NASSLLI 2014

Philosophical Logic Bootcamp

Justin Bledin, Johns Hopkins University

Session 1. Classical First-Order Logic and Advanced Quantifiers

Saturday June 21, 1:00pm-3:00pm

- Introduction
- Review of Sentential Logic
- Review of First-Order Logical
- Generalized Quantifiers
- Substitutional Quantifiers
- Plural Quantifiers

Session 2. Recursion Theory, Gödel's Incompleteness Theorems, and Nonfirstorderizability

Saturday June 21, 3:00pm-5:00pm

- Computability Theory
- Gödel's First Incompleteness Theorem
- Gödel's Second Incompleteness Theorem
- Nonfirstorderizability

Session 3. Modal Logic and Its Applications

Sunday June 22, 9:30am-12:30pm

- Propositional Modal Logic
- The Modal Zoo
- Temporal Logic
- Counterfactuals
- Deontic Logic
- Epistemic Logic

Session 4. Logical Consequence: Against Truth Preservation

Sunday June 22, 2:30am-5:30pm

- Intuitionistic Logic
- Field's Truth-Theoretic Argument
- Informational View of Logic

NASSLLI 2014

Classical First-Order Logic and Advanced Quantifiers

Justin Bledin, Johns Hopkins University

1.1. Introduction

What is logic?

Largely pre-theoretic starting point:

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

1.1. Introduction

For example, these arguments are good:

(P1) Testudo is the University of Maryland mascot.
(C) Testudo or Jay is the University of Maryland mascot.

(P1) Terrapins live in fresh or brackish water.
(P2) Testudo is a terrapin.
(C) Testudo lives in fresh or brackish water.

1.1. Introduction

These arguments are not:

(P1) Testudo or Jay is the University of Maryland mascot.
(C) Testudo is the University of Maryland mascot.

(P1) No terrapins live in saltwater.
(P2) Testudo is a terrapin.
(C) Testudo lives in saltwater.

1.1. Introduction

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

The *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 the logical form* of the sentences in the argument, and not due to the meaning of any non-logical symbols.

1.1. Introduction

These intuitions all come together in our first definition:

Def 1.1. The argument from $\varphi_1, \ldots, \varphi_n$ to ψ is *logically valid* if and only if it is impossible for each of $\varphi_1, \ldots, \varphi_n$ to be true and for ψ to be false by virtue of their 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, we will call the truth preservation view itself into question. But for the time being, we will accept Def 1.1.

1.1. Introduction

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 1.2. The argument from $\varphi_1, ..., \varphi_n$ to ψ 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 ψ 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 Def 1.1 is cashed out by quantifying over all models.

1.1. Introduction

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 logically 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.

1.2. Review of Sentential Logic

Def 1.3. The *language of sentential logic* $\mathcal{L}_{SENT}$ has the following syntax:

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

Note that this language is a bit redundant since $\bot$ 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 remaining Boolean connectives $\supset$ and $\equiv$ 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 well-formed sentences in $\mathcal{L}_{SENT}$.

1.2. Review of Sentential Logic

Def 1.4. A *model* $\mathcal{M} = \langle \mathcal{V} \rangle$ for $\mathcal{L}_{SENT}$ consists only of a valuation function $\mathcal{V} : At_{\mathcal{L}_{SENT}} \mapsto \{T, F\}$ 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.

1.2. Review of Sentential Logic

Def 1.5. A recursive specification of *truth-in-a-model* lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT}} \mapsto \{T, F\}$ 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 formed with these constants depends only on the truth value of the constituent subsentence(s).

1.2. Review of Sentential Logic

Def 1.6. The argument from $\varphi_1, ..., \varphi_n$ to ψ is *tautologically valid*, $\{\varphi_1, ..., \varphi_n\} \models_{SENT} \psi$, just in case there is no model $\mathcal{M}$ for $\mathcal{L}_{SENT}$ where $[\![\varphi_1]\!]_{\mathcal{M}} = ... = [\![\varphi_n]\!]_{\mathcal{M}} = T$ and $[\![\psi]\!]_{\mathcal{M}} = F$.

The term '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 1.7. The sentence φ is a *tautology* just in case $\emptyset \models_{SENT} \varphi$.

Def 1.8. The sentences φ and ψ are *tautologically equivalent* just in case both $\{\varphi\} \models_{SENT} \psi$ and $\{\psi\} \models_{SENT} \varphi$.

And so forth.

1.2. Review of Sentential Logic

(P1) Testudo is the University of Maryland mascot.
(C) Testudo or Jay is the University of Maryland mascot.

(P1) T
(C) $T \vee J$

This argument is tautologically valid.

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

Thus, $\{T\} \models_{SENT} T \vee J$.

1.2. Review of Sentential Logic

(P1) Testudo or Jay is the University of Maryland mascot.
(C) Testudo is the University of Maryland mascot.

