\Diamond\neg$ and $\Diamond \equiv \neg\Box\neg$ |

**Def 2.2.2.** A *proof* in **K** is a sequence of sentences $\langle\varphi_1, ..., \varphi_n\rangle$ in $S_{\mathcal{L}_{SENT_\Diamond}}$ such that for all $k \leq n$ one of the following conditions is met:

$\varphi_k$ is an axiom

$\exists i, j \leq k(\varphi_i \text{ is } (\varphi_j \supset \varphi_k))$

$\exists i \leq k(\varphi_k \text{ is } \Box\varphi_i)$

$\exists i \leq k(\varphi_k$ is obtained from $\varphi_i$ according to Duality)

If there is a proof in **K** ending with $\varphi$, then $\vdash_{\mathbf{K}} \varphi$.

**Example:** $\vdash_{\mathbf{K}} \Box(A \wedge B) \supset \Box A$.

| | | |
|---|---|---|
| 1 | $(A \wedge B) \supset A$ | PL |
| 2 | $\Box((A \wedge B) \supset A)$ | Nec 1 |
| 3 | $\Box((A \wedge B) \supset A) \supset (\Box(A \wedge B) \supset \Box A)$ | K |
| 4 | $\Box(A \wedge B) \supset \Box A$ | MP 3,2 |

**Example:** $\vdash_{\mathbf{K}} (\Box A \lor \Box B) \supset \Box(A \lor B)$.

| | | |
|---|---|---|
| 1 | $A \supset (A \lor B)$ | PL |
| 2 | $\Box(A \supset (A \lor B))$ | Nec 1 |
| 3 | $\Box(A \supset (A \lor B)) \supset (\Box A \supset \Box(A \lor B))$ | K |
| 4 | $\underbrace{\Box A}_{X} \supset \underbrace{\Box(A \lor B)}_{Z}$ | MP 3,2 |
| 5 | $B \supset (A \lor B)$ | PL |
| 6 | $\Box(B \supset (A \lor B))$ | Nec 5 |
| 7 | $\Box(B \supset (A \lor B)) \supset (\Box B \supset \Box(A \lor B))$ | K |
| 8 | $\underbrace{\Box B}_{Y} \supset \underbrace{\Box(A \lor B)}_{Z}$ | MP 7,6 |
| 9 | $(X \supset Z) \supset ((Y \supset Z) \supset ((X \lor Y) \supset Z))$ | PL |
| 10 | $(Y \supset Z) \supset ((X \lor Y) \supset Z)$ | MP 9,4 |
| 11 | $(\Box A \lor \Box B) \supset \Box(A \lor B)$ | MP 10,8 |

**Thm 2.2.1 (Soundness Theorem for K).** $\vdash_{\mathbf{K}} \varphi$ only if $\models_{SENT_\Diamond} \varphi$.

**Thm 2.2.2 (Completeness Theorem for K).** $\models_{SENT_\Diamond} \varphi$ only if $\vdash_{\mathbf{K}} \varphi$.

The proof of Completeness involves building a *canonical model for* **K**. See most modal logic textbooks for details.

**Def 2.2.3.** $\Gamma \vdash_{\mathbf{K}} \psi$ iff there are sentences $\varphi_1, ..., \varphi_n \in \Gamma$ such that $\vdash_{\mathbf{K}} (\varphi_1 \wedge ... \wedge \varphi_n) \supset \psi$.

**Thm 2.2.3 (Strong Soundness/Completeness Theorem for K).** $\Gamma \vdash_{\mathbf{K}} \varphi$ iff $\Gamma \models_{SENT_\Diamond} \varphi$.

N. B. $\Gamma$ can be infinite (Def 2.1.4 must be extended).

There are a host of *normal* modal logics extending **K** with additional axiom schemata:

| | |
|---|---|
| **T** | $\Box\varphi \supset \varphi$ |
| **D** | $\Box\varphi \supset \Diamond\varphi$ |
| **4** | $\Box\varphi \supset \Box\Box\varphi$ |
| **5** | $\Diamond\varphi \supset \Box\Diamond\varphi$ |
| **B** | $\varphi \supset \Box\Diamond\varphi$ |

$\mathbf{K}\varphi_1 ... \varphi_n$ is the weakest logic obtained by extending **K** with the axiom schemata $\varphi_1, ..., \varphi_n$.

For instance, **KD45** extends **K** with **D**, **4**, and **5**.

**KT4** and **KT5** are abbreviated **S4** and **S5** respectively.

One of the most beautiful parts of the theory of modal logic is the tight correspondence between the above axioms and structural constraints on the accessibility relation in Kripke models.

**Def 2.2.4.** A *frame* $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ consists of a nonempty set $\mathcal{W}$ of possible worlds and a binary accessibility relation $\mathcal{R} \subseteq \mathcal{W} \times \mathcal{W}$ between worlds. A model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ *is based on frame* $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$.

**Def 2.2.5.** The sentence $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$ is *valid on frame* $\mathcal{F}$, $\models_{\mathcal{F}} \varphi$, just in case there is no pointed model $\mathcal{M}, w$ based on $\mathcal{F}$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = F$.

We can prove correspondence lemmas like these:

**Lem 2.2.1.** For each $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$, $\models_{\mathcal{F}} \Box\varphi \supset \varphi$ iff $\mathcal{R}$ is *reflexive*.

Recall: $\mathcal{R}$ is reflexive just in case $\forall w(w\mathcal{R}w)$.

**Lem 2.2.2.** For each $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$, $\models_{\mathcal{F}} \Box\varphi \supset \Diamond\varphi$ iff $\mathcal{R}$ is *serial*.

Recall: $\mathcal{R}$ is serial just in case $\forall w \exists v(w\mathcal{R}v)$.

**Lem 2.2.3.** For each $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$, $\models_{\mathcal{F}} \Box\varphi \supset \Box\Box\varphi$ iff $\mathcal{R}$ is *transitive*.

Recall: $\mathcal{R}$ is transitive just in case $\forall w, v, u((w\mathcal{R}v \wedge v\mathcal{R}u) \supset w\mathcal{R}u)$.

**Proof of Lem 2.2.1.**

For the right-to-left direction, consider an arbitrary pointed model $\mathcal{M}, w$ based on $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is reflexive, and suppose $[\![\Box\varphi]\!]^w_{\mathcal{M}} = T$.

Since $w\mathcal{R}w$, $[\![\varphi]\!]^w_{\mathcal{M}} = T$, so $[\![\Box\varphi \supset \varphi]\!]^w_{\mathcal{M}} = T$.

Since $\mathcal{M}, w$ was arbitrary, $\models_{\mathcal{F}} \Box\varphi \supset \varphi$.

For the left-to-right direction, we prove the contrapositive.

Consider a frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *not* reflexive—that is, $\neg w\mathcal{R}w$ for some $w \in \mathcal{W}$.

Define a model $\mathcal{M}$ based on $\mathcal{F}$ by setting $\mathcal{V}(A, w) = F$ but $\mathcal{V}(A, x) = T$ for all $x \neq w$ (the valuation on other sentence letters is unimportant).

$[\![\Box A]\!]^w_{\mathcal{M}} = T$ but $[\![A]\!]^w_{\mathcal{M}} = F$, so $[\![\Box A \supset A]\!]^w_{\mathcal{M}} = F$.

Thus, $\not\models_{\mathcal{F}} \Box\varphi \supset \varphi$.

**Proof of Lem 2.2.2.**

For the right-to-left direction, consider an arbitrary pointed model $\mathcal{M}, w$ based on $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is serial, and suppose $[\![\Box\varphi]\!]^w_{\mathcal{M}} = T$.

Given the seriality of $\mathcal{R}$, $w\mathcal{R}v$ for some $v \in \mathcal{W}$.

Since $[\![\varphi]\!]^v_{\mathcal{M}} = T$, $[\![\Diamond\varphi]\!]^w_{\mathcal{M}} = T$, so $[\![\Box\varphi \supset \Diamond\varphi]\!]^w_{\mathcal{M}} = T$.

Since $\mathcal{M}, w$ was arbitrary, $\models_{\mathcal{F}} \Box\varphi \supset \Diamond\varphi$.

For the left-to-right direction, we prove the contrapositive.

Consider a frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *not* serial—that is, some $w \in \mathcal{W}$ is a dead-end state.

Since all $\Box$-sentences are true at a dead-end state and all $\Diamond$-sentences are false, $[\![\Box A \supset \Diamond A]\!]^w_{\mathcal{M}} = F$ for any $\mathcal{M}$ based on $\mathcal{F}$.

Thus, $\not\models_{\mathcal{F}} \Box\varphi \supset \Diamond\varphi$.

Such correspondence results can then be used to show that the various proof systems extending **K** are sound and complete with respect to validity on different kinds of frames.

**Thm 2.2.4 (Soundness and Completeness Theorem for KT).**
$\vdash_{\mathbf{KT}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *reflexive*.

**Thm 2.2.5 (Soundness and Completeness Theorem for KD).**
$\vdash_{\mathbf{KD}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *serial*.

**Thm 2.2.6 (Soundness and Completeness Theorem for K4).**
$\vdash_{\mathbf{K4}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *transitive*.

**Thm 2.2.7 (Soundness and Completeness Theorem for K5).**
$\vdash_{\mathbf{K5}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *Euclidean*.

Note: $\mathcal{R}$ is Euclidean just in case $\forall w, v, u((w\mathcal{R}v \wedge w\mathcal{R}u) \supset v\mathcal{R}u)$.

**Thm 2.2.8 (Soundness and Completeness Theorem for KB).**
$\vdash_{\mathbf{KB}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is *symmetric*.

Recall: $\mathcal{R}$ is symmetric just in case $\forall w, v(w\mathcal{R}v \supset v\mathcal{R}w)$.

**Thm 2.2.9 (Soundness and Completeness Theorem for S4).**
$\vdash_{\mathbf{S4}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is reflexive and transitive.

**Thm 2.2.10 (Soundness and Completeness Theorem for S5).**
$\vdash_{\mathbf{S5}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is reflexive and Euclidean.

How are these logics related?

The following partial ordering is based on proof-theoretic strength:

**Def 2.2.6.** $L_1$ is a *sublogic* of $L_2$, $L_1 \leq L_2$, just in case $\vdash_{L_1} \varphi$ implies $\vdash_{L_2} \varphi$ for each $\varphi \in S_{\mathcal{L}_{SENT_\Diamond}}$. $L_1$ is a *proper sublogic* of $L_2$, $L_1 < L_2$, just in case $L_1 \leq L_2$ but $L_2 \not\leq L_1$.

**For example:**

**Lem 2.2.4.** $\mathbf{KD} < \mathbf{KT}$.

**Proof of Lem 2.2.4.**

To show that $\mathbf{KD} \leq \mathbf{KT}$, it suffices to show that $\vdash_{\mathbf{KT}} \mathbf{D}$:

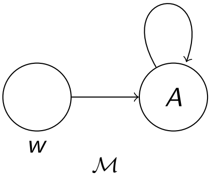
| | | |
|---|---|---|
| 1 | $\Box\varphi \supset \varphi$ | **T** Axiom |
| 2 | $\Box\neg\varphi \supset \neg\varphi$ | **T** Axiom |
| 3 | $(\Box\neg\varphi \supset \neg\varphi) \supset (\varphi \supset \neg\Box\neg\varphi)$ | PL |
| 4 | $\varphi \supset \neg\Box\neg\varphi$ | MP 3,2 |
| 5 | $\varphi \supset \Diamond\varphi$ | Duality 4 |
| 6 | $(\Box\varphi \supset \varphi) \supset ((\varphi \supset \Diamond\varphi) \supset (\Box\varphi \supset \Diamond\varphi))$ | PL |
| 7 | $(\varphi \supset \Diamond\varphi) \supset (\Box\varphi \supset \Diamond\varphi)$ | MP 6,1 |
| 8 | $\Box\varphi \supset \Diamond\varphi$ | MP 7,5 |

Alternatively, you can appeal to the fact that $\mathcal{R}$ is reflexive only if $\mathcal{R}$ is serial along with Thms 2.2.4 and 2.2.5.

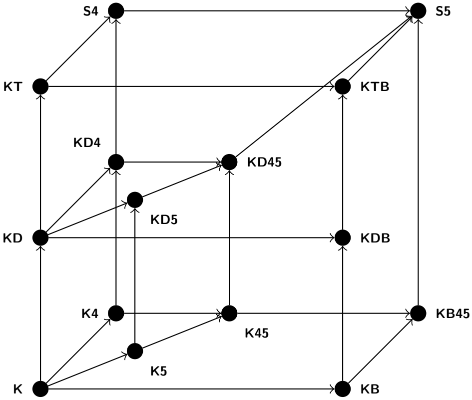
**Proof of Lem 2.2.4 continued.**

To show that **KT** $\not\leq$ **KD**, it suffices to show that $\not\models_{\mathcal{F}}$ **T** for some frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$ where $\mathcal{R}$ is serial by Thm 2.2.5.

Note that $[\![\Box A \supset A]\!]^{w}_{\mathcal{M}} = F$ in the serial model $\mathcal{M}$ below:



The landscape looks like this:

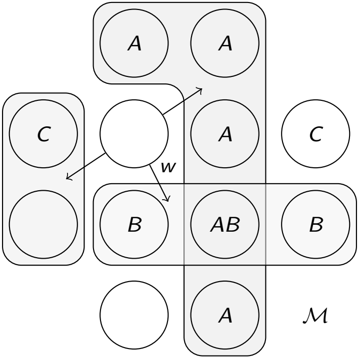


There are also many *non-normal* modal logics that lie below **K**.

In investigating these logics, it will be useful to replace the standard Kripkean semantics with an alternative *neighborhood semantics* on which some principles of **K** can fail to hold.

Instead of a binary accessibility relation between worlds in $\mathcal{W}$, our models will now include a relation between worlds and *neighborhoods* or *propositions*—sets of worlds in $2^{\mathcal{W}}$.

**Def 2.2.7.** A *neighborhood model* $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ for $\mathcal{L}_{SENT_\Diamond}$ consists of a nonempty set $\mathcal{W}$ of possible worlds, a valuation function $\mathcal{V} : At_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \to \{T, F\}$ mapping each sentence letter $p \in At_{\mathcal{L}_{SENT_\Diamond}}$ and world $w \in \mathcal{W}$ to a truth value, and a binary relation $\mathcal{R} \subseteq \mathcal{W} \times 2^{\mathcal{W}}$ between worlds and (possibly empty) sets of worlds.




The relation $\mathcal{R}$ determines a *neighborhood function* $\mathcal{N} : \mathcal{W} \to 2^{2^{\mathcal{W}}}$ where $\mathcal{N}(w) = \{X : w\mathcal{R}X\}$ is the set of neighborhoods around world $w \in \mathcal{W}$.

**Def 2.2.8.** $[\varphi]_{\mathcal{M}} = \{w : [\![\varphi]\!]^w_{\mathcal{M}} = T\}$ is the *proposition expressed* by $\varphi$ in $\mathcal{M}$.

**Def 2.2.9.** The following recursive definition of truth lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \to \{T, F\}$ for $\mathcal{L}_{SENT_\Diamond}$:

$$\begin{array}{rcl}
[\![p]\!]^w_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p, w) = T \\
[\![\bot]\!]^w_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\Diamond\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\neg\varphi]_{\mathcal{M}} \notin \mathcal{N}(w) \\
[\![\Box\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\varphi]_{\mathcal{M}} \in \mathcal{N}(w)
\end{array}$$

**Example:**

Given the previous model,

$[\![\Box A]\!]^w_{\mathcal{M}} = ?$

$[\![\Box B]\!]^w_{\mathcal{M}} = ?$

$[\![\Box(A \wedge B)]\!]^w_{\mathcal{M}} = ?$

**Def 2.2.8.** $[\varphi]_{\mathcal{M}} = \{w : [\![\varphi]\!]^{w}_{\mathcal{M}} = T\}$ is the *proposition expressed* by $\varphi$ in $\mathcal{M}$.

**Def 2.2.9.** The following recursive definition of truth lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \to \{T, F\}$ for $\mathcal{L}_{SENT_\Diamond}$:

$$
\begin{array}{rcl}
[\![p]\!]^{w}_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p, w) = T \\
[\![\bot]\!]^{w}_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^{w}_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{w}_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^{w}_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{w}_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^{w}_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^{w}_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{w}_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^{w}_{\mathcal{M}} = T \\
[\![\Diamond\varphi]\!]^{w}_{\mathcal{M}} = T & \text{iff} & [\neg\varphi]_{\mathcal{M}} \notin \mathcal{N}(w) \\
[\![\Box\varphi]\!]^{w}_{\mathcal{M}} = T & \text{iff} & [\varphi]_{\mathcal{M}} \in \mathcal{N}(w)
\end{array}
$$

**Example:**

Given the previous model,

$[\![\Box A]\!]^{w}_{\mathcal{M}} = T$.

$[\![\Box B]\!]^{w}_{\mathcal{M}} = T$.

$[\![\Box(A \wedge B)]\!]^{w}_{\mathcal{M}} = F$.

**Def 2.2.10.** The *minimal non-normal modal logic* **E** has the following rules and axioms:

| | |
|---|---|
| (PL) | All sentences with tautological form are axioms |
| (MP) | From $\varphi$ and $\varphi \supset \psi$ infer $\psi$ |
| (RE) | From $\varphi \equiv \psi$ infer $\Box\varphi \equiv \Box\psi$ |
| (Duality) | Expressions involving $\Box$ and $\Diamond$ are interchangeable according to the dualities $\Box \equiv \neg\Diamond\neg$ and $\Diamond \equiv \neg\Box\neg$ |

**Thm 2.2.11 (Soundness and Completeness Theorem for E).**
$\vdash_{\mathbf{E}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each neighborhood frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$.

Other non-normal modal logics can be generated by extending **E** with **T**, **D**, **4**, **5**, **B**, and the following axiom schemata:

| | |
|---|---|
| **M** | $\Box(\varphi \wedge \psi) \supset (\Box\varphi \wedge \Box\psi)$ |
| **C** | $(\Box\varphi \wedge \Box\psi) \supset \Box(\varphi \wedge \psi)$ |
| **N** | $\Box\neg\bot$ |

There is also a nice correspondence between the above axiom schemata and constraints on $\mathcal{N}$.

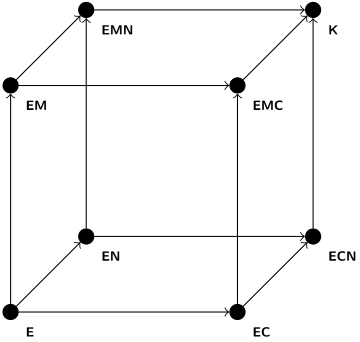
**Thm 2.2.12 (Soundness and Completeness Theorem for EM).** $\vdash_{\mathbf{EM}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$ where $\mathcal{N}$ is *closed under supersets*—that is, $(X \in \mathcal{N}(w) \wedge X \subseteq Y) \supset Y \in \mathcal{N}(w)$ for all $w \in \mathcal{W}$.

**Thm 2.2.13 (Soundness and Completeness Theorem for EC).** $\vdash_{\mathbf{EC}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$ where $\mathcal{N}$ is *closed under intersections*—that is, $X, Y \in \mathcal{N}(w) \supset X \cap Y \in \mathcal{N}(w)$ for all $w \in \mathcal{W}$.

**Thm 2.2.14 (Soundness and Completeness Theorem for EN).** $\vdash_{\mathbf{EN}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$ where $\mathcal{N}$ *contains the unit*—that is, $\mathcal{W} \in \mathcal{N}(w)$ for all $w \in \mathcal{W}$.

Adding both **M** and **C** to **E** is effectively adding the axiom schema (K). Adding **N** to **E** is effectively adding the rule (Nec). So **EMCN**=**K**.

The landscape looks like this:



**Exercise 2.2.1**

Establish the following:

$\vdash_{\mathbf{K4}} (\Box A \lor \Box B) \supset \Box(\Box A \lor \Box B)$.