(P1) $T \vee J$
(C) T

This argument is tautologically invalid.

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

Thus, $\{T \vee J\} \not\models_{SENT} T$.

1.2. Review of Sentential Logic

(P1) Terrapins live in fresh or brackish water.
(P2) Testudo is a terrapin.
(C) Testudo lives in fresh or brackish water.

(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 $\mathcal{L}_{SENT}$.

1.3. Review of First-Order Logic

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

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

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

n-ary functions $\{f_1^n, f_2^n, ...\}$ for each $n \in \mathbb{N}$

n-ary predicates $\{P_1^n, P_2^n, ...\}$ for each $n \in \mathbb{N}$

Def 1.9. 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 1.10. $\mathcal{L}_{FOL}$ has the following syntax for formulae:

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

1.3. Review of First-Order Logic

Def 1.11. The *scope* of a quantifier is the formula that directly follows it.

Def 1.12. 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*.

Def 1.13. A formula containing free variables is *open*. A formula sans free variables is a *closed* formula or *sentence*.

$At_{\mathcal{L}_{FOL}}$ contains sentences of the form $P_i^n t_{j_1} \ldots t_{j_n}$ and $t_i = t_j$.
$S_{\mathcal{L}_{FOL}}$ is the set of well-formed sentences in $\mathcal{L}_{FOL}$.

1.3. Review of First-Order Logic

Def 1.14. 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 $\mathcal{D}$.

$$\begin{aligned} \mathcal{I}: \quad & \{c_1, c_2, ...\} \mapsto \mathcal{D} \\ & \{f_1^n, f_2^n, ...\} \mapsto \mathcal{D}^{\mathcal{D}^n} \text{ for each } n \in \mathbb{N} \\ & \{P_1^n, P_2^n, ...\} \mapsto \mathcal{D}^n \text{ for each } n \in \mathbb{N} \end{aligned}$$

1.3. Review of First-Order Logic

Def 1.15. An *assignment* $g : \{x_1, x_2, ...\} \mapsto \mathcal{D}$ is a function mapping each variable to a member of $\mathcal{D}$.

Def 1.16. The *extension* $[\![t_i]\!]^g_{\mathcal{M}}$ of each term of $\mathcal{L}_{FOL}$ in $\mathcal{M}$ under assignment g is determined recursively as follows:

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

1.3. 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 1.17. *Truth-in-a-model under an assignment* is determined by the following recursive clauses:

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

where $g[x_i \to d](x_i) = d$ and $g(x_j) = g[x_i \to d](x_j)$ for $j \neq i$.

1.3. Review of First-Order Logic

$[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{FOL}} \mapsto \{T, F\}$ is now easily defined:

Def 1.18. $[\![\varphi]\!]_{\mathcal{M}} = T$ just in case $[\![\varphi]\!]^{g}_{\mathcal{M}} = T$ for all/some g.

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

Def 1.19. The argument from $\varphi_1, ..., \varphi_n$ to ψ is *logically valid*, $\{\varphi_1, ..., \varphi_n\} \models \psi$, 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 validity extensionally captures the informal target notion.

1.3. Review of First-Order Logic

(P1) Terrapins live in fresh or brackish water.
(P2) Testudo is a terrapin.
(C) Testudo lives in fresh or brackish water.

(P1) $\forall x(Tx \supset (Fx \vee Bx))$
(P2) Tt
(C) $Ft \vee Bt$

This argument is logically valid.

For any model $\mathcal{M}$, $[\![\forall x(Tx \supset (Fx \vee Bx))]\!]_{\mathcal{M}} = T$ only if $\mathcal{I}(t) \notin \mathcal{I}(T)$ or $\mathcal{I}(t) \in \mathcal{I}(F)$ or $\mathcal{I}(t) \in \mathcal{I}(B)$.

$[\![Tt]\!]_{\mathcal{M}} = T$ only if $\mathcal{I}(t) \in \mathcal{I}(T)$.

So $\mathcal{I}(t) \in \mathcal{I}(F)$ or $\mathcal{I}(t) \in \mathcal{I}(B)$. That is, $[\![Ft \vee Bt]\!]_{\mathcal{M}} = T$.

Thus, $\{\forall x(Tx \supset (Fx \vee Bx)), Tt\} \models Ft \vee Bt$.

1.3. Review of First-Order Logic

(P1) No terrapins live in saltwater.
(P2) Testudo is a terrapin.
(C) Testudo lives in saltwater.

(P1) $\neg\exists x(Tx \wedge Sx)$
(P2) Tt
(C) St

This argument is logically invalid.

Countermodel: $\mathcal{D} = \{\text{Testudo}\}$, $\mathcal{I}(t) = \text{Testudo}$, $\mathcal{I}(T) = \{\text{Testudo}\}$, and $\mathcal{I}(S) = \emptyset$.

Thus, $\{\neg\exists x(Tx \wedge Sx), Tt\} \not\models St$.

1.3. Review of First-Order Logic

Logicians have also developed elegant proof systems for establishing validity without having to unpack the semantic clauses in Def 1.17.

In the Fitch-style system $\mathcal{F}$ in Barwise and Etchemendy [1999], we can prove $Ft \vee Bt$ from $\{\forall x(Tx \supset (Fx \vee Bx)), Tt\}$ as follows:

1	$\forall x(Tx \supset (Fx \vee Bx))$	Premise
2	Tt	Premise
3	$Tt \supset (Ft \vee Bt)$	$\forall$ Elim:1
4	$Ft \vee Bt$	$\supset$ Elim: 3,2

Def 1.20. $\{\varphi_1, ..., \varphi_n\} \vdash_{\mathcal{F}} \psi$ just in case there is a proof in $\mathcal{F}$ that starts with premises $\varphi_1, ..., \varphi_n$ and concludes with ψ.

$\{\forall x(Tx \supset (Fx \vee Bx)), Tt\} \vdash_{\mathcal{F}} Ft \vee Bt$.

1.3. Review of First-Order Logic

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

Thm 1.1 (Soundness Theorem for $\mathcal{F}$). $\{\varphi_1, ..., \varphi_n\} \vdash_{\mathcal{F}} \psi$ only if $\{\varphi_1, ..., \varphi_n\} \models \psi$.

Thm 1.2 (Completeness Theorem for $\mathcal{F}$). $\{\varphi_1, ..., \varphi_n\} \models \psi$ only if $\{\varphi_1, ..., \varphi_n\} \vdash_{\mathcal{F}} \psi$.

By Soundness, $\{\forall x(Tx \supset (Fx \vee Bx)), Tt\} \models Ft \vee Bt$.

1.3. 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 $\models$ in Def 1.19 extensionally adequate?

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

1.4. Generalized Quantifiers

In $\mathcal{L}_{FOL}$, a quantifier Qx operates on a single formula. If the quantifier binds all of the free occurrences of x in its scope, then the result is a sentence.

(1) Some farmer sweats.

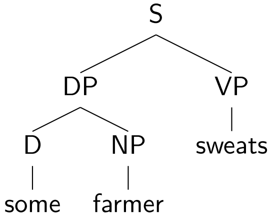
$\exists x(Fx \wedge Sx)$

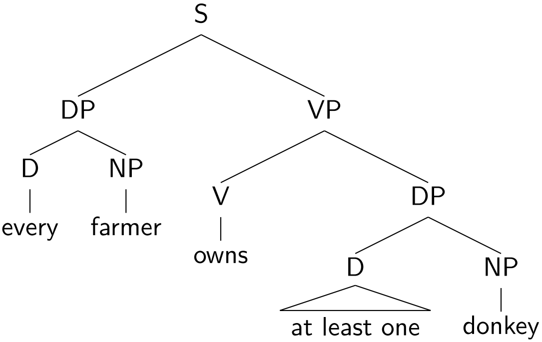
(2) Every farmer owns at least one donkey.

$\forall x(Fx \supset \exists y(Dy \wedge Oxy))$

1.4. Generalized Quantifiers

But natural language suggests a different approach. Grammatically, quantifiers like 'every', 'some', 'most', 'three' are *determiners* that modify a noun phrase. The resulting determiner phrase can combine with a verb phrase to form a sentence:





1.4. Generalized Quantifiers

This phrase structure suggests that a quantifier Qx should operate on *two* formulae.

(1) Some farmer sweats.

$some_x(Fx, Sx)$

(2) Every farmer owns at least one donkey.

$all_x(Fx, some_y(Dy, Oxy))$

(3) At least two donkeys are thirsty.

$at\text{-}least\text{-}two_x(Dx, Tx)$

(4) Most farmers are poor.

$most_x(Fx, Px)$

1.4. Generalized Quantifiers

The semantics of binary quantifiers is straightforward:

Def 1.21. *Truth-in-a-model under an assignment* for formulae involving the above binary quantifiers is determined with the following recursive clauses:

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

1.4. Generalized Quantifiers

It might seem that the difference between unary and binary quantification is not that significant since binary quantifiers can be defined in terms of $\forall$ and $\exists$ (bracketing off vague quantifiers like *few* and context-sensitive quantifiers like *enough*).

$$\mathit{some}_x(\varphi(x), \psi(x)) \equiv \exists x(\varphi(x) \wedge \psi(x))$$

$$\mathit{all}_x(\varphi(x), \psi(x)) \equiv \forall x(\varphi(x) \supset \psi(x))$$

$$\mathit{at\text{-}least\text{-}two}_x(\varphi(x), \psi(x)) \equiv \exists x \exists y(x \neq y \wedge \varphi(x) \wedge \varphi(y) \wedge \psi(x) \wedge \psi(y))$$

But what about most_x (read: $\mathit{more\text{-}than\text{-}half}_x$)? Can this be defined using the standard unary quantifiers?