$\vdash_{\mathbf{K5}} \Diamond\Box\Box A \supset \Box\Box A$.

**Exercise 2.2.2**

Prove the following lemma:

**Lem. K4** $<$ **KB5**.

**Exercise 2.2.3**

Prove the following lemmas:

**Lem.** For each $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$, $\models_{\mathcal{F}} \Box\varphi \supset \Box\Box\varphi$ iff $\mathcal{R}$ is *transitive*.

**Lem.** For each $\mathcal{F} = \langle \mathcal{W}, \mathcal{R} \rangle$, $\models_{\mathcal{F}} \Diamond\Box\varphi \supset \Box\Diamond\varphi$ iff $\mathcal{R}$ is *convergent*.

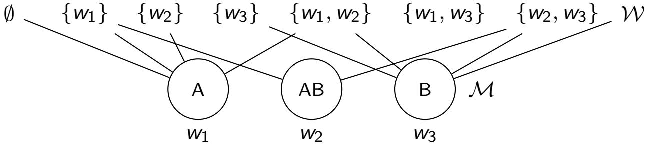
Note: $\mathcal{R}$ is convergent just in case
$\forall w, v, u((w\mathcal{R}v \wedge w\mathcal{R}u) \supset \exists x(v\mathcal{R}x \wedge u\mathcal{R}x))$.

**Exercise 2.2.4**

Consider the following neighborhood model $\mathcal{M}$:

$\mathcal{W} = \{w_1, w_2, w_3\}$
$\mathcal{N}(w_1) = \{\emptyset, \{w_1\}, \{w_2\}, \{w_1, w_2\}\}$
$\mathcal{N}(w_2) = \{\{w_1\}, \{w_2, w_3\}\}$
$\mathcal{N}(w_3) = \{\{w_3\}, \{w_1, w_2\}, \{w_2, w_3\}, \mathcal{W}\}$
$\mathcal{V}(A, w_1) = \mathcal{V}(A, w_2) = T$, $\mathcal{V}(A, w_3) = F$
$\mathcal{V}(B, w_1) = F$, $\mathcal{V}(B, w_2) = \mathcal{V}(B, w_3) = T$

$\mathcal{M}$ can be represented as follows:




In which worlds are the following sentences true?

$\Box B$, $\Box\Box(A \wedge \neg B)$, $\Diamond(\Box A \vee \Box B)$, $\Box(\Box\Box(A \wedge \neg B) \wedge \Diamond(\Box A \vee \Box B))$

**Exercise 2.2.5**

Prove that the rule (RE) preserves validity over all neighborhood frames.

# Philosophical Logic

## 2.3 Temporal Logic

Johns Hopkins University, Spring 2015

Our first application is *temporal logic*.

We will be working mostly with the following polymodal language:

**Def 2.3.1.** The *Priorean language* $\mathcal{L}_t$ has this syntax:

$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \mathsf{G}\varphi \mid \mathsf{H}\varphi$

Read $\mathsf{G}\varphi$ as 'Henceforth $\varphi$' and $\mathsf{H}\varphi$ as 'Hitherto $\varphi$'.

The duals of G and H are defined as follows: $\mathsf{F} \equiv \neg\mathsf{G}\neg$ and $\mathsf{P} \equiv \neg\mathsf{H}\neg$.

**Def 2.3.2.** The *mirror image* of $\varphi$ is the sentence obtained from $\varphi$ by switching G and H operators.

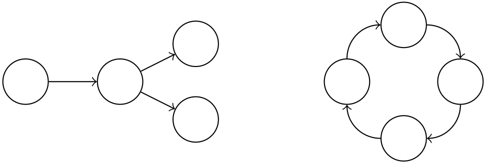
**Example:**

$(\mathsf{F}\varphi \vee \mathsf{H}\varphi) \wedge (\mathsf{G}\varphi \vee \mathsf{P}\varphi)$ is the mirror image of $(\mathsf{P}\varphi \vee \mathsf{G}\varphi) \wedge (\mathsf{H}\varphi \vee \mathsf{F}\varphi)$.

Models for $\mathcal{L}_t$ are Kripke models based on a restricted class of frames:

**Def 2.3.3.** A *flow of time* is a frame $\mathcal{F} = \langle \mathcal{T}, < \rangle$ where the *precedence relation* $<$ between points of time in $\mathcal{T}$ is transitive and irreflexive—that is, $\forall t, t', t''((t < t' < t'') \supset t < t'')$ and $\forall t(t \not< t)$.

This rules out circular time. While the left frame is a flow of time, the right frame is not (following convention, the arrows required for transitivity are omitted).



The recursive specification of truth in a pointed flow of time model is just the standard semantics supplemented with a backwards-looking clause for H:

**Def 2.3.4.** The following recursive clauses lift $\mathcal{V}$ to the complete interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_t} \times \mathcal{T} \mapsto \{T, F\}$ for $\mathcal{L}_t$:

$$
\begin{array}{rcl}
[\![p]\!]^t_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p,t) = T \\
[\![\bot]\!]^t_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^t_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^t_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^t_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^t_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^t_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^t_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^t_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^t_{\mathcal{M}} = T \\
[\![\mathsf{G}\varphi]\!]^t_{\mathcal{M}} = T & \text{iff} & \forall t' \in \{t' : t < t'\}([\![\varphi]\!]^{t'}_{\mathcal{M}} = T) \\
[\![\mathsf{H}\varphi]\!]^t_{\mathcal{M}} = T & \text{iff} & \forall t' \in \{t' : t' < t\}([\![\varphi]\!]^{t'}_{\mathcal{M}} = T)
\end{array}
$$

Many structural properties of time can be pinned down with sentences in the Priorean language. Here are some examples:

**Lem 2.3.1.** $\models_{\mathcal{F}} \mathsf{F}\varphi \supset \mathsf{G}(\mathsf{P}\varphi \vee \varphi \vee \mathsf{F}\varphi)$ iff time has a *non-branching future* $(\forall t, t', t''((t < t' \wedge t < t'') \supset (t' = t'' \vee t' < t'' \vee t'' < t')))$.

**Lem 2.3.2.** $\models_{\mathcal{F}} \mathsf{F}\neg\bot$ iff time has *no end* $(\forall t \exists t'(t < t'))$.

**Lem 2.3.3.** $\models_{\mathcal{F}} \mathsf{H}\bot \vee \mathsf{P}\mathsf{H}\bot$ iff time has *beginning points* $(\forall t, t'(t < t' \supset \exists t''(t'' < t' \wedge \neg\exists t'''(t''' < t'')))$.

**Lem 2.3.4.** $\models_{\mathcal{F}} \mathsf{F}\varphi \supset \mathsf{F}\mathsf{F}\varphi$ iff time is *dense* $(\forall t, t'(t < t' \supset \exists t''(t < t'' < t')))$.

As we will see in a moment, many properties also have correspondents within the restricted class of *linear* flows of time that do not branch in the future or past $(\forall t, t'(t < t' \vee t' < t \vee t = t'))$.

**Def 2.3.5.** The *minimal temporal logic* $\mathbf{K}_t$ has the following rules and axioms:

| | |
|---|---|
| (PL) | All sentences with tautological form are axioms |
| (MP) | From $\varphi$ and $\varphi \supset \psi$ infer $\psi$ |
| (TG) | From $\varphi$ infer $\mathsf{G}\varphi$<br>From $\varphi$ infer $\mathsf{H}\varphi$ |
| (DB) | $\mathsf{G}(\varphi \supset \psi) \supset (\mathsf{G}\varphi \supset \mathsf{G}\psi)$ is an axiom schema<br>$\mathsf{H}(\varphi \supset \psi) \supset (\mathsf{H}\varphi \supset \mathsf{H}\psi)$ is an axiom schema |
| (4) | $\mathsf{G}\varphi \supset \mathsf{GG}\varphi$ is an axiom schema |
| (CV) | $(\varphi \supset \mathsf{GP}\varphi) \wedge (\varphi \supset \mathsf{HF}\varphi)$ is an axiom schema |

**Thm 2.3.1.** $\mathbf{K}_t$ is sound and complete with respect to the class of all flows of time.

**Def 2.3.6.** The logic **Lin** is obtained by adding correspondents for non-branching future and past to $\mathbf{K}_t$:

| | |
|---|---|
| NB Future | $\forall t, t', t''((t < t' \wedge t < t'') \supset (t' = t'' \vee t' < t'' \vee t'' < t'))$ |
| | $\mathsf{F}\varphi \supset \mathsf{G}(\mathsf{P}\varphi \vee \varphi \vee \mathsf{F}\varphi)$ |
| NB Past | $\forall t, t', t''((t' < t \wedge t'' < t) \supset (t' = t'' \vee t' < t'' \vee t'' < t'))$ |
| | $\mathsf{P}\varphi \supset \mathsf{H}(\mathsf{F}\varphi \vee \varphi \vee \mathsf{P}\varphi))$ |

**Thm 2.3.2.** **Lin** is sound and complete with respect to the class of all linear flows of time.

**Def 2.3.7.** The logic **Lin.**$\mathbb{N}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

| | |
|---|---|
| Beginning of Time | $\forall t, t'(t < t' \supset \exists t''(t'' < t' \wedge \neg\exists t'''(t''' < t''))$<br>$\mathsf{H}\bot \vee \mathsf{PH}\bot$ |
| No End of Time | $\forall t \exists t'(t < t')$<br>$\mathsf{F}\neg\bot$ |
| Finite Intervals | $\forall t, t'(\exists^{\text{finite}} t''(t < t'' < t'))$<br>$(\mathsf{G}(\mathsf{G}\varphi \supset \varphi) \supset (\mathsf{FG}\varphi \supset \mathsf{G}\varphi))$<br>$(\mathsf{H}(\mathsf{H}\varphi \supset \varphi) \supset (\mathsf{PH}\varphi \supset \mathsf{H}\varphi))$ |

**Thm 2.3.3.** **Lin.**$\mathbb{N}$ is sound and complete with respect to $\langle \mathbb{N}, < \rangle$.

**Def 2.3.8.** The logic **Lin.**$\mathbb{Z}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

| | |
|---|---|
| No Beginning of Time | $\forall t \exists t'(t' < t)$ <br> $\mathsf{P}\neg\bot$ |
| No End of Time | $\forall t \exists t'(t < t')$ <br> $\mathsf{F}\neg\bot$ |
| Finite Intervals | $\forall t, t'(\exists^{\text{finite}} t''(t < t'' < t'))$ <br> $(\mathsf{G}(\mathsf{G}\varphi \supset \varphi) \supset (\mathsf{F}\mathsf{G}\varphi \supset \mathsf{G}\varphi))$ <br> $(\mathsf{H}(\mathsf{H}\varphi \supset \varphi) \supset (\mathsf{P}\mathsf{H}\varphi \supset \mathsf{H}\varphi))$ |

**Thm 2.3.4.** **Lin.**$\mathbb{Z}$ is sound and complete with respect to $\langle \mathbb{Z}, < \rangle$.

**Def 2.3.9.** The logic **Lin.**$\mathbb{Q}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

| | |
|---|---|
| No Beginning of Time | $\forall t \exists t'(t' < t)$<br>$\mathsf{P}\neg\bot$ |
| No End of Time | $\forall t \exists t'(t < t')$<br>$\mathsf{F}\neg\bot$ |
| Density | $\forall t, t'(t < t' \supset \exists t''(t < t'' < t')$<br>$\mathsf{F}\varphi \supset \mathsf{FF}\varphi$ |

**Thm 2.3.5.** **Lin.**$\mathbb{Q}$ is sound and complete with respect to $\langle \mathbb{Q}, < \rangle$.

**Def 2.3.10.** The logic **Lin.**$\mathbb{R}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

| | |
|---|---|
| No Beginning of Time | $\forall t \exists t'(t' < t)$<br>$\mathsf{P}\neg\bot$ |
| No End of Time | $\forall t \exists t'(t < t')$<br>$\mathsf{F}\neg\bot$ |
| Density | $\forall t, t'(t < t' \supset \exists t''(t < t'' < t')$<br>$\mathsf{F}\varphi \supset \mathsf{FF}\varphi$ |
| Dedekind Continuity | $\forall X(\forall t, t'((t \in X \wedge t' \notin X) \supset t < t') \supset$<br>$\exists t''(\forall t, t'((t \neq t'' \neq t' \wedge t \in X \wedge t' \notin X) \supset$<br>$(t < t'' < t'))))$<br>$(\mathsf{FH}\varphi \wedge \mathsf{F}\neg\varphi \wedge \mathsf{G}(\neg\varphi \supset \mathsf{G}\neg\varphi)) \supset$<br>$\mathsf{F}((\varphi \wedge \mathsf{G}\neg\varphi) \vee (\neg\varphi \wedge \mathsf{H}\varphi))$ |

**Thm 2.3.6. Lin.**$\mathbb{R}$ is sound and complete with respect to $\langle \mathbb{R}, < \rangle$.

We can extend the Priorean language $\mathcal{L}_t$ with other temporal operators.

The *progressive* operator $\Pi$ is intended to capture something being in progress:

$$[\![\Pi\varphi]\!]^t_{\mathcal{M}} = T \quad \text{iff} \quad \begin{array}{l} \exists t', t''(t' < t < t'' \wedge \\ \forall t''' \in \{t''' : t' < t''' < t''\}([\![\varphi]\!]^{t'''}_{\mathcal{M}} = T)) \end{array}$$

The *nexttime* or *tomorrow* operator $\mathcal{X}$ allows one to talk about the next state of a process ($\mathcal{X}$ applies only to discrete flows of time):

$$[\![\mathcal{X}\varphi]\!]^t_{\mathcal{M}} = T \quad \text{iff} \quad [\![\varphi]\!]^{t+1}_{\mathcal{M}} = T$$

The dyadic operators $\mathcal{S}$ and $\mathcal{U}$ are intended to formalize 'since' and 'until' respectively:

$$[\![\mathcal{S}\varphi\psi]\!]^t_{\mathcal{M}} = T \quad \text{iff} \quad \begin{array}{l} \exists t' < t([\![\varphi]\!]^{t'}_{\mathcal{M}} = T) \wedge \\ \forall t'' \in \{t'' : t' < t'' < t\}([\![\psi]\!]^{t''}_{\mathcal{M}} = T) \end{array}$$

$$[\![\mathcal{U}\varphi\psi]\!]^t_{\mathcal{M}} = T \quad \text{iff} \quad \begin{array}{l} \exists t' > t([\![\varphi]\!]^{t'}_{\mathcal{M}} = T) \wedge \\ \forall t'' \in \{t'' : t < t'' < t'\}([\![\psi]\!]^{t''}_{\mathcal{M}} = T) \end{array}$$

It can be shown (using bisimulation) that adding any of these operators to $\mathcal{L}_t$ increases expressive power.

Another motivation for extending $\mathcal{L}_t$ arises with branching time.

If time branches in the future, then 'It will be the case that...' is ambiguous.

On a strong reading, it will be the case that $\varphi$ just in case $\varphi$ happens *come what may*; that is, $\varphi$ occurs in every possible future.

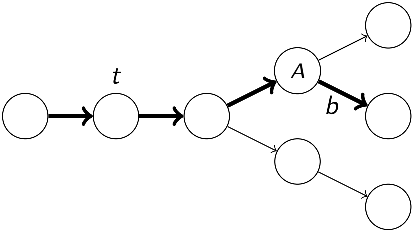
On a weaker reading, it will be the case that $\varphi$ just in case $\varphi$ occurs in the *actual* future.

These two readings correspond to the Peircean and Ockhamist schools respectively.

Let us now consider flows of time that are *trees*—there must be a path along $< \cup >$ between any two times $t, t' \in \mathcal{T}$ but the flow does not branch in the past.

A *branch* $b$ is a maximal linearly ordered subset of $\mathcal{T}$.

If $t \in b$, then $t$ *lies on* $b$ and $b$ *passes through* $t$.



On the Peircean model, the Priorean language $\mathcal{L}_t$ is supplemented with an operator $\mathsf{F}_\Box$ with the following truth clause:

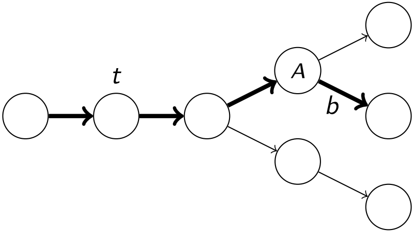
$$[\![\mathsf{F}_\Box\varphi]\!]^t_{\mathcal{M}} = T \quad \text{iff} \quad \forall b(t \in b \supset \exists t' \in \{t' \in b : t < t'\}([\![\varphi]\!]^{t'}_{\mathcal{M}} = T))$$

That is, $\mathsf{F}_\Box\varphi$ is true at $t$ in $\mathcal{M}$ just in case there is some future time $t'$ lying on every branch $b$ passing through $t$ where $\varphi$ is true at $t'$.

On the Ockhamist model, sentences are evaluated for truth relative *both* to a time $t \in \mathcal{T}$ *and* to a branch $b \subseteq \mathcal{T}$ where $t \in b$ (the actual history):

$$
\begin{aligned}
[\![p]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad \mathcal{V}(p,t) = T \\
[\![\bot]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad 0 = 1 \\
[\![\neg\varphi]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad [\![\varphi]\!]^{t,b}_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad [\![\varphi]\!]^{t,b}_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^{t,b}_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad [\![\varphi]\!]^{t,b}_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^{t,b}_{\mathcal{M}} = T \\
[\![\mathsf{G}\varphi]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad \forall t' \in \{t' \in b : t < t'\}([\![\varphi]\!]^{t',b}_{\mathcal{M}} = T) \\
[\![\mathsf{H}\varphi]\!]^{t,b}_{\mathcal{M}} = T &\quad \text{iff} \quad \forall t' \in \{t' \in b : t' < t\}([\![\varphi]\!]^{t',b}_{\mathcal{M}} = T)
\end{aligned}
$$

F and P are still defined as follows: $\mathsf{F} \equiv \neg\mathsf{G}\neg$ and $\mathsf{P} \equiv \neg\mathsf{H}\neg$.




**Example:**

In the Peircean framework, $[\![ \mathsf{F}_{\Box} A ]\!]^{t}_{\mathcal{M}} = F$.

In the Ockhamist framework, $[\![ \mathsf{F} A ]\!]^{t,b}_{\mathcal{M}} = T$.

In addition to the temporal operators G and H, the Ockhamist language also includes another necessity modal $\Box$ that quantifies over branches:

$$[\![\Box\varphi]\!]^{t,b}_{\mathcal{M}} = T \quad \text{iff} \quad \forall c(t \in c \supset [\![\varphi]\!]^{t,c}_{\mathcal{M}} = T)$$

The following equivalence holds: $[\![\mathsf{F}_{\Box}\varphi]\!]^{t}_{\mathcal{M}} = [\![\Box\mathsf{F}\varphi]\!]^{t,b}_{\mathcal{M}}$.

So the Peircean language can be regarded as a fragment of the Ockhamist language.

**Exercise 2.3.1**

Prove the following lemma:

**Lem.** For each $\mathcal{F} = \langle \mathcal{T}, < \rangle$, $\models_{\mathcal{F}} \mathsf{F}\varphi \supset \mathsf{F}\mathsf{F}\varphi$ iff time is *dense*.

## Exercise 2.3.2

In the temporal setting, we can also introduce the notion of bisimulation between models:

**Def.** Given $\mathcal{M} = \langle \mathcal{T}^{\mathcal{M}}, <^{\mathcal{M}}, \mathcal{V}^{\mathcal{M}} \rangle$ and $\mathcal{N} = \langle \mathcal{T}^{\mathcal{N}}, <^{\mathcal{N}}, \mathcal{V}^{\mathcal{N}} \rangle$, a *temporal bisimulation* between $\mathcal{M}, s$ and $\mathcal{N}, t$ is a binary relation $\mathcal{Z} \subseteq \mathcal{T}^{\mathcal{M}} \times \mathcal{T}^{\mathcal{N}}$ such that $s\mathcal{Z}t$ and for all times $x \in \mathcal{T}^{\mathcal{M}}$ and $y \in \mathcal{T}^{\mathcal{N}}$, if $x\mathcal{Z}y$ then:

(atomic harmony) For all $p \in At_{\mathcal{L}_t}$, $\mathcal{V}^{\mathcal{M}}(p, x) = \mathcal{V}^{\mathcal{N}}(p, y)$.

(zig) If $x <^{\mathcal{M}} z$, then there exists $z' \in \mathcal{T}^{\mathcal{N}}$ such that $y <^{\mathcal{N}} z'$ and $z\mathcal{Z}z'$.

(zag) If $y <^{\mathcal{N}} z'$, then there exists $z \in \mathcal{T}^{\mathcal{M}}$ such that $x <^{\mathcal{M}} z$ and $z'\mathcal{Z}z$.

(converse zig) If $z <^{\mathcal{M}} x$, then there exists $z' \in \mathcal{T}^{\mathcal{N}}$ such that $z' <^{\mathcal{N}} y$ and $z\mathcal{Z}z'$.

(converse zag) If $z' <^{\mathcal{N}} y$, then there exists $z \in \mathcal{T}^{\mathcal{M}}$ such that $z <^{\mathcal{M}} x$ and $z'\mathcal{Z}z$.

We say that $\mathcal{M}, s$ and $\mathcal{N}, t$ are *temporally bisimilar*: $\mathcal{M}, s \leftrightarroweq \mathcal{N}, t$.

**Exercise 2.3.2 continued**

This Invariance Lemma also holds (where modal equivalence applies to the Priorean language $\mathcal{L}_t$):

**Lem.** $\mathcal{M}, s \leftrightarroweq \mathcal{N}, t$ only if $\mathcal{M}, s \leftrightsquigarrow \mathcal{N}, t$.

Using this lemma, prove that the 'since' operator $\mathcal{S}$ is undefinable in $\mathcal{L}_t$:

$$\begin{array}{lll} [\![\mathcal{S}\varphi\psi]\!]^t_{\mathcal{M}} = T & \text{iff} & \exists t' < t([\![\varphi]\!]^{t'}_{\mathcal{M}} = T) \wedge \\ & & \forall t'' \in \{t'' : t' < t'' < t\}([\![\psi]\!]^{t''}_{\mathcal{M}} = T) \end{array}$$

# Philosophical Logic

## 2.4 Deontic Logic

Johns Hopkins University, Spring 2015

Our next application is *deontic logic* which is concerned with obligation, permission, prohibition, and related normative concepts.

**Def 2.4.1.** The *deontic language* $\mathcal{L}_d$ extends the basic sentential language with obligation and permission operators:

$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \mathsf{O}\varphi \mid \mathsf{P}\varphi$

Read $\mathsf{O}\varphi$ as 'It ought to be the case that $\varphi$' and $\mathsf{P}\varphi$ as 'It is permissible that $\varphi$'.

O and P are interdefinable: $\mathsf{O} \equiv \neg\mathsf{P}\neg$ and $\mathsf{P} \equiv \neg\mathsf{O}\neg$.

The semantics for $\mathcal{L}_d$ is the standard one based on Kripke models where $w\mathcal{R}v$ just in case $v$ is a *deontically ideal* world relative to $w$.

Intuitively, $\mathsf{O}\varphi \supset \varphi$ is invalid; that is, $\mathcal{R}$ needn't be reflexive.

Intuitively, $\mathsf{O}\varphi \supset \mathsf{P}\varphi$ is valid; that is, $\mathcal{R}$ should be serial.

At first glance, then, it seems that deontic logic should be at least as strong as **KD** but shouldn't validate **T**.

**Def 2.4.2.** *Standard Deontic Logic* (SDL) is the logic **KD**:

| | |
|---|---|
| (PL) | All sentences with tautological form are axioms |
| (MP) | From $\varphi$ and $\varphi \supset \psi$ infer $\psi$ |
| ($\mathrm{Nec}_d$) | From $\varphi$ infer $\mathrm{O}\varphi$ |
| ($\mathrm{K}_d$) | $\mathrm{O}(\varphi \supset \psi) \supset (\mathrm{O}\varphi \supset \mathrm{O}\psi)$ is an axiom schema |
| ($\mathrm{D}_d$) | $\mathrm{O}\varphi \supset \mathrm{P}\varphi$ is an axiom schema |
| (Duality) | Expressions involving O and P are interchangeable according to the dualities $\mathrm{O} \equiv \neg\mathrm{P}\neg$ and $\mathrm{P} \equiv \neg\mathrm{O}\neg$ |

Intuitively, $\mathrm{O}(\mathrm{O}\varphi \supset \varphi)$ is also valid; while $\mathcal{R}$ needn't be reflexive, this relation should be *shift reflexive*: $\forall w, v(w\mathcal{R}v \supset v\mathcal{R}v)$.

**Def 2.4.3.** $\mathrm{SDL}^+$ is the logic obtained by supplementing **KD** with the axiom schema $\mathrm{O}(\mathrm{O}\varphi \supset \varphi)$.

In lieu of treating O as a primitive operator, Anderson [1956] and Kanger [1957] proposed reducing deontic logic to ordinary alethic logic.

$\mathrm{O}\varphi \equiv \Box(D \supset \varphi)$ (alternatively: $\mathrm{O}\varphi \equiv \Box(\neg\varphi \supset S)$)

where $D$ designates that all normative requirements have been met (and $S$ designates that a sanction has been imposed).

If the logic of $\Box$ is **K** plus the axiom $\Diamond D$, then the logic of O is SDL.

Here are some theorems of the combined logic of $\Box$ and O:

- $\mathrm{O}D$
- $\Box\varphi \supset \mathrm{O}\varphi$
- $\Box(\varphi \supset \psi) \supset (\mathrm{O}\varphi \supset \mathrm{O}\psi)$
- $\neg\Diamond(\mathrm{O}\varphi \wedge \mathrm{O}\neg\varphi)$
- $\mathrm{O}\varphi \supset \Diamond\varphi$ ('ought' implies 'can')

If the logic of $\Box$ is **KT** plus the axiom $\Diamond D$, then the logic of O is SDL$^{+}$.

$\mathsf{SDL}/\mathsf{SDL}^{+}$ have their share of problems.

**Conflicting Obligations.** SDL rules out conflicting obligations:

| | | |
|---|---|---|
| 1 | $(O\varphi \wedge O\neg\varphi) \supset O(\varphi \wedge \neg\varphi)$ | $C_d$ |
| 2 | $\neg O(\varphi \wedge \neg\varphi)$ | From $D_d$ |
| 3 | $\neg(O\varphi \wedge O\neg\varphi)$ | PL 1,2 |

N. B. In this section, I will present only proof sketches in SDL.

However, such conflicts arguably occur:

(1) I ought to fight in the war (since I signed a contract to do so).

(2) I ought not to fight in the war (since the war is unjust).

**Response:** Abandon $\mathbf{C}_d$ and work with neighborhood semantics.

**Response:** Abandon $\mathbf{D}_d$.

**Response:** Deny the possibility of conflicting obligations. Allow for different kinds of 'ought' (moral, prudential, all-things-considered, etc.) and deny that conflict can arise for any particular 'ought'.

**Free Choice Permission.** The following inference seems good:

(P1) You may have the whiskey or the gin.
(C) You may have the whiskey and you may have the gin.

However, this inference is invalidated by SDL.

**Response:** Appeal to Gricean conversational implicature.

**Response:** Abandon the standard semantics for $\vee$.

**Ross' Paradox.** The following inferences seem terrible:

(P1) You ought to mail the letter.
(C) You ought to mail the letter or burn it.

(P1) You may have the whiskey.
(C) You may have the whiskey or the gin.

However, these inferences are validated by SDL given the axiom $\mathbf{M}_d$.

**Response:** Abandon $\mathbf{M}_d$ and work with neighborhood semantics.

**Response:** Explain the oddness of the inferences in pragmatic terms.

**Good Samaritan Paradox.** Consider the following argument (Prior [1958]):

(P1) It ought to be that Jones helps Smith who has been robbed.
(C) It ought to be that Smith has been robbed.

This is terrible but comes out valid in SDL:

| | | |
|---|---|---|
| 1 | $\mathsf{O}(H \wedge R)$ | P1 |
| 2 | $\mathsf{O}R$ | $\mathsf{M}_d$, PL 1 |

**Response:** Deny that $\mathrm{O}(H \wedge R)$ is a good translation of P1.

**Response:** Abandon $\mathbf{M}_d$ and work with neighborhood semantics.

**Paradox of Epistemic Obligation.** Consider the following argument (Aqvist [1967]):

(P1) There is a fire.
(P2) If there is a fire, it ought to be that the firefighter knows that there is a fire.
(C) It ought to be that there is a fire.

This is terrible but comes out valid in SDL:

| | | |
|---|---|---|
| 1 | $F$ | P1 |
| 2 | $F \supset \mathrm{O}\mathrm{K}_f F$ | P2 |
| 3 | $\mathrm{O}\mathrm{K}_f F$ | PL 2,1 |
| 4 | $\mathrm{K}_f F \supset F$ | Factivity of $\mathrm{K}_f$ |
| 5 | $\mathrm{O}\mathrm{K}_f F \supset \mathrm{O}F$ | $\mathrm{Nec}_d$, $\mathrm{K}_d$, PL 4 |
| 6 | $\mathrm{O}F$ | PL 5,3 |

**Response:** Abandon $\mathbf{K}_d$ and work with neighborhood semantics.

**Response:** Abandon $\mathrm{Nec}_d$ (but note that this is applied only to $\mathsf{K}_f F \supset F$ which is presumably a validity).

**Chisholm's Paradox.** The following statements appear consistent and pairwise logically independent (Chisholm [1963]):

(P1) It ought to be that Jones goes to help his neighbors.
(P2) It ought to be that Jones tells his neighbors he is coming if he is going to help them.
(P3) If Jones doesn't go to help, it ought to be that he doesn't tell his neighbors he is coming.
(P4) Jones doesn't go help his neighbors.

However, these statements are inconsistent in SDL:

| | | |
|---|---|---|
| 1 | $\mathsf{O}H$ | P1 |
| 2 | $\mathsf{O}(H \supset T)$ | P2 |
| 3 | $\neg H \supset \mathsf{O}\neg T$ | P3 |
| 4 | $\neg H$ | P4 |
| 5 | $\mathsf{O}H \supset \mathsf{O}T$ | $\mathsf{K}_d$, PL 2 |
| 6 | $\mathsf{O}T$ | PL 5,1 |
| 7 | $\mathsf{P}T$ | $\mathsf{D}_d$, PL 6 |
| 8 | $\neg\mathsf{O}\neg T$ | Duality 7 |
| 9 | $\mathsf{O}\neg T$ | PL 3,4 |
| 10 | $\perp$ | PL 8,9 |

**Response:** Translate P2 and P3 as $H \supset \mathrm{O}T$ and $\neg H \supset \mathrm{O}\neg T$ respectively. But then P2 follows from P4 so the statements are not independent.

**Response:** Translate P2 and P3 as $\mathrm{O}(H \supset T)$ and $\mathrm{O}(\neg H \supset \neg T)$ respectively. But then P3 follows from P1 so the statements are not independent.

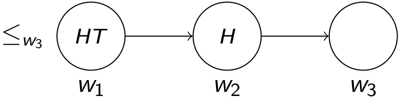
**Response:** Replace the unary obligation and permission operators with the dyadic operators $\mathsf{O}(\psi/\varphi)$ and $\mathsf{P}(\psi/\varphi)$.

Read $\mathsf{O}(\psi/\varphi)$ as 'It ought to be the case that $\psi$ *given that* $\varphi$' and $\mathsf{P}(\psi/\varphi)$ as 'It is permissible that $\psi$ *given that* $\varphi$'.

The semantics for these operators is similar to the semantics for counterfactuals in using an ordering on worlds (Lewis [1973]):

$v \leq_w u$ just in case $v$ is as good as $u$ relative to $w$.

Translating the premises as $\mathsf{O}(H/\neg\bot)$, $\mathsf{O}(T/H)$, $\mathsf{O}(\neg T/\neg H)$, and $\neg H$, these are all true at $w_3$ in the model below:




**Response:** Keep the unary deontic operators but replace the material conditional with a more sophisticated intensional conditional.

**Miners Paradox.** Consider a scenario where ten miners are trapped inside shaft A or shaft B, but you do not know which. Floodwaters are approaching these shafts. Armed with some sandbags, you can block either shaft A or shaft B, but not both shafts. If you block the shaft with the miners, then all ten live. If you block the shaft without the miners and divert the water into the other shaft containing them, then all ten die. If you block neither shaft, then the water will flow into both shafts, killing only the single miner who is lowest down. What should you do?

(Kolodny and MacFarlane [2010] credit this example to Derek Parfit who in turn credits Donald Regan)

Answer: You ought to block neither shaft.

But you might reason as follows:

| | | |
|---|---|---|
| 1 | $InA \vee InB$ | P |
| 2 | $InA \supset \mathrm{O}(Bl A)$ | P |
| 3 | $InB \supset \mathrm{O}(Bl B)$ | P |
| 4 | $InA \supset (\mathrm{O}(Bl A) \vee \mathrm{O}(Bl B))$ | PL 2 |
| 5 | $InB \supset (\mathrm{O}(Bl A) \vee \mathrm{O}(Bl B))$ | PL 3 |
| 6 | $\mathrm{O}(Bl A) \vee \mathrm{O}(Bl B)$ | PL 1,4,5 |

Where does this reasoning go wrong?

Kolodny and MacFarlane [2010] canvas many possible responses.

**Exercise 2.4.1**

Provide a pointed neighborhood model for the deontic language $\mathcal{L}_d$ in which $\mathsf{O}A \wedge \mathsf{O}\neg A$ is true.

**Exercise 2.4.2**

Prove that the following arguments are not valid over neighborhood models for $\mathcal{L}_d$.

(P1) $\mathrm{O}A$
(C) $\mathrm{O}(A \vee B)$

(P1) $\mathrm{P}A$
(C) $\mathrm{P}(A \vee B)$

**Exercise 2.4.3**

How would you respond to the Miners Paradox? Argue for your preferred solution in a couple of paragraphs.

# Philosophical Logic

## 2.5 Epistemic Logic

Johns Hopkins University, Spring 2015

Another important application is *epistemic logic* which is concerned with the individual and collective knowledge states of groups of agents, and how these knowledge states evolve when new information comes to light.

**Def 2.5.1.** The *epistemic language* $\mathcal{L}_e$ extends the basic sentential language with knowledge operators for each agent in Agt$=\{a, b, c, ...\}$:

$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \mathsf{K}_a\varphi$

Read $\mathsf{K}_a\varphi$ as 'Agent *a* knows that $\varphi$'.

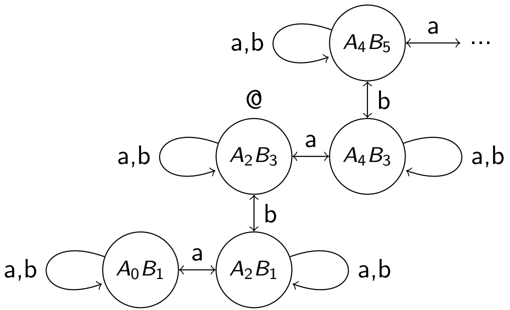
Read the dual $\mathsf{K}_a^*\varphi \equiv \neg\mathsf{K}_a\neg\varphi$ as 'It is compatible with what Agent *a* knows that $\varphi$'.

**Def 2.5.2.** A model $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a\}_{a \in \mathrm{Agt}}, \mathcal{V} \rangle$ for $\mathcal{L}_e$ is a standard Kripke model with an epistemic accessibility relation for each agent in Agt, where $w\mathcal{R}_a v$ just in case $v$ is epistemically possible for Agent $a$ in $w$; that is, Agent $a$'s knowledge in $w$ leaves open $v$.

@ will designate the actual world.

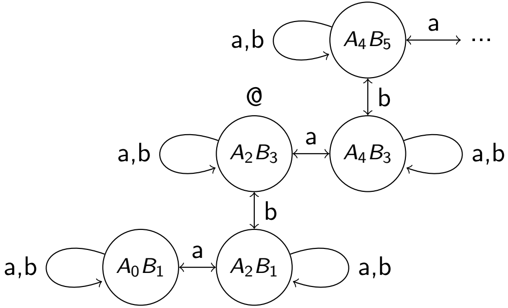
**Example:**

Suppose that Agent $a$ has 2 written on her forehead and Agent $b$ has 3 written on his forehead. Each agent can see the other's forehead but they do not know the number on their own forehead. They are told by a reliable source that the numbers on their foreheads are $n$ and $n + 1$ for some $n \in \mathbb{N}$.




where $A_n$ designates that Agent *a* has *n* on her forehead, and $B_n$ designates that Agent *b* has *n* on his forehead.

Note that each *a*-arrow and *b*-arrow reflects Agent *a*'s and Agent *b*'s *ignorance* respectively.

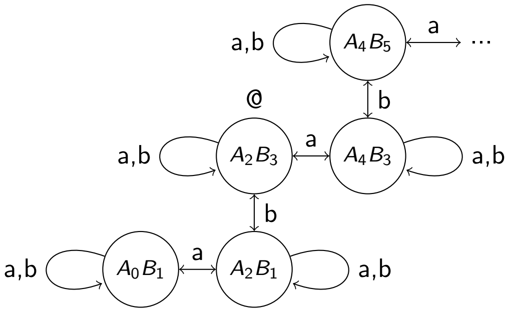



In this model,

$[\![\mathsf{K}_a B_3 \wedge \mathsf{K}_b A_2]\!]^{@}_{\mathcal{M}} = ?$

$[\![\mathsf{K}_a \neg B_5 \wedge \mathsf{K}_b \neg B_5]\!]^{@}_{\mathcal{M}} = ?$

$[\![\mathsf{K}_a \mathsf{K}_b \neg B_5]\!]^{@}_{\mathcal{M}} = ?$




In this model,

$[\![\mathsf{K}_a B_3 \wedge \mathsf{K}_b A_2]\!]^{@}_{\mathcal{M}} = T.$

$[\![\mathsf{K}_a \neg B_5 \wedge \mathsf{K}_b \neg B_5]\!]^{@}_{\mathcal{M}} = T.$

$[\![\mathsf{K}_a \mathsf{K}_b \neg B_5]\!]^{@}_{\mathcal{M}} = F.$

Since knowledge is factive, the **T** axiom $\mathsf{K}_a\varphi \supset \varphi$ is valid (hence the reflexive loops for each agent at each world in the above model—these are typically left implicit).

The validity of other axioms is more controversial.

The **4** axiom is the famous KK-principle: $\mathsf{K}_a\varphi \supset \mathsf{K}_a\mathsf{K}_a\varphi$ (also known as *positive introspection*).

While philosophers reject **5** (also known as *negative introspection*) and **B**, economists and computer scientists typically assume that the logic of knowledge is **S5**.

Williamson [2000] against the KK-principle:

Suppose that you are looking at a tree in the distance. If $H_n$ designates that the tree is $n$ inches tall, then $H_{1000}$. Given that your eyesight is imperfect, $\mathsf{K}\neg H_n \supset \neg H_{n+1}$, and you can come to know this by reflecting on your visual limitations.

But Williamson argues that KK then leads to trouble:

| | | |
|---|---|---|
| 1. | $\mathsf{K}\neg H_{500}$ | PL |
| 2. | $\mathsf{K}\neg H_{500} \supset \neg H_{501}$ | Assumption |
| 3. | $\mathsf{K}(\mathsf{K}\neg H_{500} \supset \neg H_{501})$ | Assumption |
| 4. | $\mathsf{KK}\neg H_{500}$ | 4 Axiom 1 |
| 5. | $\mathsf{KK}\neg H_{500} \supset \mathsf{K}\neg H_{501}$ | K Axiom 3 |
| 6. | $\mathsf{K}\neg H_{501}$ | PL 5,4 |

Repeating this reasoning, you can infer $\mathsf{K}\neg H_{1000}$, so $\neg H_{1000}$ by the T Axiom. This contradicts $H_{1000}$.