1.4. Generalized Quantifiers

Barwise and Cooper [1981] prove that there is no sentence in $S_{\mathcal{L}_{FOL}}$ that is equivalent to $most_{x_i}(\varphi, \psi)$.

In fact, they prove something stronger. If we add the unary quantifier M to $\mathcal{L}_{FOL}$ where

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

$most_{x_i}(\varphi, \psi)$ is still not definable in the expanded language.

It is often concluded that 'Most' is *essentially binary*. But Barwise and Cooper have shown only that $most_{x_i}(\varphi, \psi)$ is not capturable by M.

1.4. Generalized Quantifiers

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

If 'Most' is logical vocabulary, then arguments like this reveal that the formal consequence relation $\models$ in Def 1.19 does not explicate our informal concept of validity.

Some other quantifiers that are not definable in $\mathcal{L}_{FOL}$:

'There are at least as many...as...'

'There are finitely many...'

'There are infinitely many...'

'The are an even number of...'

See Barwise and Feferman [1985].

1.5. Substitutional Quantifiers

So far, we have been loose about the *mention* of an expression like ‘it’ and the *use* of it.

When mentioning an expression in this section, we will use ordinary quotation marks and Quinean corner quotes.

Maryland is a state. ‘Maryland’ is the name of this state. “Maryland” denotes the name of the state of Maryland. And so forth.

If ‘φ’ and ‘ψ’ denote the sentences ‘College Park is in Maryland’ and ‘Baltimore is in Maryland’ respectively, then ‘$\ulcorner \varphi \wedge \psi \urcorner$’ formalizes the sentence ‘College Park and Baltimore are in Maryland’.

$\ulcorner \varphi \wedge \psi \urcorner = \varphi \frown \text{‘} \wedge \text{’} \frown \psi$.

1.5. Substitutional Quantifiers

Armed with this distinction between use and mention, we can also distinguish between *objectual* '$\forall$','$\exists$' and *substitutional* 'Π','Σ' quantifiers.

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

'$\Pi x Px$' and '$\Sigma x Px$' involve substitutional quantification. To evaluate these sentences for truth, we evaluate the embedded open formula 'Px' for truth after replacing 'x' with closed terms in $\mathcal{L}$.

1.5. Substitutional Quantifiers

$$\begin{array}{lll} [\![\ulcorner\forall x_i\varphi\urcorner]\!]^g_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{g[x_i\rightarrow d]}_{\mathcal{M}} = T \text{ for every } d \in \mathcal{D} \\ [\![\ulcorner\exists x_i\varphi\urcorner]\!]^g_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^{g[x_i\rightarrow d]}_{\mathcal{M}} = T \text{ for some } d \in \mathcal{D} \end{array}$$

Compare:

Def 1.22. *Truth-in-a-model under an assignment* for formulae involving substitutional quantifiers is determined with these recursive clauses:

$$\begin{array}{lll} [\![\ulcorner\Pi x_i\varphi\urcorner]\!]^g_{\mathcal{M}} = T & \text{iff} & [\![\varphi[x_i \rightarrow t']]\!]^g_{\mathcal{M}} = T \text{ for every closed term } t' \text{ in } \mathcal{L} \\ [\![\ulcorner\Sigma x_i\varphi\urcorner]\!]^g_{\mathcal{M}} = T & \text{iff} & [\![\varphi[x_i \rightarrow t']]\!]^g_{\mathcal{M}} = T \text{ for some closed term } t' \text{ in } \mathcal{L} \end{array}$$

where $\varphi[x_i \rightarrow c]$ is the result of replacing each occurrence of 'x_i' in φ with the closed term t'.

1.5. Substitutional Quantifiers

Consider a language with only '0', '1', and '$\times$'. Given the standard model,

$[\![\ulcorner \Pi x_i(x_i \times x_i = x_i)\urcorner]\!]_{\mathcal{M}} = T$ but $[\![\ulcorner \forall x_i(x_i \times x_i = x_i)\urcorner]\!]_{\mathcal{M}} = F$.

$[\![\ulcorner \exists x_i(x_i \times x_i \neq x_i)\urcorner]\!]_{\mathcal{M}} = T$ but $[\![\ulcorner \Sigma x_i(x_i \times x_i \neq x_i)\urcorner]\!]_{\mathcal{M}} = F$.

1.5. Substitutional Quantifiers

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.

1.5. Substitutional Quantifiers

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

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 '$\exists x(\neg(x$ exists$))$'.