We can also define some interesting notions of collective knowledge.

For instance, we might introduce this *everyone in $X$ knows* operator $\mathsf{E}_X$ (where $X \subseteq \mathrm{Agt}$):

$$[\![\mathsf{E}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall a \in X([\![\mathsf{K}_a\varphi]\!]^w_{\mathcal{M}} = T)$$

If $|X|$ is finite, then $\mathsf{E}_X$ is clearly definable in $\mathcal{L}_e$: $\mathsf{E}_X\varphi \equiv \bigwedge_{a\in X} \mathsf{K}_a\varphi$.

We might also introduce this *someone in $X$ knows* operator $\mathsf{S}_X$:

$$[\![\mathsf{S}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \exists a \in X([\![\mathsf{K}_a\varphi]\!]^w_{\mathcal{M}} = T)$$

If $|X|$ is finite, then $\mathsf{S}_X\varphi \equiv \bigvee_{a\in X} \mathsf{K}_a\varphi$.

More interestingly, we might introduce the notions of *common knowledge* and *distributed knowledge*.

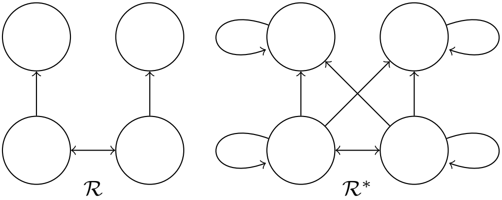
**Def 2.5.3.** Something is *common knowledge* among a group of agents $X$ iff everyone in $X$ knows it and everyone in $X$ knows that everyone in $X$ knows it, and so forth:

$$[\![\mathsf{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad [\![\mathsf{E}_X\varphi]\!]^w_{\mathcal{M}} = [\![\mathsf{E}_X\mathsf{E}_X\varphi]\!]^w_{\mathcal{M}} = ... = T$$

Our Kripke models afford a more elegant truth clause:

$$[\![\mathsf{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \begin{array}{l}\text{for all } v \in \mathcal{W}\text{, if } v \text{ is reachable from } w\\ \text{in a finite number of steps along any } \mathcal{R}_a\\ \text{where } a \in X\text{, then } [\![\varphi]\!]^v_{\mathcal{M}} = T\end{array}$$

Given a relation $\mathcal{R}$, let $\mathcal{R}^*$ be the *reflexive transitive closure* of $\mathcal{R}$; that is, $\mathcal{R}^*$ is the relation obtained from $\mathcal{R}$ by adding reflexive loops and whatever is required for transitivity.



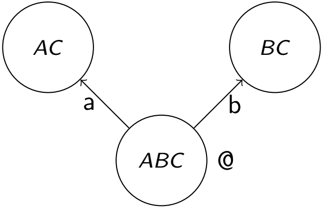

The truth clause for $\mathsf{C}_X$ can be restated thus:

$$[\![\mathsf{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w(\textstyle\bigcup_{a\in X}\mathcal{R}_a)^* v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)$$

**Def 2.5.4.** Something is *distributed knowledge* among a group of agents $X$ iff, roughly, the agents would know it were they to share all of their individual knowledge:

$$[\![\mathrm{D}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w \bigcap_{a \in X} \mathcal{R}_a v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)$$

Informally, $v$ is an epistemic possibility post-sharing in $w$ just in case $v$ is epistemically possible for each member of $X$ in $w$. If any agent's individual knowledge rules out $v$, then the agents' distributed knowledge will also rule out $v$.



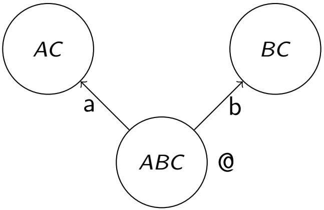

In this model,

$[\![ \mathsf{E}_X (A \vee B) ]\!]^{@}_{\mathcal{M}} = ?$

$[\![ \mathsf{S}_X A ]\!]^{@}_{\mathcal{M}} = ?$

$[\![ \mathsf{C}_X C ]\!]^{@}_{\mathcal{M}} = ?$

$[\![ \mathsf{D}_X (A \wedge B) ]\!]^{@}_{\mathcal{M}} = ?$




In this model,

$[\![\mathsf{E}_X(A \lor B)]\!]^{@}_{\mathcal{M}} = T$.

$[\![\mathsf{S}_X A]\!]^{@}_{\mathcal{M}} = T$.

$[\![\mathsf{C}_X C]\!]^{@}_{\mathcal{M}} = T$.

$[\![\mathsf{D}_X(A \land B)]\!]^{@}_{\mathcal{M}} = T$.

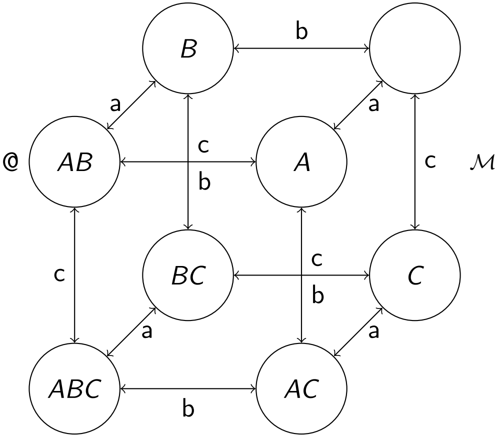
Let us close with a nice application of epistemic logic: the Muddy Children Puzzle.

Three children $a$, $b$, and $c$ have been playing outside in the mud.

Let $A$, $B$, and $C$ designate that $a$, $b$, and $c$ have mud on their forehead respectively.

In fact, $A$ and $B$ but $\neg C$.

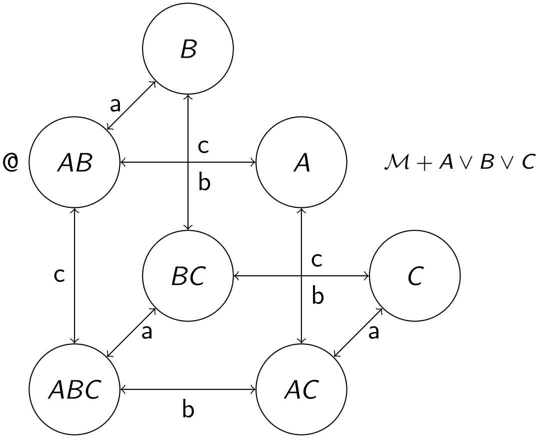
Here is the initial epistemic model:



Returning home, their mother says that at least one of the children has mud on their forehead. How does the model change?

**Def 2.5.5.** Given $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a\}_{a \in \mathsf{Agt}}, \mathcal{V} \rangle$, the *model updated with* $\varphi$ is $\mathcal{M} + \varphi = \langle \mathcal{W} + \varphi, \{\mathcal{R}_a + \varphi\}_{a \in \mathsf{Agt}}, \mathcal{V} + \varphi \rangle$ where:

$\mathcal{W} + \varphi = \{w \in \mathcal{W} : [\![\varphi]\!]^w_{\mathcal{M}} = T\}$
$\mathcal{R}_a + \varphi$ is the restriction of $\mathcal{R}_a$ to $\mathcal{W} + \varphi$
$\mathcal{V} + \varphi$ is the restriction of $\mathcal{V}$ to $\mathcal{W} + \varphi$

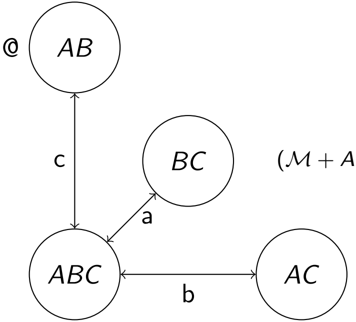


@
$B$
$AB$
$A$
$BC$
$C$
$ABC$
$AC$
a
b
c
$\mathcal{M} + A \vee B \vee C$

She then asks each of the children to step forward if they know whether they have mud on their forehead.

Since $[\![\neg \mathsf{K}_a A \wedge \neg \mathsf{K}_a \neg A]\!]^{@}_{\mathcal{M}+A\vee B\vee C} = T$, $[\![\neg \mathsf{K}_b B \wedge \neg \mathsf{K}_b \neg B]\!]^{@}_{\mathcal{M}+A\vee B\vee C} = T$, and $[\![\neg \mathsf{K}_c C \wedge \neg \mathsf{K}_c \neg C]\!]^{@}_{\mathcal{M}+A\vee B\vee C} = T$, none of the children step forward.

This provides new information.




$(\mathcal{M} + A \vee B \vee C) + \neg \mathsf{K}_a A \wedge \neg \mathsf{K}_a \neg A \ldots$

The mother again asks each of the children to step forward if they know whether they are dirty.

Since $[\![\mathsf{K}_a A]\!]^{@}_{\mathcal{M}+\dots} = [\![\mathsf{K}_b B]\!]^{@}_{\mathcal{M}+\dots} = T$, $a$ and $b$ step forward.

But since $[\![\neg\mathsf{K}_c C \wedge \neg\mathsf{K}_c\neg C]\!]^{@}_{\mathcal{M}+\dots} = T$, $c$ does not.

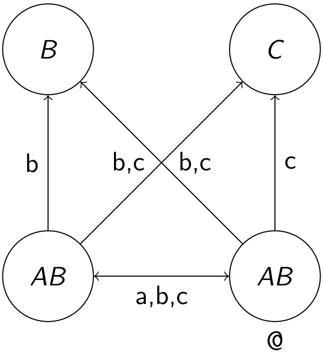
Again, this provides new information.

@ $AB$

In fact, only @ remains open after this last update.

Next time the mother asks her question, $c$ steps forward as well.

**Exercise 2.5.1**




Are the following sentences true or false at @ in this model? Justify your answers by appealing to the semantics of collective knowledge operators.

$\mathsf{S}_{\{a,b,c\}}A \wedge B$

$\mathsf{C}_{\{a,b\}}A$

$\mathsf{D}_{\{b,c\}}B$

**Exercise 2.5.2**

Agents *a*, *b*, and *c* are sitting around a table on which the Jack of spades lies face down.

The agents each know that the card is one of the four Jacks but they do not know which one. Draw an epistemic model that represents the knowledge in this initial situation.

Now suppose that agent *a* picks up the card and (without showing it to the other agents) looks at the Jack of spades. He then returns the card face down to the table. Draw an epistemic model that represents the knowledge in this situation.

Now suppose that agent *a* announces that he will tell agent *b* the color of the card. He then pulls *b* aside and tells him that the card is black. Agent *c* sees this all happening but remains in the dark about the color of the card. Draw an epistemic model that represents the knowledge in this situation.

**Exercise 2.5.3**

Describe what will happen in these variations on the Muddy Children Puzzle. Draw sequences of epistemic models to represent how the knowledge state of each child changes as new information comes to light.

The mother first announces that at least one of the children is clean and then proceeds, as before, to repeatedly ask each of the children to step forward if they know whether they have mud on their forehead.

The mother first announces that child *a* is dirty and then proceeds, as before, to repeatedly ask each of the children to step forward if they know whether they have mud on their forehead.

# Philosophical Logic

## 3.1 Tarski on Truth

Johns Hopkins University, Spring 2015

# Introduction

Tarski's goal in 'The Concept of Truth in Formalized Languages' [1936]:

"*to construct*—with reference to a given language—*a materially adequate and formally correct definition of the term 'true sentence'*." (p. 152)

Materially adequate: Spelled out in Convention T.

Formally correct: The definition does not make use of any terms whose sense is not entirely clear or that presupposes the notion of truth itself.

Tarski thinks that the classical *correspondence* conception of truth is quite clear and intelligible. But attempts to define the term 'true sentence' more precisely have been fruitless and often led to paradox.

The extension of 'true sentence' depends on the particular language under consideration. So Tarski considers whether we can define truth for different kinds of languages.

# Colloquial Language

Beginning with everyday or colloquial language, he first pursues a *semantical definition*:

**Def 3.1.1.** A true sentence is one which says that the state of affairs is so and so, and the state of affairs indeed is so and so.

Compare Aristotle: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or of what is not that it is not, is true."

Tarski thinks that Def 3.1.1 is intelligible but imprecise.

# Colloquial Language

Sharpening:

- $x$ is a true sentence if and only if $p$

where we substitute in the place of '$p$' a sentence and in the place of '$x$' a name for this sentence.

Using *quotation-mark names*:

- 'It is snowing' is a true sentence if and only if it is snowing.

Using *structural-descriptive names*:

- An expression consisting of three words of which the first is composed of the two letters I and Te (in that order) the second of the two letters I and Es (in that order) and the third of the seven letters Es, En, O, Double-U, I, En, and Ge (in that order), is a true sentence if and only if it is snowing.

# Colloquial Language

Can we define truth by providing sentences of this kind?

**Problem:** The Liar.

Let '$c$' abbreviate 'The first declarative sentence on this slide'.

*$c$ is not a true sentence.*

Empirical fact:

($\alpha$) '$c$ is not a true sentence' is identical with $c$.

Instance of (2):

($\beta$) '$c$ is not a true sentence' is a true sentence if and only if $c$ is not a true sentence.

By applying Identity Elimination, we obtain this equivalence:

$c$ is a true sentence if and only if $c$ is not a true sentence.

# Colloquial Language

**Problem:** In some situations, we cannot even indicate for a given name the sentence denoted by this name.

The first sentence that will be printed in 3000 is a true sentence if and only ???

N. B. This example modifies one given by Tarski. If you worry that the definite description 'The first sentence that will be printed in 3000' is not a *name*, replace this with a co-denoting proper name of your choosing.

# Colloquial Language

Suppose that we ban sentences involving the expression ‘true sentence’ from the substitution class for ‘$p$’. Can we then define truth for colloquial language?

Next attempt:

**Def 3.1.2.** For all $p$, ‘$p$’ is a true sentence if and only if $p$.

# Colloquial Language

**Problem:** Sentences can have names besides their quotation-mark name.

**Problem:** The year 3000 worry remains.

# Colloquial Language

Next attempt:

**Def 3.1.3.** For all $x$, $x$ is a true sentence if and only if, for a certain $p$, $x$ is identical with '$p$' and $p$.

This definition accommodates the fact that sentences can have multiple names.

This definition also accommodates the fact that we cannot always indicate for a given name the sentence denoted by this name.

Tarski worries about the quantification into quotation in Def 3.1.3 but he suggests that we treat "$p$" as a *quotation-function* mapping sentences to quotation-names of those sentences.

N. B. Recall that $\Sigma$ allows us to quantify into quotes.

# Colloquial Language

**Problem**: Quotation-functions lead to paradox.

N. B. Alternatively, substitutional quantification leads to paradox.

Let '$c$' abbreviate 'The first declarative sentence on this slide'.

*For all p, if c is identical with the sentence 'p', then it is not the case that p.*

Empirical fact:

($\alpha$) 'For all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$' is identical with $c$.

General principle:

($\beta$) For all $p$ and $q$, if '$p$' is identical with '$q$', then $p$ if and only if $q$.

This leads to trouble.

# Colloquial Language

Assume: For all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$.

Instantiating this with $c$ itself: If $c$ is identical with the sentence 'For all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$', then it is not the case that for all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$.

Given $(\alpha)$: It is not the case that for all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$.

Contradiction.

# Colloquial Language

So: For some $p$, $c$ is identical with '$p$' and $p$.

Let $p^*$ witness this existential: $c$ is identical with '$p^*$' and $p^*$.

Given $(\alpha)$: 'For all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$' is identical with '$p^*$'.

Given $(\beta)$: $p^*$ if and only for all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$.

Given $p^*$: For all $p$, if $c$ is identical with the sentence '$p$', then it is not the case that $p$.

Contradiction.

# Colloquial Language

Upshot: “Our discussions so far entitle us in any case to say that *the attempt to construct a correct semantical definition of the expression ‘true sentence’ meets with very real difficulties*.” (p. 162)

Tarski next considers the possibility of giving a *structural definition* of truth for colloquial language along these lines:

**Def 3.1.4.** $x$ is a true sentence if and only if $x$ has certain structural properties or can be obtained from such and such structurally described expressions by means of such and such structural transformations.

**Example:** If a true sentence consists of four parts, of which the first is the word ‘if’, the second is a true sentence, the third is the word ‘then’, then the fourth part is a true sentence.

But Tarski claims that this approach is hopeless for natural language since natural language is “not something finished, closed, or bounded by clear limits”. (p. 164)

# Colloquial Language

In any case, Tarski thinks that the original Liar paradox is devastating for the prospect of defining truth for colloquial language.

A characteristic feature of natural language is its *universality*:

"It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about anything at all, we can also speak about it in colloquial language'." (p. 164)

As a result, we cannot block the Liar.

# Colloquial Language

Conclusion:

"No consistent language can exist for which the usual laws of logic hold and which at the same time satisfies the following conditions:

(I) for any sentence which occurs in the language a definite name of this sentence also belongs to this language

(II) every expression formed from (2) by replacing the symbol '$p$' by any sentence of the language and the symbol '$x$' by a name of this sentence is to be regarded as a true sentence of this language

(III) in the language in question an empirically established premiss having the same meaning as $(\alpha)$ can be formulated and accepted as a true sentence." (p. 165)

# Formal Languages: The Positive

Though Tarski argues that truth is undefinable for colloquial language, an important positive part of his project is to show how to define truth for many formal languages.

*object language:* the language under investigation.

*metalanguage:* the language in which this investigation (or *metatheory*) is carried out.

Given the universality of colloquial language, the distinction between object language and metalanguage collapses. But formal languages are not universal. So the prospect of defining truth for these languages in a stronger metalanguage is not hopeless.

# Formal Languages: The Positive

To start off, Tarski considers the *language of the calculus of classes*.

The lexicon of this object language consists of the following 'constants' and variables:

- Negation sign '$N$'
- Disjunction sign '$A$'
- Universal quantifier '$\Pi$'
- Inclusion sign '$I$'
- Variables '$x_{,}$', '$x_{,,}$', '$x_{,,,}$', ...

Sample sentential functions: '$NIx_{,}x_{,,}$', '$\Pi x_{,}\Pi x_{,,}AIx_{,}x_{,,}Ix_{,,}x_{,}$'.

The calculus of classes itself consists of axioms of sentential logic and class-theoretic axioms involving the inclusion sign '$I$'.

# Formal Languages: The Positive

The metalanguage has two kinds of expressions.

First, the metalanguage has expressions of a general logical character:

- ‘not’, ‘or’, ‘for all’, ‘$\subseteq$’
- Expressions of sentential calculus: ‘and’, ‘if...then...’, ...
- Expressions of first-order calculus: ‘for some $x$’, ‘is identical with’, ...
- Further class-theoretic expressions: ‘is an element of’, ‘null class’, ‘infinite class’, ‘power of a class’, ‘ordered pair’, ‘sequence’, ...
- Arithmetic expressions: ‘0’, ‘$+$’, ‘$>$’, ...

Note that every sentence in the object language can be translated into the metalanguage.

**Example:** ‘$\Pi x_{\prime} I x_{\prime} x_{\prime}$’ gets translated as ‘for all $a$, $a \subseteq a$’.

# Formal Languages: The Positive

Second, the metalanguage has these structural-descriptive expressions:

- 'the negation sign' (abbr: '*ng*')
- 'the sign of logical sum' (abbr: '*sm*')
- 'the sign of the universal quantifier' (abbr: '*un*')
- 'the inclusion sign' (abbr: '*in*')
- 'the $k$th variable' (abbr: '$v_k$')
- 'the expression consisting of $x$ followed by $y$' (abbr: '$x \frown y$')
- 'expression'

Note that every expression in the object language has a structural-descriptive name in the metalanguage.

**Example:** $(((un \frown v_1) \frown in) \frown v_1) \frown v_1$ names '$\Pi x_{\prime} I x_{\prime} x_{\prime}$'.

# Formal Languages: The Positive

Time to define truth.

**Convention T.** A formally correct definition of the symbol '$Tr$', formulated in the metalanguage, will be called an adequate definition of truth if it has the following consequences:

($\alpha$) all sentence which are obtained from the expression '$x \in Tr$ if and only if $p$' by substituting for the symbol '$x$' a structural-descriptive name of any sentence of the language in question and for the symbol '$p$' the expression which forms the translation of this sentence into the metalanguage

($\beta$) the sentence 'for any $x$, if $x \in Tr$ then [$x$ is a sentence]' (p. 187-8)

That is, our definition of '$Tr$' should imply *T-sentences* like this:

$(((un \frown v_1) \frown in) \frown v_1) \frown v_1 \in Tr$ iff for all $a$, $a \subseteq a$.

# Formal Languages: The Positive

If the object language had only finitely many sentences, then defining truth would be easy.

**Def 3.1.5.** $x \in Tr$ iff $(x = x_1 \wedge p_1) \vee ...(x = x_n \wedge p_n)$.

But the language of the calculus of classes has infinitely many sentences and the metalanguage does not allow sentences of infinite length.

We must define truth in some other way.

# Formal Languages: The Positive

The metatheory itself has both general logical axioms and specific axioms having to do with expressions of the object language (p. 173-4).

**Def 3.1.6.** $x = \iota_{k,l}$ iff $x = (in \frown v_k) \frown v_l$.

**Def 3.1.7.** $x = \overline{y}$ iff $x = ng \frown y$.

**Def 3.1.8.** $x = y + z$ iff $x = (sm \frown y) \frown z$.

**Def 3.1.9.** $x = \bigcap_k y$ iff $x = (un \frown v_k) \frown y$.

# Formal Languages: The Positive

**Def 3.1.10.** $x$ is a *sentential function* iff $x$ is an expression which satisfies one of the following four conditions:

($i$) $x = \iota_{k,l}$ where $k, l \in \mathbb{N}$

($ii$) $x = \overline{y}$ where $y$ is a sentential function

($iii$) $x = y + z$ where $y$ and $z$ are sentential functions

($iv$) $x = \bigcap_k y$ where $k \in \mathbb{N}$ and $y$ is a sentential function

# Formal Languages: The Positive

**Def 3.1.11.** $v_k$ is a *free variable* of the sentential function $x$ iff $k \in \mathbb{N}$ and $x$ is a sentential function satisfying one of the following four conditions:

$(i)$ $x = \iota_{k,l}$ or $x = \iota_{l,k}$ where $l \in \mathbb{N}$

$(ii)$ $x = \overline{y}$ where $y$ is a sentential function and $v_k$ is free in $y$

$(iii)$ $x = y + z$ or $x = z + y$ where $y$ and $z$ are sentential functions and $v_k$ is free in $y$

$(iv)$ $x = \bigcap_l y$ where $l \in \mathbb{N}$, $l \neq k$, $y$ is a sentential function, and $v_k$ is free in $y$

**Def 3.1.12.** $x$ is a *sentence* $(x \in S)$ iff $x$ is a sentential function with no free variables.

# Formal Languages: The Positive

To define truth, Tarski first defines the more general concept of *satisfaction* of a sentential function by given objects.

**Def 3.1.13.** The sequence $f$ *satisfies* the sentential function $x$ iff $f$ is an infinite sequence of classes and $x$ is a sentential function and these meet one of the following four conditions:

($i$) $x = \iota_{k,l}$ and $f_k \subseteq f_l$

($ii$) $x = \overline{y}$ and $f$ does not satisfy $y$

($iii$) $x = y + z$ and $f$ satisfies $y$ or satisfies $z$

($iv$) $x = \bigcap_k y$ and any infinite sequence of classes $g$ which differs from $f$ in at most the $k$-th place satisfies $y$

N. B. Recall our modern recursive definition of truth relative to an assignment function.

# Formal Languages: The Positive

**Example:**

The sequence $f$ satisfies $\bigcap_2 \iota_{1,2}$ iff...

any infinite sequence of classes $g$ which differs from $f$ in at most the 2nd place satisfies $\iota_{1,2}$ iff...

$f_1 \subseteq g_2$ for any infinite sequence of classes $g$ iff...

$f_1$ is the null class.

# Formal Languages: The Positive

**Def 3.1.14.** $x \in Tr$ iff $x \in S$ and every infinite sequence of classes satisfies $x$.

"*We have succeeded in doing for the language of the calculus of classes what we have tried in vain to do for colloquial language: namely to construct a formally correct and materially adequate semantical definition of the expression 'true sentence'.*" (p. 208-9)

# Formal Languages: The Positive

**Thm 3.1.1 (Principle of Contradiction).** For all $x \in S$, $x \notin Tr$ or $\overline{x} \notin Tr$.

**Thm 3.1.2 (Principle of Excluded Middle).** For all $x \in S$, $x \in Tr$ or $\overline{x} \in Tr$.

Tarski also defines related notions of *truth in an individual domain a*, *truth in an individual domain with k elements*, and *truth in every individual domain*.

*Truth simpliciter* is truth in the domain consisting of all classes.

Note that for Tarski, the object language is already interpreted.

# Formal Languages: The Positive

Tarski also sketches how to define truth for other formal languages.

**Def 3.1.15.** Two expressions belong to the same *semantical category* if the following two conditions hold:

($i$) there is a sentential function which contains one of these expressions

($ii$) no sentential function which contains one of these expressions ceases to be a sentential function if this expression is replaced in it by the other

**Examples:** sentential functions, names of individuals, names of two-termed relations between individuals.

# Formal Languages: The Positive

Every semantical category has an *order* $k \in \mathbb{N}$.

Categories of 1st order: names of individuals, variables ranging over individuals.

Categories of 2nd order: names of classes of individuals, variables ranging over classes of individuals, names of two-, three-, and many-termed relations between individuals, ...

And so forth.

# Formal Languages: The Positive

Focusing just on variables, formal languages can be divided into four categories:

**Type I:** All variables belong to the same semantical category.

**Type II:** $n > 1$ semantical categories of variables.

**Type III:** Infinitely many categories but all have order $\leq n$.

**Type IV:** Infinitely many categories of arbitrarily high order.

Languages of Type I-III are *languages of finite order*.

Languages of Type IV are *languages of infinite order*.

# Formal Languages: The Positive

**Example of Type I:** The language of the calculus of classes.

**Example of Type II:** The language of two-termed relations.

'$x_{\prime}$', '$x_{\prime\prime}$', '$x_{\prime\prime\prime}$', ... range over individuals.

'$X_{\prime}$', '$X_{\prime\prime}$', '$X_{\prime\prime\prime}$', ... range over binary relations between individuals.

**Example of Type III:** The language of many-termed relations.

'$x_{\prime}$', '$x_{\prime\prime}$', '$x_{\prime\prime\prime}$', ... range over individuals.

'$X'_{\prime}$', '$X'_{\prime\prime}$', '$X'_{\prime\prime\prime}$', ... range over 1-ary relations (classes).

'$X''_{\prime}$', '$X''_{\prime\prime}$', '$X''_{\prime\prime\prime}$', ... range over 2-ary relations.

And so forth.

# Formal Languages: The Positive

Tarski sketches how to define truth for languages of finite order.

For languages of Type I and II, we can use the *method of many-rowed sequences* and define satisfaction as an *n*-termed relation holding between sentential functions and sequences of objects of different categories.

For all finite order languages, we can use the *method of semantical unification of the variables* and assimilate the objects that all of the variables range over to objects that fall under the range of variables of a single semantical category (this *unifying category* cannot be of lower order than the semantical category of any of the variables in the language).

But neither approach works for languages of infinite order—assuming that we cannot use variables of infinite order. In the Postscript, Tarski allows for such variables.

# Formal Languages: The Negative

**Example of Type IV:** The language of the general theory of classes.

'$X^{\prime}_{\prime}$', '$X^{\prime}_{\prime\prime}$', '$X^{\prime}_{\prime\prime\prime}$', ... range over individuals.

'$X^{\prime\prime}_{\prime}$', '$X^{\prime\prime}_{\prime\prime}$', '$X^{\prime\prime}_{\prime\prime\prime}$', ... range over classes of individuals.

'$X^{\prime\prime\prime}_{\prime}$', '$X^{\prime\prime\prime}_{\prime\prime}$', '$X^{\prime\prime\prime}_{\prime\prime\prime}$', ... range over classes of classes of individuals.

And so forth.

Atomic sentential functions are of the form '$XY$' where in the place of '$X$' and '$Y$' variables of the $n+1$-st order and $n$-th order can be substituted respectively.

Can we define truth for this language in some other way?

# Formal Languages: The Negative

**Thm 3.1.3.** In whatever way the symbol '*Tr*', denoting a class of expressions, is defined in the metatheory, it will be possible to derive from it the negation of one of the sentences which were described in the conditions ($\alpha$) in Convention T. So, assuming that the metatheory is consistent, it is impossible to construct an adequate definition of truth in the sense of Convention T on the basis of the metatheory.

The proof uses techniques from Gödel. See p. 247–251 for more details.

Nice gloss: "The semantics of the language could not be established as a part of its morphology." (p. 254)

# Formal Languages: The Negative

Despite the undefinability of truth in the metalanguage, Tarski suggests that we can still take an *axiomatic* approach to truth by introducing an undefined truth predicate '*Tr*' and establishing its properties with axioms.

Indeed, we can consistently adopt every equivalence in condition ($\alpha$) in Convention T as an axiom.

# Formal Languages: The Negative

**Thm 3.1.4.** For arbitrary $k \in \mathbb{N}$, we can define '*Tr*' in the metalanguage which has among its consequences every equivalence in condition $(\alpha)$ in Convention T in which in the place of '*p*' sentences with variables of at most the *k*-th order occur.

Since the fragment of the object language containing variables of at most the *k*-th order can be thought of as a language of finite order, this theorem follows from Tarski's investigation of such languages.

**Thm 3.1.5.** If the metatheory is consistent, then we can add every equivalence in condition $(\alpha)$ in Convention T as an axiom and maintain consistency.

This follows from the previous theorem and Compactness.

# Conclusion

Since Tarski's results can be extended to other semantical concepts, he concludes (p. 266):

- *The semantics of any formalized language of finite order can be built up as part of the morphology of language, based on correspondingly constructed definitions.*
- *It is impossible to establish the semantics of the formalized languages of infinite order in this way.*
- *But the semantics of any formalized language of infinite order can be established as an independent science based upon its own primitive concepts and its own axioms, possessing as its logical foundation a system of the morphology of language.*

# Conclusion

A final remark on colloquial language (p. 267):

"The concept of truth (as well as other semantical concepts) when applied to colloquial language in conjunction with the normal laws of logic leads inevitably to confusions and contradictions. Whoever wishes, in spite of all difficulties, to pursue the semantics of colloquial language with the help of exact methods will be driven first to undertake the thankless task of a reform of this language. He will find it necessary to define its structure, to overcome the ambiguity of the terms which occur in it, and finally to split the language into a series of languages of greater and greater extent, each of which stands in the same relation to the next in which a formalized language stands to its metalanguage."

So we end with the idea of a *Tarskian hierarchy* of languages.

**Exercise 3.1.1**

Write out the T-sentences in the metalanguage for the following sentences in the language of the calculus of classes:

‘$N\Pi x_{\prime} N I x_{\prime} x_{\prime}$’

‘$\Pi x_{\prime} \Pi x_{\prime\prime} A I x_{\prime} x_{\prime\prime} I x_{\prime\prime} x_{\prime}$’

**Exercise 3.1.2**

The sequence $f$ satisfies $\overline{\iota_{2,3}} + \overline{\iota_{3,2}}$ iff ____________.

The sequence $f$ satisfies $\bigcap_1 \bigcap_2 (\iota_{1,2} + \iota_{2,1})$ iff ____________.

Fill in the blanks by recursively applying Def 3.1.13.

# Philosophical Logic

## 3.2 Kripke on Truth

Johns Hopkins University, Spring 2015

# Semantic Paradoxes

Once you start looking for them, semantic paradoxes crop up everywhere.

Jones:

(1) Most of Nixons assertions about Watergate are false.

Nixon:

(2) Everything Jones says about Watergate is true.

Suppose that (1) is the only thing that Jones says about Watergate and Nixon's other assertions about Watergate are evenly balanced between the true and the false.

Then both (1) and (2) are paradoxical:

(1) is true $\Rightarrow$ (2) is false $\Rightarrow$ (1) is false $\Rightarrow$ (2) is true $\Rightarrow$ (1) is true.

# Semantic Paradoxes

Morals:

"The versions of the Liar paradox which use empirical predicates already point up one major aspect of the problem: *many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features*." (p. 691)

"An adequate theory must allow our statements involving the notion of truth to be *risky*: they risk being paradoxical if the empirical facts are extremely (and unexpectedly) unfavorable. There can be no syntactic or semantic 'sieve' that will winnow out the 'bad' cases while preserving the 'good' ones." (p. 692)

Though Gödel showed us how to generate paradox using syntactic predicates, the paradoxes involving empirical predicates reveal just how risky our truth talk is. There is no *intrinsic* criterion for paradoxicality.

# Semantic Paradoxes

Though not paradoxical, sentences like the following are also problematic:

(3) The first numbered sentence on this slide is true.

This sentence is *ungrounded*:

(3) is true $\Rightarrow$ (3) is true $\Rightarrow$ (3) is true ...

Since the truth conditions for (3) involve truth ‘all the way down’, we cannot ascertain the truth value of this sentence.

Examples like this show that ungroundedness, like paradoxicality, can be an empirical matter. There is no syntactic or semantic criterion for ungroundedness either.

# Tarski's Hierarchy

According to Kripke, the *orthodox approach* to the semantic paradoxes is the Tarskian hierarchy of *typed* truth predicates.

$\mathcal{L}_0$: formal first-order language that can discuss its own syntax.

$\mathcal{L}_1$: $\mathcal{L}_0$ along with truth predicate $T_1(x)$ for $\mathcal{L}_0$.

$\mathcal{L}_2$: $\mathcal{L}_1$ along with truth predicate $T_2(x)$ for $\mathcal{L}_1$.

And so forth.

Of course, natural language contains just one word 'true', not a sequence of distinct expressions 'true$_1$', 'true$_2$', 'true$_3$', ... But the idea is that our ordinary truth talk is systematically ambiguous. The 'level' of a truth predicate is determined by the context of use and speaker intentions.

# Tarski's Hierarchy

**Problem:** The Tarskian approach is unfaithful to the facts. When Jones and Nixon utter (1) and (2) respectively, they do not attach subscripts, explicit or implicit, to 'true' and 'false'.

Indeed, if Jones is not aware of everything Nixon has said about Watergate, then he does not know the levels of Nixon's relevant utterances. If Jones is forced to assign a level in advance, he will not know how high a level to choose.

Kripke's point is not that (1) and (2) do not have levels, but rather that the levels of these sentences are not determined by syntax, semantics, or speaker intentions.

"In some sense a statement should be allowed to seek its own level, high enough to say what it intends to say. It should not have an intrinsic level fixed in advance, as in the Tarski hierarchy." (p. 696)

# Tarski's Hierarchy

**Problem:** It is difficult for the Tarskian to respect linguistic intuitions.

Dean:

(4) All of Nixons assertions about Watergate are false.

Nixon:

(5) Everything Dean says about Watergate is false.

Suppose that Dean has said something true about Watergate other than (4) and Nixon has said something true about Watergate other than (5). Then (4) and (5) are both false. Pretty straightforward.

But things are not at all straightforward for the Tarskian. In asserting (4), Dean wants to include (5) in the scope of the universal. Similarly, in asserting (5), Nixon wants to include (4) in the scope of the universal. Both cannot succeed.

# Tarski's Hierarchy

Kripke mentions some other problems with the Tarskian approach, including some technical difficulties that crop up when we consider transfinite levels.

Given all of these problems, Kripke outlines an alternative theory of truth.

This theory involves a single untyped truth predicate but allows for *truth-value gaps*. In particular, Liar sentences are neither true nor false.

# Kripke's Fixed Point

Disclaimer:

"I do not regard any proposal, including the one to be advanced here, as definitive in the sense that it gives *the* interpretation of the ordinary use of 'true', or *the* solution to the semantic paradoxes. On the contrary, I have not at the moment thought through a careful philosophical justification of the proposal, nor am I sure of the exact areas and limitations of its applicability. I do hope that the model given here has two virtues: first, that it provides an area rich in formal structure and mathematical properties; second, that to a reasonable extent these properties capture important intuitions." (p. 699)

# Kripke's Fixed Point

Preliminaries:

We will be working with interpreted first-order languages that have a finite list of 1-place predicate symbols. These languages are rich enough to express their own syntax.

Kripke allows for predicates to have only partial interpretations:

**Def 3.2.1.** A *model* $\mathcal{M} = \langle \mathcal{D}, \mathcal{I} \rangle$ consists of a nonempty domain of objects $\mathcal{D}$ and an interpretation function $\mathcal{I}$ mapping each monadic predicate $P$ to an ordered pair $\langle S_1, S_2 \rangle$ where $S_1 \subseteq \mathcal{D}$, $S_2 \subseteq \mathcal{D}$, and $S_1 \cap S_2 = \emptyset$.

$S_1$ is the *extension* of $P$.

$S_2$ is the *anti-extension* of $P$.

$P$ is *true* of objects in $S_1$, *false* of object in $S_2$, and *undefined* otherwise.

# Kripke's Fixed Point

**Def 3.2.2.** The logical connectives are handled with *Kleene's Strong 3-valued logic*:

| $\neg$ | |
|---|---|
| T | F |
| N | N |
| F | T |

| $\wedge$ | T | N | F |
|---|---|---|---|
| T | T | N | F |
| N | N | N | F |
| F | F | F | F |

| $\vee$ | T | N | F |
|---|---|---|---|
| T | T | T | T |
| N | T | N | N |
| F | T | N | F |

| $\supset$ | T | N | F |
|---|---|---|---|
| T | T | N | F |
| N | T | N | N |
| F | T | T | T |

| $\equiv$ | T | N | F |
|---|---|---|---|
| T | T | N | F |
| N | N | N | N |
| F | F | N | T |

The semantics for $\forall$ and $\exists$ generalizes that for $\wedge$ and $\vee$ respectively.

The term '3-valued logic' is a bit unfortunate. 'N' is not an extra truth value but rather the absence of a truth value.

# Kripke's Fixed Point

Begin with a language $L$ whose monadic predicates are fully interpreted.

Add the monadic predicate $T(x)$ to $L$. The language $\mathcal{L}\langle S_1, S_2\rangle$ results from interpreting $T(x)$ as $\langle S_1, S_2\rangle$ and the other predicates as before.

$S_1' = \{d \in \mathcal{D} : d$ codes a true sentence in $\mathcal{L}\langle S_1, S_2\rangle\}$.

$S_2' = \{d \in \mathcal{D} : d$ codes a false sentence in $\mathcal{L}\langle S_1, S_2\rangle$ or fails to code a sentence at all$\}$.

**Def 3.2.3.** $\phi(\langle S_1, S_2\rangle) = \langle S_1', S_2'\rangle$.

**Def 3.2.4.** $\langle S_1, S_2\rangle$ is a *fixed point* iff $\phi(\langle S_1, S_2\rangle) = \langle S_1, S_2\rangle$.

If $T(x)$ is a truth predicate for $\mathcal{L}\langle S_1, S_2\rangle$, then $\langle S_1, S_2\rangle$ is a fixed point.

# Kripke's Fixed Point

Here is the actual fixed point construction:

$\mathcal{L}_0 = \mathcal{L}\langle S_{1,0}, S_{2,0}\rangle = \mathcal{L}\langle\emptyset, \emptyset\rangle$. ($T(x)$ is completely undefined.)