1.5. Substitutional Quantifiers

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

(P1) $\neg$(p exists)
(C) $\Sigma x \neg$(x exists)

We must still understand (P1) in such a way that this premise comes out true. We might treat '$\neg$(p exists)' as an unstructured sentence that does not have ordinary constant-predicate form.

If the argument involving 'Σ' is logically valid, then we must move beyond first-order logic.

1.5. Substitutional Quantifiers

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 'Σx(James thinks x is a great writer)'.

1.6. Plural Quantifiers

The language of first-order logic $\mathcal{L}_{FOL}$ has other expressive limitations.

(5) Some critics admire only one another. (Frege-Geach)

This sentence can be symbolized using second-order quantifiers:

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

However, this sentence cannot be formalized in $\mathcal{L}_{FOL}$. We will later prove that (5) is *nonfirstorderizable*.

1.6. Plural Quantifiers

Some other examples discussed by Boolos [1984]:

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

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

(7) 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 these sentences can enter into deductive arguments, they provide another motivation for moving beyond first-order logic.

1.6. Plural Quantifiers

There is a fine line between the firstorderizable and nonfirstorderizable.

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

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

(9) 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.

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

1.6. Plural Quantifiers

The nonfirstorderizable sentences suggest that we should add second-order quantifiers to our language. But how should we interpret the quantified statement $\exists X Xa$?

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 $\exists X Xa$ 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.

1.6. Plural Quantifiers

Boolos [1994] resists such interpretations.

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

We sometimes want to say things about sets such as 'There are some sets of which every set that is not a member of itself is one'. But such sentences are problematic if they entail the existence of a set of all sets.

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

1.6. Plural Quantifiers

"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." ([1998], p. 72)

NASSLLI 2014

Recursion Theory, Gödel's Incompleteness Theorems, and Nonfirstorderizability

Justin Bledin, Johns Hopkins University

2.1. Computability Theory

To prove Gödel's Incompleteness Theorems, we will some background in recursion or computability theory.

Def 2.1. A function $f : \mathbb{N}^n \mapsto \mathbb{N}$ is *effectively computable* just in case there is a finite list of instructions that in principle make it possible to determine the value $f(x_1, ..., x_n)$ for each input.

Effective computability is an intuitive informal notion. But logicians have developed various equivalent formal notions—Turing computability, recursive computability, abacus computability—that purportedly capture this intuitive notion.

We will focus exclusively on recursive computability.

The thesis that effective and recursive computability extensionally coincide is *Church's Thesis*.

2.1. Computability Theory

zero function: $z(x) = 0$

successor function: $s(x) = x + 1$

identity/projection functions: $id_i^n(x_1, ..., x_i, ...x_n) = x_i$

These functions are the *basic functions*.

composition: $h(x_1, ..., x_n) = f(g_1(x_1, ..., x_n), ..., g_m(x_1, ..., x_n))$

(primitive) recursion: $h(x, 0) = f(x)$, $h(x, s(y)) = g(x, y, h(x, y))$

For example, addition and multiplication are defined recursively as follows:

$x + 0 = x$, $x + s(y) = s(x + y)$

$x \times 0 = x$, $x \times s(y) = x + (x \times y)$

Def 2.2. The functions obtainable from the basic functions by composition and recursion are the *primitive recursive functions*.

2.1. Computability Theory

minimization:

$$h(x_1, ..., x_n) = \begin{cases} y & f(x_1, ..., x_n, y) = 0 \text{ and for all } t < y, \\ & f(x_1, ..., x_n, y) \text{ is defined and } \neq 0 \\ \text{undefined} & \text{if there is no such } y \end{cases}$$

Def 2.3. The (total and partial) functions obtainable from the basic functions by composition, recursion, and minimization are the *recursive functions*.

2.1. Computability Theory

Def 2.4. A set $X \subseteq \mathbb{N}^n$ is *effectively decidable* if and only if its characteristic function is effectively computable.

Def 2.5. A set $X \subseteq \mathbb{N}^n$ is *recursively decidable*, or *recursive*, if and only if its characteristic function is recursive.

By Church's Thesis, a set is effectively decidable if and only if this set is recursive.

2.2. Gödel's First Incompleteness Theorem

Time to introduce the two main ingredients required to prove Gödel's Incompleteness Theorems.

The first ingredient is a coding from formal logical syntax to the natural numbers. One of Gödel's nice ideas was that by coding linguistic expressions as natural numbers, a theory of arithmetic can discuss its own syntax.

Def 2.6. A *Gödel numbering* $g : \mathcal{E} \mapsto \mathbb{N}$ is a 1-1 effectively computable function from a set $\mathcal{E}$ of expressions to natural numbers such that it is effectively decidable whether a natural number is the *Gödel number* of some expression in $\mathcal{E}$ and, if so, it is effectively computable which expression it is the Gödel number of.