If $\alpha + 1$ is a successor ordinal, $\mathcal{L}_{\alpha+1} = \mathcal{L}\phi(\langle S_{1,\alpha}, S_{2,\alpha}\rangle)$.

If $\lambda$ is a limit ordinal, $\mathcal{L}_\lambda = \mathcal{L}\langle\bigcup_{\beta<\lambda} S_{1,\beta}, \bigcup_{\beta<\lambda} S_{2,\beta}\rangle$.

# Kripke's Fixed Point

**Def 3.2.5.** $\langle S_1^*, S_2^* \rangle$ *extends* $\langle S_1, S_2 \rangle$ ($\langle S_1^*, S_2^* \rangle \geq \langle S_1, S_2 \rangle$) iff $S_1 \subseteq S_1^*$ and $S_2 \subseteq S_2^*$.

**Lem 3.2.1.** $\phi$ is a *monotone (order-preserving) operation* on $\leq$: if $\langle S_1^*, S_2^* \rangle \geq \langle S_1, S_2 \rangle$ then $\phi(\langle S_1^*, S_2^* \rangle) \geq \phi(\langle S_1, S_2 \rangle)$.

That is, if the interpretation of $T(x)$ is extended, no truth values previously established will change or become undefined.

**Lem 3.2.2.** For any ordinals $\alpha$ and $\beta$, if $\alpha > \beta$ then $\langle S_{1,\alpha}, S_{2,\alpha} \rangle \geq \langle S_{1,\beta}, S_{2,\beta} \rangle$.

Since neither the extension nor anti-extension of $T(x)$ ever decreases as $\alpha$ increases, we must eventually reach a *minimal* fixed point; there is an ordinal level $\sigma$ such that $\phi(\langle S_{1,\sigma}, S_{2,\sigma} \rangle) = \langle S_{1,\sigma}, S_{2,\sigma} \rangle$.

While this resembles the Tarskian hierarchy, Kripke's construction involves a single growing predicate $T(x)$. The languages in the hierarchy are not the primary objects of interest, but are better and better approximations to the minimal language with its own truth predicate.

# Kripke's Fixed Point

**Def 3.2.6.** $\varphi$ is *grounded* if it has a truth value in the minimal fixed point language $\mathcal{L}\langle S_{1,\sigma}, S_{2,\sigma}\rangle$; otherwise $\varphi$ is *ungrounded*.

Both Liar sentences and intuitively 'ungrounded' sentences like (3) are ungrounded in Kripke's model.

**Def 3.2.7.** If $\varphi$ is grounded, then the *level* of $\varphi$ is the smallest ordinal $\alpha$ such that $\varphi$ has a truth value in $\mathcal{L}\langle S_{1,\alpha}, S_{2,\alpha}\rangle$.

Sentences seek their own level. Consider $\forall x(P(x) \supset T(x))$ and suppose that the extension of the empirical predicate $P(x)$ consists entirely of grounded true sentences of levels 2, 4, and 13. Then this sentence will be grounded with level 14.

Concluding Remarks:

"The present approach certainly does not claim to give a universal language, and I doubt that such a goal can be achieved. First, the induction defining the minimal fixed point is carried out in a set-theoretic metalanguage, not in the object language itself. Second, there are assertions we can make about the object language which we cannot make in the object language. For example, Liar sentences are *not true* in the object language, in the sense that the inductive process never makes them true; but we are precluded from saying this in the object language by our interpretation of negation and the truth predicate. If we think of the minimal fixed point, say under the Kleene valuation, as giving a model of natural language, then the sense in which we can say, in natural language, that a Liar sentence is not true must be thought of as associated with some later state in the development of natural language, one in which speakers reflect on the generation process leading to the minimal fixed point. It is not itself a part of that process. The necessity to ascend to a metalanguage may be one of the weaknesses of the present theory. The ghost of the Tarski hierarchy is still with us." (p. 714)

"On the basis of the fact that the goal of a universal language seems elusive, some have concluded that truth-gap approaches, or any approaches that attempt to come closer to natural language than does the orthodox approach, are fruitless. I hope that the fertility of the present approach, and its agreement with intuitions about natural language in a large number of instances, cast doubt upon such negative attitudes." (p. 715)

## Exercise 3.2.1

Suppose that the intrepretation of $T(x)$ is $\langle S_1, S_2 \rangle$ where $\langle S_1, S_2 \rangle$ is *not* a fixed point of $\varphi$; that is, $\varphi(\langle S_1, S_2 \rangle) \neq \langle S_1, S_2 \rangle$. Briefly explain why $T(x)$ is ill-suited to serve as a truth predicate for the language $\mathcal{L}(\langle S_1, S_2 \rangle)$.

**Exercise 3.2.2**

Prove Lem 3.2.2 via transfinite induction by establishing the following:

- $\langle S_{1,1}, S_{2,1}\rangle \geq \langle S_{1,0}, S_{2,0}\rangle$.
- For any successor ordinal $\alpha + 1$, if $\langle S_{1,\alpha}, S_{2,\alpha}\rangle \geq \langle S_{1,\beta}, S_{2,\beta}\rangle$ for all $\beta < \alpha$, then $\langle S_{1,\alpha+1}, S_{2,\alpha+1}\rangle \geq \langle S_{1,\alpha}, S_{2,\alpha}\rangle$.
- For any limit ordinal $\lambda$ and $\beta < \lambda$, $\langle S_{1,\lambda}, S_{2,\lambda}\rangle \geq \langle S_{1,\beta}, S_{2,\beta}\rangle$.

**Exercise 3.2.3**

Suppose that $\varphi$ is a true sentence with level 5, $\psi$ is a false sentence with level 8, and $\xi$ is ungrounded.

Let $\ulcorner\varphi\urcorner \in \mathcal{D}$ designate the code of $\varphi$.

Are the following sentences grounded? What is the truth value and level of each grounded sentence?

$T(\ulcorner\varphi \supset \psi\urcorner)$

$(\varphi \vee \psi) \supset \xi$

$\xi \supset T(\ulcorner T(\ulcorner\varphi\urcorner)\urcorner)$

# Philosophical Logic

## 3.3 The Space of Truth Theories

Johns Hopkins University, Spring 2015

In his useful survey article, Leitgeb [2007] lists many desiderata for theories of truth.

Unfortunately, no truth theory can meet them all. So Leitgeb considers theories that meet maximal sets of these desiderata.

**Desiderata (a):** Truth should be expressed by a predicate.

If truth is a predicate, then we can concatenate it with proper names, definite descriptions, or variables ranging over names for sentences to form constructions like these:

(1) $Tr$('$2+2=4$').

(2) $Tr$(the $x$ such that $x$ is the last sentence spoken by Caesar).

(3) For all $x$ and $y$, if $y$ is the negation of $x$, then $Tr(y)$ iff not $Tr(x)$.

If truth is captured using the operator 'it is true that', then Leitgeb thinks that constructions like (2) and (3) are problematic. What if we do not know the last sentence spoken by Caesar?

N. B. We can formalize (3) using a truth operator and substitutional quantifiers.

**Desiderata (b):** If a theory of truth is added to mathematical or empirical theories, it should be possible to prove the latter true.

For example, suppose that we add a truth theory (TT) on top of Peano Arithmetic (PA).

The following sentence is a theorem of PA:

(4) $\forall x \forall y (x + y = y + x)$.

So the following sentence should be a theorem of PA + TT:

(5) $Tr(`\forall x \forall y (x + y = y + x)')$.

In fact, the following general sentence should be a theorem:

(6) $\forall x$(If $x$ is provable from PA, then $Tr(x)$).

N. B. Given a few additional assumptions, this effectively requires that we can prove the consistency of PA in the extended theory PA + TT.

**Desiderata (c):** The truth predicate should not be subject to any type restrictions.

That is, no Tarskian hierarchy.

Leitgeb finds the arguments in Kripke [1975] convincing.

**Desiderata (d):** T-biconditionals should be derivable unrestrictedly.

This is Tarski's [1936] requirement of *material adequacy*.

Alternatively, one might demand, as Field [2006] does, that $Tr(x)$ and the denotation of $x$ are intersubstitutable in transparent (non-quotational, etc.) contexts.

Note that there is still room to disagree about the kind of conditional used in the T-biconditionals.

**Desiderata (e):** Truth should be compositional.

Leitgeb's argument for compositionally seems to be this:

Premise: Meaning should be compositional (because of learnability).

Premise: The meaning of a sentence is the condition under which this sentence is true.

Conclusion: Truth should be compositional.

N. B. Compare this with Tarski [1936] who gives a recursive definition of satisfaction (and then truth) because of the finite length restrictions in the metalanguage.

N. B. Should a disquotationalist/deflationist care about the compositionality of truth?

**Desiderata (f):** The theory should allow for standard interpretations.

"When we use singular terms in order to express the truth or falsity of sentences as in our examples [(1), (2), (3)] from above, we intend these singular terms to refer to sentences, i.e., finite sequences of signs, and we intend the truth predicate to express a property of these sequences. A theory of truth should not exclude this standard interpretation, for otherwise the theory could not be understood as speaking about the very objects that it was designed to refer to. Put differently: a theory of truth does not only have to be consistent (of course it has to be!), it also should not mess up its intended ontological commitments." (p. 281)

N. B. Will a disquotationalist/deflationist interpret a truth theory in this way?

**Desiderata (g):** The outer logic and the inner logic should coincide.

*Outer logic:* the logic in force outside the truth predicate.

*Inner logic:* the logic in force inside the truth predicate.

For example, a truth theory might have a classical outer logic and imply all tautologies while still implying a sentence $\neg Tr(x)$ where $x$ denotes a classical tautology.

"Whatever reasons there might be for preferring one logic over another, if they apply to linguistic contexts outside the applications of truth predicates, why should they not equally apply to contexts within such applications? Every discrepancy between the outer and the inner logic of a theory of truth would indicate that our calling a sentence true somehow changes the logic that governs our understanding of this sentence. This is definitely questionable. Hence such discrepancies are—ceteris paribus—to be voted out." (p. 282)

**Desiderata (h):** The outer logic should be classical.

Leitgeb gives a kind of Quinean holistic argument here:

"Suppose that two theories of truth are formally and philosophically equally useful and elegant, but while the one theory combines classical logic with a sophisticated axiom system for '*Tr*', the other theory compensates its obvious and plausible truth-theoretic axioms by deviating from classical logic in some sophisticated manner. It seems that in such a case the former theory should be preferred, if only because the principle of minimal mutilation tells us to be as conservative as possible, and fiddling with the semantic principles for '*Tr*' seems to be less 'costly' than deforming our standard understanding of 'not', 'or', and the other logical constants." (p. 283)

But no truth theory can satisfy all of these desiderata.

Liar sentences show (a) + (b) + (c) + (d) + (h) to be inconsistent.

While Tarski abandoned (c), Leitgeb takes (a) + (b) + (c) for granted.

So we must abandon (d) or (h).

If we want to hold onto classical logic, we must abandon (d).

But abandoning (d) alone is insufficient. McGee [1985] shows that (a) + (b) + (c) + (e) + (f) + (g) + (h) are inconsistent.

Kripke construed with classical outer logic and non-classical Strong Kleene inner logic, Feferman: (a) + (b) + (c) + (e) + (f) + (h).

McGee, Friedman & Sheard: (a) + (b) + (c) + (e) + (g) + (h).

Kripke with Supervaluation, Cantini: (a) + (b) + (c) + (f) + (g) + (h).

Alternatively, we can keep the full Tarskian T-schema by abandoning classical logic.

Field: (a) + (b) + (c) + (d) + (e) + (f) + (g).

**Exercise 3.3.1**

Which desiderata listed by Leitgeb do you find the most important? Argue for your preferred maximal consistent subset of $\{$(a),...,(h)$\}$ in a couple of paragraphs. You need not get into the details of any specific truth theory.

# Philosophical Logic

## 4.1 Truth Preservation View of Logic

Johns Hopkins University, Spring 2015

# Tarski on Logical Consequence

Tarksi [1936] opens his seminal essay on logical consequence with this remark on the subject of mathematical logic:

"The concept of *logical consequence* is one of those whose introduction into the field of strict formal investigation was not a matter of arbitrary decision on the part of this or that investigator; in defining this concept, efforts were made to adhere to the common usage of the language of everyday life.

# Tarski on Logical Consequence

But these efforts have been confronted with the difficulties which usually present themselves in such cases. With respect to the clarity of its content the common concept of consequence is in no way superior to the other concepts of everyday language. Its extension is not sharply bounded and its usage fluctuates. Any attempt to bring into harmony all possible vague, sometimes contradictory, tendencies which are connected with the use of this concept, is certainly doomed to failure. We must reconcile ourselves from the start to the fact that every precise definition of this concept will show arbitrary features to a greater or less degree." (p. 409)

# Tarski on Logical Consequence

Later in his essay, Tarski elaborates on what he takes to be the *common concept of consequence*:

"Consider any class $K$ of sentences and a sentence $X$ which follows from the sentences of this class. From an intuitive standpoint it can never happen that both the class $K$ consists only of true sentences and the sentence $X$ is false." (p. 414)

Reading 'it can never happen' as introducing a kind of modality, the informal target notion of consequence *necessarily preserves truth*.

# Tarski on Logical Consequence

Tarski continues:

"Moreover, since we are concerned here with the concept of logical, i.e., *formal*, consequence, and thus with a relation which is to be uniquely determined by the form of the sentences between which it holds, this relation cannot be influenced in any way by empirical knowledge, and in particular by knowledge of the objects to which the sentence $X$ or the sentences of the class $K$ refer. The consequence relation cannot be affected by replacing the designations of the objects referred to in these sentences by the designations of any other objects." (p. 415)

So the informal target notion of consequence necessarily preserves truth *by virtue of logical form and not due to the meaning/reference of any extra-logical constants*.

N. B. Tarski is considering an interpreted language.

# Tarski on Logical Consequence

Aside: Before moving on, you might want to ask yourself whether this common concept of consequence really is so *common*—at least outside the community of mathematicians and philosophers.

What is our real target?

# Tarski on Logical Consequence

Tarski's goal is to formulate a precise *semantic* account of logical consequence.

Before doing this, he first considers and rejects *syntactic* accounts of consequence.

**Syntactic view:** a sentence $X$ is a consequence of a class $K$ of sentences iff $X$ can be obtained from sentences in $K$ via certain rules of inference.

N. B. In all deductive systems? In some deductive system? In *sound* deductive systems?

# Tarski on Logical Consequence

**Problem:** There is a theory $T$ with theorems such as this:

($A_0$) 0 possesses property $P$.

($A_1$) 1 possesses property $P$.

($A_2$) 2 possesses property $P$.

And so on for each $n \in \mathbb{N}$.

But $T$ does not prove:

($A$) Every natural number possesses property $P$.

That is, $T$ is $\omega$-*inconsistent*.

# Tarski on Logical Consequence

According to Tarski, $(A)$ is a consequence of $(A_0)$, $(A_1)$, $(A_2)$, ...

"Yet intuitively it seems certain that the universal sentence $A$ follows in the usual sense from the totality of particular sentences $A_0$, $A_1$, ..., $A_n$, .... Provided all these sentences are true, the sentence $A$ *must* also be true." (p. 411, my emphasis)

Upshot: A syntactic concept of consequence based on standard inference rules is extensionally inadequate.

# Tarski on Logical Consequence

**Response:** Formulate new rules of inference such as the *rule of infinite induction*.

**Counter-response:** Gödel's First Incompleteness Theorem shows that any system of rules will always undergenerate:

"In every deductive theory (apart from certain theories of a particularly elementary nature), however much we supplement the ordinary rules of inference by new purely structural rules, it is possible to construct sentences which follow, in the usual sense, from the theorems of this theory, but which nevertheless cannot be proved in this theory on the basis of the accepted rules of inference." (p. 412-3)

# Tarski on Logical Consequence

Turning to the semantic approach:

"I should like to sketch here a method which, it seems to me, enables us to construct an adequate definition of the concept of consequence for a comprehensive class of formalized languages. I emphasize, however, that the proposed treatment of the concept of consequence makes no very high claim to complete originality. The ideas involved in this treatment will certainly seem to be something well known, or even something of his own, to many a logician who has given close attention to the concept of consequence and has tried to characterize it more precisely. Nevertheless it seems to me that only the methods which have been developed in recent years for the establishment of scientific semantics, and the concepts defined with their aid, allow us to present these ideas in exact form." (p. 414)

# Tarski on Logical Consequence

First attempt:

$(F)$ If, in the sentences of the class $K$ and in the sentence $X$, the constants—apart from the purely logical constants—are replaced by any other constants (like signs being everywhere replaced by like signs), and if we denote the class of sentences thus obtained from $K$ by $K'$, and the sentence obtained from $X$ by $X'$, then the sentence $X'$ must be true provided only that all sentences of the class $K'$ are true.

# Tarski on Logical Consequence

**Example:**

Consider a language with the following non-logical constants:

Names: ‘Mark Twain’, ‘Samuel Clemens’, ‘Virginia Woolf’.

Predicates: ‘is British’, ‘is a writer’.

# Tarski on Logical Consequence

| | |
|---|---|
| $K$ : | Mark Twain or Virginia Woolf is British |
| | Mark Twain is not British |
| $X$ : | Virginia Woolf is British |

Is $X$ a logical consequence of $K$ according to (F)? **Yes.**

A non-logical constant swap:

'MT' $\Rightarrow$ 'VW', 'VW' $\Rightarrow$ 'SC', 'is British' $\Rightarrow$ 'is a writer'.

| | |
|---|---|
| $K'$ : | Virginia Woolf or Samuel Clemens is a writer |
| | Virginia Woolf is not a writer |
| $X'$ : | Samuel Clemens is a writer |

$X'$ must be true provided that both sentences of $K'$ are true.

# Tarski on Logical Consequence

$K$ : Mark Twain is a writer
$X$ : Samuel Clemens is a writer

Is $X$ a logical consequence of $K$ according to $(F)$? **No.**

A non-logical constant swap:

'MT' $\Rightarrow$ 'VW', 'SC' $\Rightarrow$ 'MT', 'is a writer' $\Rightarrow$ 'is British'.

$K'$ : Virginia Woolf is British
$X'$ : Mark Twain is British

$K'$ is true but $X'$ is false.

# Tarski on Logical Consequence

**Problem:** As Tarski himself recognizes, $(F)$ is not a sufficient condition for $X$ to be a consequence of the class $K$ (though he thinks that $(F)$ is a necessary condition).

"It may, and it does, happen—it is not difficult to show this by considering special formalized languages—that the sentence $X$ does not follow in the ordinary sense from the sentences of the class $K$ although the condition $(F)$ is satisfied. This condition may in fact be satisfied only because the language with which we are dealing does not possess a sufficient stock of extra-logical constants. The condition $(F)$ could be regarded as sufficient for the sentence $X$ to follow from the class $K$ only if the designations of all possible objects occurred in the language in question. This assumption, however, is fictitious and can never be realized." (p. 415-6)

# Tarski on Logical Consequence

**Example:**

Consider a language with the following non-logical constants:

Names: 'Mark Twain', 'Samuel Clemens'.

Predicates: 'is a writer'.

If $(F)$ were a sufficient condition for logical consequence, then 'Samuel Clemens is a writer' would be a consequence of 'Mark Twain is a writer'.

# Tarski on Logical Consequence

Second attempt:

Given a class $L$ of sentences, let $L'$ be the class of sentential functions obtained by replacing every extra-logical constant by a corresponding variable, like constants being replaced by like variables, and unlike by unlike.

**Example:**

$L$ : Virginia Woolf or Samuel Clemens is a writer
Virginia Woolf is not a writer
Samuel Clemens is a writer

$L'$ : $Xx$ or $Xy$
not-$Xx$
$Xy$

# Tarski on Logical Consequence

**Def 4.1.1.** A *model* of a class $L$ of sentences is an arbitrary sequence of objects which satisfies every sentential function of the class $L'$.

N. B. These objects can be high-order classes.

**Example:**

Given the sequence of variables $\langle X, x, y \rangle$ appearing in $L'$ above, here are some models for $L$:

$\langle \{\text{Mark Twain}\}, \text{Virginia Woolf}, \text{Mark Twain} \rangle$

$\langle \{\text{Hulk Hogan}, \text{Andre the Giant}\}, \text{Vladimir Putin}, \text{Hulk Hogan} \rangle$

# Tarski on Logical Consequence

**Def 4.1.2.** The sentence $X$ *follows logically* from the sentences of the class $K$ if and only if every model of the class $K$ is also a model of the sentence $X$.

“It seems to me that everyone who understands the content of the above definition must admit that it agrees quite well with common usage. This becomes still clearer from its various consequence. In particular, it can be proved, on the basis of this definition, that every consequence of true sentences must be true, and also that the consequence relation which holds between given sentences is completely independent of the sense of the extra-logical constants which occur in these sentences.” (p. 417)

# Tarski on Logical Consequence

As Tarski himself recognizes, an outstanding issue is how to demarcate the logical versus extra-logical terms of a language.

Though Tarski does not think that this division of terms is arbitrary, he suggests that there might be no objective grounds for drawing it, and so we might *relativize* logical notions to a particular choice of which expressions count as 'logical'.

# Tarski on Logical Consequence

“Perhaps it will be possible to find important objective arguments which will enable us to justify the traditional boundary between logical and extra-logical expressions. But I also consider it to be quite possible that investigations will bring no positive results in this direction, so that we shall be compelled to regard such concepts as ‘logical consequence’, ‘analytical statement’, and ‘tautology’ as relative concepts which must, on each occasion, be related to a definite, although in greater or less degree arbitrary, division of terms into logical and extra-logical. The fluctuation in the common usage of the concept of consequence would—in part at least—be quite naturally reflected in such a compulsory situation.” (p. 420)

# Comparison with Modern Definition

Etchemendy [1988] discusses two differences between Tarski's Def 4.1.2 and the modern model-theoretic definition of consequence.

The first difference is inessential.

Whereas we now take models to be interpretations of a language, Tarski was working with interpreted languages.

However, replacing extra-logical constants with variables and then considering various objects that satisfy the resulting sentential functions is effectively the same thing as reinterpreting the extra-logical constants.

# Comparison with Modern Definition

The second difference is far more significant.

While our models have a *variable domain* $\mathcal{D}$, Etchemendy claims that Tarski was working with an intended interpretation with *fixed domain*.

As a result, sentences like $\exists x \exists y (x \neq y)$ come out logically true on Tarski's account.

N. B. This is a matter of considerable exegetical debate.

**Exercise 4.1.1**

Do the following classes of sentences have models? Variabilize the non-logical terms and provide models when they exist.

$L_1$ : Virginia Woolf is a writer and Mark Twain is a painter
Mark Twain is a writer and Virginia Woolf is a painter

$L_2$ : Neither Virginia Woolf nor Samuel Clemens is a writer
Mark Twain is not a painter
If Mark Twain is a painter, then Samuel Clemens is a writer

$L_3$ : Virginia Woolf is a writer iff Samuel Clemens is a writer
Samuel Clemens is a writer iff Mark Twain is not a writer
Virginia Woolf is not a writer iff Mark Twain is not a writer

**Exercise 4.1.2**

Demonstrate that $(F)$ is not a sufficient condition for logical consequence by considering a language whose only non-logical constants are one name and two predicate symbols.

# Philosophical Logic

## 4.2 Relevance Logic

Johns Hopkins University, Spring 2015

Relevance logicians famously reject the following implication:

**Ex Falso Quodlibet (EFQ):** $\varphi \wedge \neg\varphi \models \psi$.

Traditionally, this rule was rejected because the premise is not relevant to the conclusion.

**The Old Relevantism:** While truth preservation is a necessary condition for validity, it is not sufficient. The truth preservation requirement must be supplemented with a requirement of relevance.

The Lewis Argument for EFQ:

| | | |
|---|---|---|
| 1 | $\varphi \wedge \neg\varphi$ | Premise |
| 2 | $\varphi$ | $\wedge$ Elim: 1 |
| 3 | $\neg\varphi$ | $\wedge$ Elim: 1 |
| 4 | $\varphi \vee \psi$ | $\vee$ Intro: 2 |
| 5 | $\psi$ | Disjunctive Syllogism: 4,3 |

Relevance logicians must abandon $\wedge$-Elim, $\vee$-Intro, DS, or Transitivity of Entailment.

Aside from rejecting $\wedge$-Elim, all of these options have been endorsed by some logician or another.

The fathers of relevance logic, Anderson and Belnap [1975], reject DS.

Anderson and Belnap develop a *logic of first-degree entailment* $\mathbf{E}_{fde}$.

**Def 4.2.1.** The logical connectives are handled with the following *4-valued logic*:

| $\neg$ | |
|---|---|
| $\{\,\}$ | $\{\,\}$ |
| $\{F\}$ | $\{T\}$ |
| $\{T\}$ | $\{F\}$ |
| $\{T,F\}$ | $\{T,F\}$ |

| $\wedge$ | $\{\,\}$ | $\{F\}$ | $\{T\}$ | $\{T,F\}$ |
|---|---|---|---|---|
| $\{\,\}$ | $\{\,\}$ | $\{F\}$ | $\{\,\}$ | $\{F\}$ |
| $\{F\}$ | $\{F\}$ | $\{F\}$ | $\{F\}$ | $\{F\}$ |
| $\{T\}$ | $\{\,\}$ | $\{F\}$ | $\{T\}$ | $\{T,F\}$ |
| $\{T,F\}$ | $\{F\}$ | $\{F\}$ | $\{T,F\}$ | $\{T,F\}$ |

And similarly for the other sentential connectives.

**Def 4.2.2.** $\varphi$ is *true* iff $\varphi$ is $\{T\}$ or $\{T,F\}$.

**Def 4.2.3.** $\varphi$ is *false* iff $\varphi$ is $\{F\}$ or $\{T,F\}$.

**Def 4.2.4.** $\varphi \Rightarrow \psi$ is a *tautological entailment* iff:

- Any assignment of truth values to sentence letters that makes $\varphi$ true makes $\psi$ true.
- Any assignment of truth values to sentence letters that makes $\psi$ false makes $\varphi$ false.

**Def 4.2.5.** The argument from $\varphi_1, ..., \varphi_n$ to $\psi$ is *logically valid* iff $\varphi_1 \wedge ... \wedge \varphi_n \Rightarrow \psi$ is a tautological entailment.

Both EFQ and DS come out invalid in this system.

Why relevance?

Anderson and Belnap [1975] appeal to:

- The Natural Light of Reason.
- The logical tradition since Aristotle.
- The common sense of ‘naive freshmen’.

Burgess [1983] rejects all of these appeals.

Regarding the Natural Light:

"The appeal to Reason...is pointless, since what seems to one side the *lumen naturals* seems to the other side an *ignis fatuus*." (p. 42)

Regarding the logical tradition:

"Traditional logic manuals may list a fallacy of 'irrelevant conclusion', but they list it as an *informal* fallacy, pertaining not to formal logic, but rather to rhetoric." (p. 42)

Regarding the common sense of freshmen:

"Here we must distinguish two questions: (1) What *theory* about the nature of entailment strikes untrained undergraduates as most plausible at a first hearing? (2) What sort of logical system best agrees with the *practice* of nonspecialists in constructing and evaluating purported proofs? ... Question 1 might be investigated by holding a rally of the freshman class, shouting out the rival slogans 'Preserve the Truth!' and 'Make Logic Relevant!', and recording reaction on an applause-meter. What the outcome of such an *ad populum* appeal would be I neither know nor care. Question 2 might be investigated by looking over the graded examination papers from a large underclass course in mathematics, physics, or economics. The outcome, I submit, vindicates the classical position that [EFQ] expresses a fact about deducibility." (p. 43)

Hilbert's advice on how to 'prove' Fermat's Theorem: Keep on calculating until you miscalculate.

Burgess also presents some common sense instances of DS that relevantists have difficulty accounting for.

**Example.** Suppose that X has an insurance policy that pays off if X loses either an arm or a leg. And suppose moreover that one knows both that X is receiving payments and that he hasn't lost an arm. 'Well, then,' one concludes, 'he must have lost a leg'.

Some relevantists claim that such arguments are really instances of some other relevantistically acceptable schema (like the fission version of DS).

Other 'hardheaded' or 'true' relevantists reject DS and concede that logic sometimes conflicts with common sense.

But common sense was one of the original motivations for relevantism.

Later relevantists did not see relevance and truth as two separate requirements on consequence but rather saw relevance as something needed to ensure truth preservation.

**The New Relevantism:** Given the existence of true sentences whose negations are also true, a requirement of relevance is subsumed in the requirement to preserve truth.

Lewis [1982] discusses three versions of relevance logic.

**Logic of Entailment (E)**: 4-valued, validity preserves truth.

N. B. What about preserving non-falsity?

**R-mingle (RM)**: 3-valued (no $\{\}$), validity preserves truth + truth only.

**Logic of Paradox (LP)**: 3-valued (no $\{\}$), validity preserves truth.

Can a sentence be true and false?

Routley and Priest: Yes, in the literal sense.

Some candidates:

- Paradoxical sentences like the Liar.
- Conflicting implications of a legal code.

These examples alone motivate only the 3-valued logics RM or LP.

Lewis:

"The reason we should reject this proposal is simple. No truth does have, and no truth could have, a true negation. Nothing is, and nothing could be, literally both true and false. This we know for certain, and *a priori*, and without any exception for especially perplexing subject matters. The radical case for relevance should be dismissed just because the hypothesis it requires us to entertain is inconsistent.

That may seem dogmatic. And it is: I am affirming the very thesis that Routley and Priest have called into question and—contrary to the rules of debate—I decline to defend it. Further, I concede that it is indefensible against their challenge. They have called so much into question that I have no foothold on undisputed ground. So much the worse for the demand that philosophers always must be ready to defend their theses under the rules of debate." (p. 434-5)

Dunn: Sentences can be true and false according to some corpus of information (someone's system of beliefs, a data bank or almanac or encyclopedia, or even a work of fiction).

If we want to quarantine inconsistencies, we can use a relevance logic for determining truth and falsity in a corpus.

This motivates the four-valued logic E.

Lewis: Though a sentence and its negation can be true in a corpus, such inconsistencies are quarantined not by restrictions of relevance but rather by the method of *fragmentation*.

"I speak from experience as the repository of a mildly inconsistent corpus. I used to think that Nassau Street ran roughly east-west; that the railroad nearby ran roughly north-south; and that the two were roughly parallel. (By "roughly" I mean "to within 20°".) So each sentence in an inconsistent triple was true according to my beliefs, but not everything was true according to my beliefs. Now, what about the blatantly inconsistent conjunction of the three sentences?

I say that it was not true according to my beliefs. My system of beliefs was broken into (overlapping) fragments. Different fragments came into action in different situations, and the whole system of beliefs never manifested itself all at once. The first and second sentences in the inconsistent triple belonged to—were true according to—different fragments; the third belonged to both. The inconsistent conjunction of all three did not belong to, was in no way implied by, and was not true according to, any one fragment. That is why it was not true according to my system of beliefs taken as a whole. Once the fragmentation was healed, straightway my beliefs changed: now I think that Nassau Street and the railroad both run roughly northeast-southwest.

I think the same goes for other corpora in which inconsistencies are successfully quarantined. The corpus is fragmented. Something about the way it is stored, or something about the way it is used, keeps it from appearing all at once. It appears now as one consistent corpus, now as another. The disagreements between the fragments that appear are the inconsistencies of the corpus taken as a whole. We avoid trouble with such inconsistencies (and similar trouble with errors that do not destroy consistency) by not reasoning from mixtures of fragments. Something is true according to the corpus if and only if it is true according to some one fragment thereof. So we have no guarantee that implication preserves truth according to the corpus, unless all the premises come from a single fragment. What follows from two or more premises drawn from disagreeing fragments may be true according to no fragment, hence not true according to the corpus." (p. 436)

‘Fallacies of relevance’ such as DS can fail to preserve truth according to an inconsistent corpus. But so can conjunction introduction. Only one-premise implications can be trusted not to mix fragments.

“I am inclined to think that when we are forced to tolerate inconsistencies in our beliefs, theories, stories, etc., we quarantine the inconsistencies entirely by fragmentation and not at all by restrictions of relevance. In other words, truth according to any single fragment is closed under unrestricted classical implication.” (p. 437)

Lewis does suggest a role for relevance logic.

We might say that an ambiguous sentence is both *true-osd* and *false-osd* (true/false on some disambiguation).

This motivates the 3-valued logics RM and LP.

**Exercise 4.2.1**

Complete Def 4.2.2 by proving 4-valued truth tables for $\vee$, $\supset$, and $\equiv$.

**Exercise 4.2.2**

Prove that EFQ and DS are invalid in **E**, **RM**, and **LP**.

# Philosophical Logic

## 4.3 Intuitionistic Logic

Johns Hopkins University, Spring 2015

Intuitionistic logicians reject the standard truth preservation view.

The central concern of logic is not *truth* but *proof*.

The Law of Excluded Middle can fail: we should not currently accept Goldbach's Conjecture $\vee$ $\neg$Goldbach's Conjecture since we do not currently have a proof of this conjecture or a proof of its negation.

Brouwer, Heyting, and Kolmogorov independently proposed the following reinterpretation of the sentential logical constants:

- There is no proof of $\bot$
- A proof of $\neg\varphi$ is a method of converting any proof of $\varphi$ into a proof of $\bot$
- A proof of $(\varphi \wedge \psi)$ is a proof of $\varphi$ and proof of $\psi$
- A proof of $(\varphi \vee \psi)$ is a proof of $\varphi$ or proof of $\psi$
- A proof of $(\varphi \sqsupset \psi)$ is a method of converting any proof of $\varphi$ into a proof of $\psi$

where $\sqsupset$ is the intuitionistic conditional.

Note that $\neg\varphi \equiv \varphi \sqsupset \bot$. In the intuitionistic language $\mathcal{L}_{ISENT}$, we will treat $\sqsupset$ as basic and $\neg$ as a derived symbol.

A Hilbert-style proof system for classical sentential logic consists of *modus ponens* and the following axioms:

- $A \supset (B \supset A)$
- $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$
- $(A \wedge B) \supset A$ and $(A \wedge B) \supset B$
- $A \supset (B \supset (A \wedge B))$
- $A \supset (A \vee B)$ and $B \supset (A \vee B)$
- $(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))$
- $\perp \supset A$
- $A \vee \neg A$

$\vdash_{\mathbf{CL}} \varphi$ designates that $\varphi$ is provable in this classical system **CL**.

We get intuitionistic sentential logic by eliminating the last axiom $A \vee \neg A$.

- $A \supset (B \supset A)$
- $(A \supset (B \supset C)) \supset ((A \supset B) \supset (A \supset C))$
- $(A \wedge B) \supset A$ and $(A \wedge B) \supset B$
- $A \supset (B \supset (A \wedge B))$
- $A \supset (A \vee B)$ and $B \supset (A \vee B)$
- $(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))$
- $\bot \supset A$

$\vdash_{\mathbf{IL}} \varphi$ designates that $\varphi$ is provable in the resulting proof system **IL**.

Is there a semantics corresponding to **IL**?

**Def 4.3.1.** An *intuitionistic model* for $\mathcal{L}_{ISENT}$ is a Kripke model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ where $\mathcal{R}$ is a *partial order* (reflexive, transitive, and antisymmetric) and the following *heredity property* holds:

For all $p \in At_{\mathcal{L}_{ISENT}}$ and $w, v \in \mathcal{W}$,

$\mathcal{V}(p, w) = T$ and $w\mathcal{R}v$ only if $\mathcal{V}(p, v) = T$.

You can think of each element $w \in \mathcal{W}$ as a stage of inquiry and a walk along $\mathcal{R}$ as a process of inquiry. The heredity property captures the assumption that when something is established, it is established for good.

**Def 4.3.2.** A recursive specification lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{ISENT}} \times \mathcal{W} \to \{T, F\}$ for $\mathcal{L}_{ISENT}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{ISENT}}$ and world $w \in \mathcal{W}$ to a value in $\{T, F\}$:

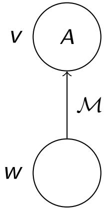
$$
\begin{array}{rcl}
[\![p]\!]^w_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p) = T \\
[\![\bot]\!]^w_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\varphi \wedge \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \sqsupset \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & \forall v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = T \text{ only if } [\![\psi]\!]^v_{\mathcal{M}} = T)
\end{array}
$$

Since $\neg\varphi \equiv \varphi \sqsupset \bot$,

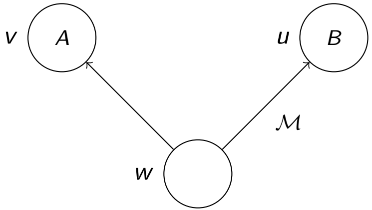
$$
[\![\neg\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = F)
$$

When $[\![\varphi]\!]^w_{\mathcal{M}} = T$, we needn't say that $\varphi$ is *true* at $w$ in $\mathcal{M}$. We can say that $w$ *forces* $\varphi$ in $\mathcal{M}$.

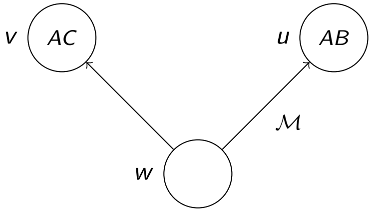
**Example:** $A \vee \neg A$ and $\neg\neg A \supset A$ are not always forced.



**Example:** $\neg(A \wedge B) \sqsupset (\neg A \vee \neg B)$ is not always forced.



**Example:** $(A \supset (B \vee C)) \supset ((A \supset B) \vee (A \supset C))$ is not always forced.



Two lemmas concerning this intuitionistic semantics:

**Lem 4.3.1.** For any intuitionistic model $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$, worlds $w, v \in \mathcal{W}$, and sentence $\varphi \in S_{\mathcal{L}_{ISENT}}$:

$[\![\varphi]\!]^w_{\mathcal{M}} = T$ and $w\mathcal{R}v$ only if $[\![\varphi]\!]^v_{\mathcal{M}} = T$.

That is, heredity extends to all sentences in $S_{\mathcal{L}_{ISENT}}$.

**Lem 4.3.2.** If $v$ is a dead-end state in $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$,

$[\![\varphi]\!]^v_{\mathcal{M}} = T$ iff $\varphi$ is classically true at $v$ in $\mathcal{M}$.

Three theorems:

**Thm 4.3.1 (Soundness and Completeness Theorem for IL).**
$\vdash_{\mathbf{IL}} \varphi$ iff $\varphi$ is valid over the class of *finite* intuitionistic Kripke models.

**Thm 4.3.2 (Disjunction Property).** $\vdash_{\mathbf{IL}} \varphi \vee \psi$ only if $\vdash_{\mathbf{IL}} \varphi$ or $\vdash_{\mathbf{IL}} \psi$.

**Thm 4.3.3 (Glivenko's Theorem).** $\vdash_{\mathbf{CL}} \varphi$ iff $\vdash_{\mathbf{IL}} \neg\neg\varphi$.

**Exercise 4.3.1**

For each of the following sentences, provide a pointed model $\mathcal{M}, w$ in which the sentence is not forced.

$(\neg\neg A \supset A) \supset (A \vee \neg A)$

$(\neg A \supset (B \vee C)) \supset ((\neg A \supset B) \vee (\neg A \supset C))$

**Exercise 4.3.2**

Prove Lem 4.3.2.

# Philosophical Logic

## 4.4 Informational View of Logic

Johns Hopkins University, Spring 2015

Deductively good arguments involving informational modal operators and the indicative conditional suggest that logically valid arguments do not preserve truth.

**Example:** (cf. Łukasiewicz [1930])

(P1) Colonel Mustard didn't do it.
(C) It's not the case that Colonel Mustard might have done it.