2.2. Gödel's First Incompleteness Theorem

Let us set up a Gödel numbering for the *language of arithmetic* $\mathcal{L}_{AR}$ with the following symbols: 0, $'$, $+$, $\times$, logical symbols, and parentheses.

There are many ways to do this. The method presented here is based on Boolos, Burgess, and Jeffrey [2002]:

Symbol	(	)	$\perp$	$\neg$	$\wedge$	$\vee$	$=$	0	$'$	$+$	$\times$	$\forall$	x_i
Code	1	3	5	7	9	11	13	15	17	19	21	23	2×5^i

Let $c(s_i)$ designate the code of symbol s_i.

If $e = s_1 \frown \ldots \frown s_n$, then $g(e) = p_1^{c(s_1)} \times \ldots \times p_n^{c(s_n)}$ where p_i is the i-th prime number.

For example, $g(0 \times 0' = 0) = 2^{15} \times 3^{21} \times 5^{15} \times 7^{17} \times 11^{13} \times 13^{15}$.

2.2. Gödel's First Incompleteness Theorem

The second ingredient is the *representability* of recursive functions in Robinson Arithmetic.

Let $\overline{m}$ be the numeral in $\mathcal{L}_{AR}$ corresponding to $m \in \mathbb{N}$.

Def 2.7 An *n*-ary function $f : \mathbb{N}^n \mapsto \mathbb{N}$ is *representable* in a theory of arithmetic T if there is a formula $\varphi(x_1, ..., x_{n+1})$ in $\mathcal{L}_{AR}$ such that if $f(m_1, ..., m_n) = m_{n+1}$ then $\vdash_T \forall x_{n+1}(\varphi(\overline{m_1}, ..., x_{n+1}) \equiv x_{n+1} = \overline{m_{n+1}})$

where $\vdash_T \varphi$ just in case φ is provable from the axioms of T in a proof system for FOL.

The formula $\varphi(x_1, ..., x_{n+1})$ *represents* f in T.

Intuitively, T can 'talk about' f with φ.

2.2. Gödel's First Incompleteness Theorem

Def 2.8. *Robinson Arithmetic* Q has the following axioms:

(Q1) $\forall x(s(x) \neq 0)$
(Q2) $\forall x \forall y(s(x) = s(y) \supset x = y)$
(Q3) $\forall x(x = 0 \vee \exists y(x = s(y)))$
(Q4) $\forall x(x + 0 = x)$
(Q5) $\forall x \forall y(x + s(y) = s(x + y))$
(Q6) $\forall x(x \times 0 = 0)$
(Q7) $\forall x \forall y(x \times s(y) = x \times y + x)$

Q is a consistent theory: (Q1)-(Q7) are all true in the standard model of arithmetic.

Thm 2.1. The recursive functions are representable in Q.

See Boolos, Burgess, and Jeffrey [2002] for proof.

2.2. Gödel's First Incompleteness Theorem

With these ingredients in hand, we can prove the engine of Gödel's proofs.

Let $\ulcorner e \urcorner = \overline{g(e)}$.

Def 2.9. The *diagonalization* of an expression e is the expression $\exists x(x = \ulcorner e \urcorner \wedge e)$.

Lem 2.1. There is a recursive function $d : \mathbb{N} \mapsto \mathbb{N}$ mapping the Gödel number of an expression to the Gödel number of its diagonalization.

The function d is effectively computable so recursive by Church's Thesis.

Thm 2.2 (Diagonalization Theorem). Let T be a theory of arithmetic in which d is representable. Then for any formula $\varphi(x)$ containing just the variable x free, there is a sentence $G \in S_{\mathcal{L}_{AR}}$ such that $\vdash_T G \equiv \varphi(\ulcorner G \urcorner)$.

2.2. Gödel's First Incompleteness Theorem

Proof of Thm 2.2. Let Dxy represent d in T. That is, if $d(m_1) = m_2$ then $\vdash_T \forall y(D\overline{m_1}y \equiv y = \overline{m_2})$.

Consider the expression $e : \exists y(Dxy \wedge \varphi(y))$.

Let $G : \exists x(x = \ulcorner e \urcorner \wedge \exists y(Dxy \wedge \varphi(y)))$.

Note that G is the diagonalization of e, so $d(g(e), g(G))$.

$\vdash_T G \equiv \exists y(D\ulcorner e \urcorner y \wedge \varphi(y))$.

Since $\vdash_T \forall y(D\ulcorner e \urcorner y \equiv y = \ulcorner G \urcorner)$, $\vdash_T G \equiv \exists y(y = \ulcorner G \urcorner \wedge \varphi(y))$.

Thus, $\vdash_T G \equiv \varphi(\ulcorner G \urcorner))$.

The Diagonalization or Fixed Point Theorem is often given this gloss: as G holds whenever $\varphi(\ulcorner G \urcorner)$ does *modulo* T, the sentence G 'says' that the property expressed by $\varphi(x)$ holds of itself.

2.2. Gödel's First Incompleteness Theorem

The Diagonalization Theorem can be used to prove a number of important results. We first need a few more definitions:

Def 2.10. *Arithmetic* is the theory T_{AR} whose theorems are all of the truths in the standard model of arithmetic.

Def 2.11. T is *consistent* iff $\nvdash_T 0 = \overline{1}$.

Def 2.12. T_1 *extends* T_2 iff every theorem of T_2 is a theorem of T_1: $\vdash_{T_2} \varphi$ only if $\vdash_{T_1} \varphi$.

T_{AR} is consistent and extends Q.

2.2. Gödel's First Incompleteness Theorem

Def 2.13. A set of expressions $\mathcal{E}$ is *recursively decidable* if and only if the set of their Gödel numbers is recursively decidable.

A theory T is *recursively decidable* if and only if the set of its theorems is recursively decidable.

Def 2.14. A set $X \subseteq \mathbb{N}$ is *definable in T* if and only if there is a formula $\varphi(x)$ containing just the variable x free such that:

$m \in X$ only if $\vdash_T \varphi(\overline{m})$
$m \notin X$ only if $\vdash_T \neg\varphi(\overline{m})$

2.2. Gödel's First Incompleteness Theorem

Thm 2.3. No consistent extension of Q (incl. Q and T_{AR}) is recursively decidable.

Proof of Thm 2.3. Suppose that T extends Q and T is consistent and recursively decidable.

Since T is recursively decidable, the set X of Gödel numbers of its theorems is recursively decidable.

Since T extends Q, the recursive characteristic function c_X of X is representable in T. That is, there is a formula $\varphi(x_1, x_2)$ in $\mathcal{L}_{AR}$ such that if $c_x(m_1) = m_2$ then $\vdash_T \forall x(\varphi(\overline{m_1}, x) \equiv x = \overline{m_2})$.

$\varphi(x, 1)$ *defines* the set X in T.

2.2. Gödel's First Incompleteness Theorem

By the Diagonalization Theorem, there is a sentence G such that $\vdash_T G \equiv \varphi(\ulcorner G \urcorner, 0)$.

$\not\vdash_T G$ only if $c_X(g(G)) = 0$ only if $\vdash_T \varphi(\ulcorner G \urcorner, 0)$ only if $\vdash_T G$.

$\vdash_T G$ only if $c_X(g(G)) = 1$ only if $\vdash_T \forall x(\varphi(\ulcorner G \urcorner, x) \equiv x = \overline{1})$.

$\vdash_T G$ only if $\vdash_T \varphi(\ulcorner G \urcorner, 0)$.

Thus, $\vdash_T G$ only if $\vdash_T 0 = \overline{1}$.

2.2. Gödel's First Incompleteness Theorem

Thm 2.4 (Church's Undecidability Theorem). The set of validities in $\mathcal{L}_{AR}$ is undecidable.

Proof of Thm 2.4. Let C be the conjunction of (Q1)-(Q7).

$\vdash_{\mathsf{Q}} \varphi$ iff $\vdash C \supset \varphi$.

Moreover, there is a recursive function $f : \mathbb{N} \mapsto \mathbb{N}$ mapping the Gödel number of each sentence $\varphi \in S_{\mathcal{L}_{AR}}$ to the Gödel number of $C \supset \varphi$.

Suppose that validity is recursively decidable—that is, the set X of Gödel numbers of validities is recursive decidable.

Then the set $\{m : f(m) \in X\}$ of Gödel numbers of theorems of Q is recursively decidable.

This contradicts Thm 2.3.

2.2. Gödel's First Incompleteness Theorem

Thm 2.5 (Tarski's Undefinability Theorem). The set of Gödel numbers of arithmetic truths is not definable in T_{AR}.

Proof of Thm 2.5. Suppose that arithmetic truth is definable in T_{AR} with the formula $Tr(x)$.

By the Diagonalization Theorem, there is a sentence G such that $\vdash_{T_{AR}} G \equiv \neg Tr(\ulcorner G \urcorner)$.

$\vdash_{T_{AR}} G$ only if $\vdash_{T_{AR}} Tr(\ulcorner G \urcorner)$ only if $\vdash_{T_{AR}} \neg G$ only if $\nvdash_{T_{AR}} G$.

$\nvdash_{T_{AR}} G$ only if $\vdash_{T_{AR}} \neg Tr(\ulcorner G \urcorner)$ only if $\vdash_{T_{AR}} G$.

2.2. Gödel's First Incompleteness Theorem

Finally, we can prove Gödel's Incompleteness Theorems. Again, we first need some more definitions:

Def 2.15. A theory T is *complete* iff $\vdash_T \varphi$ or $\vdash_T \neg\varphi$ for all $\varphi \in S_{\mathcal{L}_{AR}}$.