**Example:** (cf. McGee [1985])

(P1) If a married woman committed the murder, then if Mrs. Peacock didn't do it, it was Mrs. White.
(P2) A married woman committed the murder.
(C) If Mrs. Peacock didn't do it, it was Mrs. White.

We do well to make these arguments in both categorical and hypothetical deliberative contexts. But they arguably fail to preserve truth.

**Def 4.4.1.** The language $\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}$ has the following syntax:

$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid \Diamond\varphi \mid \Box\varphi \mid (\varphi \Rightarrow \varphi)$

where $\Rightarrow$ is the indicative conditional.

$At_{\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}} = \{A, B, ...\}$ is the set of atoms in $\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}$.
$S_{\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}}$ is the set of sentences in $\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}$.

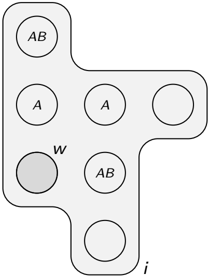
**Def 4.4.2.** A *model* $\mathcal{M} = \langle \mathcal{W}, \mathcal{V} \rangle$ for $\mathcal{L}_{SENT_{\Diamond / \Rightarrow}}$ consists of a nonempty set $\mathcal{W}$ of possible worlds and a valuation $\mathcal{V} : At_{\mathcal{L}_{SENT_{\Diamond / \Rightarrow}}} \times \mathcal{W} \to \{T, F\}$ mapping each sentence letter $p \in At_{\mathcal{L}_{SENT_{\Diamond / \Rightarrow}}}$ and world $w \in \mathcal{W}$ to a truth value.

Sentences in $S_{\mathcal{L}_{SENT_{\Diamond / \Rightarrow}}}$ will be evaluated for truth relative both to a world $w \in \mathcal{W}$ and to an *information state* $i \in 2^{\mathcal{W}}$ (a set of possible worlds).

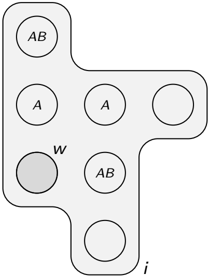
**Def 4.4.3.** A recursive specification of *truth in a model* lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}} \times \mathcal{W} \times 2^{\mathcal{W}} \to \{T, F\}$ for $\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{SENT_{\Diamond/\Rightarrow}}}$, world $w \in \mathcal{W}$, and information state $i \in 2^{\mathcal{W}}$ to a truth value:

$$\begin{array}{rcl}
[\![p]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p, w) = T \\
[\![\bot]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{w,i}_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{w,i}_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^{w,i}_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{w,i}_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^{w,i}_{\mathcal{M}} = T \\
[\![\Box\varphi]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & \forall v \in i([\![\varphi]\!]^{v,i}_{\mathcal{M}} = T) \\
[\![\Diamond\varphi]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & \exists v \in i([\![\varphi]\!]^{v,i}_{\mathcal{M}} = T) \\
[\![\varphi \Rightarrow \psi]\!]^{w,i}_{\mathcal{M}} = T & \text{iff} & \forall v \in i + \varphi([\![\psi]\!]^{v,i+\varphi}_{\mathcal{M}} = T)
\end{array}$$

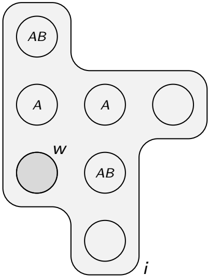
where $i + \varphi$ appearing in the clause for the indicative is the largest subset $i' \subseteq i$ such that $\forall w \in i'([\![\varphi]\!]^{w,i'}_{\mathcal{M}} = T)$.



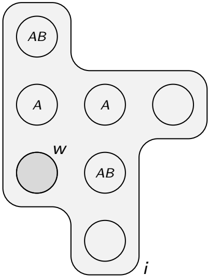

$$[\![\Box(A \vee \neg B)]\!]^{w,i}_{\mathcal{M}} = ?$$



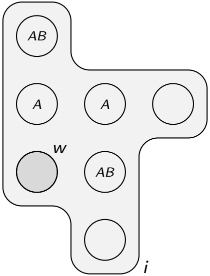

$$[\![\Box(A \vee \neg B)]\!]^{w,i}_{\mathcal{M}} = T.$$



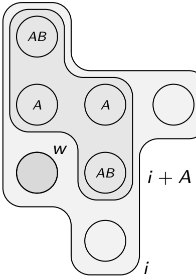

$$[\![\Diamond(A \wedge B)]\!]_{\mathcal{M}}^{w,i} = ?$$



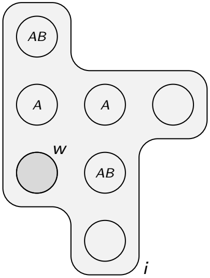

$$[\![\Diamond(A \wedge B)]\!]^{w,i}_{\mathcal{M}} = T.$$



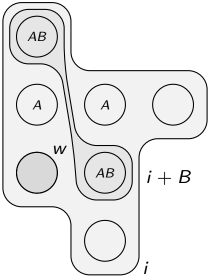

$$[\![A \Rightarrow B]\!]^{w,i}_{\mathcal{M}} = ?$$




$$[\![A \Rightarrow B]\!]^{w,i}_{\mathcal{M}} = F.$$




$$[\![ B \Rightarrow A ]\!]_{\mathcal{M}}^{w,i} = \,?$$



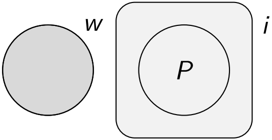

$$[\![B \Rightarrow A]\!]^{w,i}_{\mathcal{M}} = T.$$

What about the formal consequence relation?

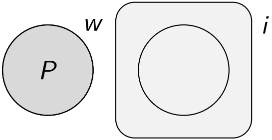
Running with the truth preservation view of logic, here is a natural first suggestion:

**Def 4.4.4.** $\{\varphi_1, ..., \varphi_n\} \models_0 \psi$ just in case there is no model $\mathcal{M}$ such that for some $w \in W$ and $i \in 2^W$, $[\![\varphi_1]\!]_{\mathcal{M}}^{w,i} = ... = [\![\varphi_n]\!]_{\mathcal{M}}^{w,i} = T$ but $[\![\psi]\!]_{\mathcal{M}}^{w,i} = F$.

However, Def 4.4.4 is too demanding. $\models_0$ invalidates some deductively good arguments.




(P1) Mrs. Peacock must have done it.
(C) Mrs. Peacock did it.



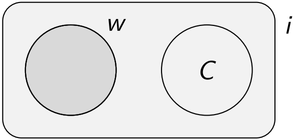

(P1) Mrs. Peacock did it.
(C) Mrs. Peacock might have done it.

Kolodny and MacFarlane [2010] offer this fix:

**Def 4.4.5.** $\{\varphi_1, ..., \varphi_n\} \models_{Tr} \psi$ just in case there is no model $\mathcal{M}$ such that for some $i \in 2^W$ and $w \in i$, $[\![\varphi_1]\!]^{w,i}_{\mathcal{M}} = ... = [\![\varphi_n]\!]^{w,i}_{\mathcal{M}} = T$ but $[\![\psi]\!]^{w,i}_{\mathcal{M}} = F$.

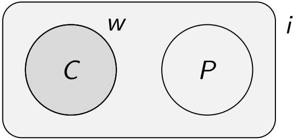
$\{\Box A\} \models_{Tr} A$ and $\{A\} \models_{Tr} \Diamond A$, so $\models_{Tr}$ supports our judgments that these implications hold.

However, $\models_{Tr}$ also invalidates some intuitively good arguments.

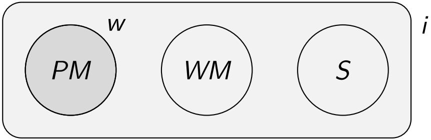



(P1) Colonel Mustard didn't do it.

(C) It's not the case that Colonel Mustard might have done it.




(P1) Either Colonel Mustard did it or Professor Plum did it.
(P2) Professor Plum didn't do it.
(C) Colonel Mustard must have done it.




(P1) If a married woman committed the murder, then if Mrs. Peacock didn't do it, it was Mrs. White.
(P2) A married woman committed the murder.
(C) If Mrs. Peacock didn't do it, it was Mrs. White.

How to respond to this mismatch?

- One might dismiss the positive evaluations of the previous three arguments as misguided.
- One might concede that these are good arguments and maintain the truth preservation view, but concede that the formal relation $\models_{Tr}$ is not up to the job.
- One might say that the three inferences are logically invalid but still good inferences.
- One might surrender the idea that logically valid arguments preserve truth. Perhaps we can understand logic in some other way such that validity and good deductive argument coincide.

Let us explore this last option.

The formal semantics suggests another formal consequence relation (cf. Veltman [1996], Yalcin [2007], Kolodny and MacFarlane [2010]):

**Def 4.4.6.** $\{\varphi_1, ..., \varphi_n\} \models_I \psi$ just in case there is no $\mathcal{M}$ such that for some $i \in 2^W$, $\forall w \in i([\![\varphi_1]\!]^{w,i}_{\mathcal{M}} = ... = [\![\varphi_n]\!]^{w,i}_{\mathcal{M}} = T)$ but $\neg\forall w \in i([\![\psi]\!]^{w,i}_{\mathcal{M}} = T)$.

Unlike the previous relations, the 'informational consequence' relation $\models_I$ does not preserve truth at an index $\langle w, i\rangle$ in all models, whether or not $w \in i$. The relation $\models_I$ holds between a set of sentences and a single sentence in $Sent_{\mathcal{L}_{\Diamond/\Rightarrow}}$ just in case all information states in all models have a particular kind of *structure*.

Each sentence $\varphi \in Sent_{\mathcal{L}_{\Diamond/\Rightarrow}}$ corresponds to a potential feature of information states:

**Def 4.4.7.** $i \rhd \varphi$ just in case $\forall w \in i(\llbracket\varphi\rrbracket^{w,i}_{\mathcal{M}} = T)$.

Let us say that when $i \rhd \varphi$, $\varphi$ is *incorporated* in $i$.

For instance:

$i \rhd A$ just in case $\forall w \in i(\llbracket A\rrbracket^{w,i}_{\mathcal{M}} = T)$; that is, $i \rhd A$ just in case every world in $i$ is an $A$-world.

$i \rhd \Diamond B$ just in case $\forall w \in i(\llbracket\Diamond B\rrbracket^{w,i}_{\mathcal{M}} = T)$; that is, $i \rhd \Diamond B$ just in case some world in $i$ is a $B$-world (assuming $i \neq \emptyset$).

Def 4.4.6 can be restated as:

$\{\varphi_1, ..., \varphi_n\} \models_I \psi$ just in case there is no $\mathcal{M}$ such that for some $i \in 2^W$, $i \rhd \varphi_1, ..., i \rhd \varphi_n$, and $i \not\rhd \psi$.

We can think of $\models_I$ as preserving not truth at an index, but incorporation in all information states.

$\models_I$ validates the good deductive arguments that give $\models_{Tr}$ a rough time.

Why is the argument valid from ‘Professor Plum didn’t do it’ to ‘It’s not the case that Professor Plum might have done it’?

- **Truth Preservation View:** because it is impossible for the former sentence to be true and for the latter to be false by virtue of logical form.
- **Informational View:** because any body of information (the content of an eyewitness’s utterances, the evening news, and so on) according to which Professor Plum didn’t do it is therefore also information according to which it’s not the case that he might have done it.

# Philosophical Logic

## 4.5 Field Against Truth Preservation

Johns Hopkins University, Spring 2015

The standard truth preservation view of logic has come under fire from other directions.

Field [2006], [2008], [2009a], [2009b], [*ms.*] has argued in recent work that this orthodoxy cannot be right. Logic, says Field, is not really about truth preservation.

His argument in a nutshell:

If we look at our best formal truth theories developed since the 1970s to handle the Liar paradox and related semantic paradoxes, these theories formulated in a language with a general untyped truth predicate $Tr(x)$ include axioms that they do not regard as true, or rules of inference that they do not regard as unrestrictedly truth preserving.

Worse, adding to truth theory $T$ either the sentence saying that all of $T$'s axioms are true or the sentence saying that all of $T$'s rules of inference preserve truth results in inconsistency.

But all of us accept, or should accept, one or the other of these theories, taking its axioms/rules to govern our inferential practices.

So if we want to hold onto the idea that logic—where 'logic' is broad enough to include the logic of $Tr(x)$—lines up with good deductive inference such that these axioms/rules are logical truths/logically valid, then we are not in a position to consistently accept that logic is about unrestricted truth preservation.

Note that Field's argument is not targeted at the mathematical project of *explicating* logical validity in terms of truth preservation in all models.

Rather, it is targeted at the philosophical approach of *defining* logical validity in terms of unrestricted truth preservation.

If we insist that logical validity is so definable, then some of the good deductive arguments that we make in our deliberations are not really logically valid.

Let us consider two truth theories to get a feel for Field's argument.

Background: No sufficiently powerful consistent truth theory $T$ that allows representations of recursive relations and whose logic is classical proves all instances of the T-Schema: $Tr(\ulcorner\varphi\urcorner) \equiv \varphi$ for $\varphi \in S_{\mathcal{L}}$.

$\ulcorner\varphi\urcorner$ is the Gödel code of $\varphi$.

Since $T$ proves $\neg Tr(\ulcorner\lambda\urcorner) \equiv \lambda$ for the Liar sentence $\lambda$, this would lead to paradox.

Therefore, truth theories must either restrict the T-Schema in some way, weaken the logic, or both.

The first theory that we will consider is a classical truth theory that gives up the full T-Schema. The second theory retains the full T-Schema but has a weaker logic.

The Kripke-Feferman Theory (KF) proves all instances of $Tr(\ulcorner\varphi\urcorner) \supset \varphi$ but not $\varphi \supset Tr(\ulcorner\varphi\urcorner)$.

Field takes the sentences $Tr(\ulcorner\varphi\urcorner) \supset \varphi$ to be among its axioms. However, KF proves that $Tr(\ulcorner\lambda\urcorner) \supset \lambda$ is not true.

| | | |
|---|---|---|
| 1 | $\neg Tr(\ulcorner\lambda\urcorner) \equiv \lambda$ | Liar Property |
| 2 | $Tr(\ulcorner\lambda\urcorner) \supset \lambda$ | Axiom |
| 3 | $Tr(\ulcorner\neg\lambda\urcorner) \supset \neg\lambda$ | Axiom |
| 4 | $Tr(\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner) \equiv Tr(\ulcorner\neg\lambda\urcorner)$ | Axiom |
| 5 | $Tr(\ulcorner Tr(\ulcorner\lambda\urcorner) \supset \lambda\urcorner) \equiv Tr(\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner) \vee Tr(\ulcorner\lambda\urcorner)$ | Axiom |
| 6 | $\lambda$ | From 1,2 |
| 7 | $\neg Tr(\ulcorner\lambda\urcorner)$ | From 1,6 |
| 8 | $\neg Tr(\ulcorner\neg\lambda\urcorner)$ | From 3,6 |
| 9 | $\neg Tr(\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner)$ | From 4,8 |
| 10 | $\neg Tr(\ulcorner Tr(\ulcorner\lambda\urcorner) \supset \lambda\urcorner)$ | From 5,7,9 |

Thus, KF $\cup$ $\{Tr(\ulcorner Tr(\ulcorner \lambda \urcorner) \supset \lambda \urcorner)\}$ is inconsistent.

One possible move is to weaken the truth theory by excluding problematic instances of $Tr(\ulcorner \varphi \urcorner) \supset \varphi$ from the axioms, but Field [2006] argues that this "seems totally against the spirit of KF [since] the whole point of KF was to insist that restrictions are only required [for $\varphi \supset Tr(\ulcorner \varphi \urcorner)$]."

A clear proposal for how to restrict KF has also never been offered.

Field's own favorite truth theories are non-classical 'paracomplete' ones that do not prove all instances of the law of excluded middle.

In order for the truth predicate $Tr(x)$ to play its useful standard role as a device for forming conjunctions and disjunctions, Field argues that $\varphi$ and $Tr(\ulcorner \varphi \urcorner)$ must be *intersubstitutable* in transparent (non-quotational, etc.) contexts.

But this intersubstitutivity combined with classical logic leads to trouble.

| | | |
|---|---|---|
| 1 | $\lambda \equiv \neg Tr(\ulcorner\lambda\urcorner)$ | Liar Property |
| 2 | $\lambda \vee \neg\lambda$ | Axiom |
| 3 | $\lambda$ | Assumption |
| 4 | $\neg Tr(\ulcorner\lambda\urcorner)$ | From 1,3 |
| 5 | $\neg\lambda$ | From 4 (Sub *Tr*) |
| 6 | $\perp$ | From 3,5 |
| 7 | $\neg\lambda$ | Assumption |
| 8 | $Tr(\ulcorner\lambda\urcorner)$ | From 1,7 |
| 9 | $\lambda$ | From 8 (Sub *Tr*) |
| 10 | $\perp$ | From 7,9 |
| 11 | $\perp$ | From 2,3-6,7-10 |

Thus, theories that are not *externally dialethethic*—that reject the existence of 'true contradictions'—and hold on to the intersubstitutivity of $\varphi$ and $Tr(\langle\varphi\rangle)$ and the classical elimination rules for $\vee$ and $\equiv$ must restrict the law of excluded middle.

Now, let $\gamma$ be a Curry sentence, that is, a sentence such that the theories under consideration all prove $\gamma \leftrightarrow (Tr(\ulcorner\gamma\urcorner) \rightarrow \perp)$, and consider the argument from $\gamma$ and $\gamma \rightarrow \perp$ to $\perp$.

Field's favored paracomplete theories restrict the introduction rule for the conditional (hence the shift in notation from $\supset$ to $\rightarrow$) but *modus ponens* holds. Still, adding the sentence to these theories saying that this particular instance of *modus ponens* preserves truth (and so the sentence saying that *modus ponens* unrestrictedly preserves truth) results in inconsistency.

| | | |
|---|---|---|
| 1 | $(Tr(\ulcorner\gamma\urcorner) \wedge Tr(\ulcorner\gamma \rightarrow \bot\urcorner)) \rightarrow Tr(\ulcorner\bot\urcorner)$ | Premise |
| 2 | $\gamma \leftrightarrow (Tr(\ulcorner\gamma\urcorner) \rightarrow \bot)$ | Curry Property |
| 3 | $(\gamma \wedge (\gamma \rightarrow \bot)) \rightarrow \bot$ | From 1 (Sub *Tr*) |
| 4 | $\gamma \leftrightarrow (\gamma \rightarrow \bot)$ | From 2 (Sub *Tr*) |
| 5 | $(\gamma \wedge \gamma) \rightarrow \bot$ | From 3,4 (Sub $\leftrightarrow$) |
| 6 | $\gamma \rightarrow \bot$ | From 5 (Sub $\leftrightarrow$) |
| 7 | $\gamma$ | From 4,6 |
| 8 | $\bot$ | From 6,7 |

But if logical validity shouldn't be defined in terms of truth preservation, then what is logic about?

Field's [2009a] answer is this:

"If logic is not the science of what [forms of inference] necessarily preserve truth, it is hard to see what the subject of logic could possibly be, if it isn't somehow connected to norms of thought." (p. 263)

His bold suggestion is that we should regard the normative component of logic as fundamental. Rather than regarding logic as a descriptive science, we should instead recognize that logic is *essentially normative*.

This is not to say that logical validity should be *defined* in terms of its normativity for thought. Field finds this tack repugnant, and argues that it is best not to define logical validity at all but to treat it as a primitive notion and illuminate its conceptual role.

Field [*ms.*] suggests the following role for validity:

"To regard an inference or argument as valid is (in large part anyway) to accept a constraint on belief: one that prohibits fully believing its premises without fully believing its conclusion. (The prohibition should be 'due to logical form': for any other argument of that form, the constraint should also prohibit fully believing the premises without fully believing the conclusion." (p. 11)

This proposal is refined in Field [2009a], §I, and [*ms.*], §2.