Def 2.16. A theory T is *axiomatizable* iff there is a decidable set of expressions that entail all of the theorems of T. If this decidable set is finite, then T is *finitely axiomatizable*.

By definition, any decidable theory is axiomatizable.

Q is finitely axiomatizable but not decidable.

Def 2.17. A theory T is ω-*inconsistent* iff for some formula $\varphi(x)$ containing just the variable x free, $\vdash_T \varphi(\overline{m})$ for all $m \in \mathbb{N}$ but $\vdash_T \exists x(\neg\varphi(x))$. T is ω-*consistent* iff T is not ω-inconsistent.

ω-consistency implies consistency.

ω-inconsistency does not imply inconsistency.

2.2. Gödel's First Incompleteness Theorem

Lem 2.2. For any axiomatizable theory T, there is a recursively decidable relation $Proof_T \subset \mathbb{N} \times \mathbb{N}$ where $Proof_T(m_1, m_2)$ just in case m_1 is the Gödel number of a proof in T of the sentence with Gödel number m_2.

The characteristic function c_{Proof_T} of $Proof_T$ is effectively computable so recursive by Church's Thesis.

Thm 2.6 (Gödel's First Incompleteness Theorem). Any ω-consistent axiomatizable extension of Q is incomplete.

2.2. Gödel's First Incompleteness Theorem

Proof of Thm 2.6. Suppose that T extends Q and T is ω-consistent and axiomatizable.

Since T extends Q and T is axiomatizable, c_{Proof_T} is representable in T. That is, there is a formula $\varphi(x_1, x_2, x_3)$ in $\mathcal{L}_{AR}$ such that if $c_{Proof_T}(m_1, m_2) = m_3$ then $\vdash_T \forall x(\varphi(\overline{m_1}, \overline{m_2}, x) \equiv x = \overline{m_3})$.

Let $Prov(y) : \exists x \varphi(x, y, \overline{1})$.

By the Diagonalization Theorem, there is a sentence G such that $\vdash_T G \equiv \neg Prov(\ulcorner G \urcorner)$.

$\vdash_T G$ only if $\vdash_T Prov(\ulcorner G \urcorner)$ only if $\vdash_T \neg G$ only if $\vdash_T 0 = \overline{1}$.

$\vdash_T \neg G$ only if $\vdash_T Prov(\ulcorner G \urcorner)$ only if $\vdash_T G$ only if $\vdash_T 0 = \overline{1}$.

(Note: $\vdash_T Prov(\ulcorner G \urcorner)$ only if $\vdash_T G$ because T is ω-consistent)

Thus, $\nvdash_T G$ and $\nvdash_T \neg G$.

2.2. Gödel's First Incompleteness Theorem

Rosser [1936] strengthens Gödel's result:

Thm 2.7. Any *consistent* axiomatizable extension of Q is incomplete.

Since T_{AR} is a complete consistent extension of Q, a corollary of Gödel's First Incompleteness Theorem is that arithmetic is not axiomatizable.

2.3. Gödel's Second Incompleteness Theorem

Def 2.18. A formula $Pr(x)$ containing just the variable x free is a *provability predicate for* T just in case it meets the following three conditions:

(P1) if $\vdash_T \varphi$ then $\vdash_T Pr(\ulcorner\varphi\urcorner)$
(P2) $\vdash_T Pr(\ulcorner\varphi \supset \psi\urcorner) \supset (Pr(\ulcorner\varphi\urcorner) \supset Pr(\ulcorner\psi\urcorner))$
(P3) $\vdash_T Pr(\ulcorner\varphi\urcorner) \supset Pr(\ulcorner Pr(\ulcorner\varphi\urcorner)\urcorner)$

$Prov(y)$ is a provability predicate.

But there are other provability predicates besides: $\exists x\varphi(x, y, \overline{1}) \wedge 0 = 0$.

Recall that when T is ω-consistent, $Prov(y)$ also satisfies this fourth condition:

(P4) if $\vdash_T Pr(\ulcorner\varphi\urcorner)$ then $\vdash_T \varphi$

2.3. Gödel's Second Incompleteness Theorem

To motivate Gödel's Second Incompleteness Theorem, recall the core of the proof of his first theorem:

$\vdash_T G \xrightarrow{(P1)} \vdash_T Pr(\ulcorner G \urcorner) \xrightarrow{(DL)} \vdash_T \neg G \rightarrow \vdash_T 0 = \overline{1}.$

$\vdash_T \neg G \xrightarrow{(DL)} \vdash_T Pr(\ulcorner G \urcorner) \xrightarrow{(P4)} \vdash_T G \rightarrow \vdash_T 0 = \overline{1}.$

Question: Can't this simple argument be run inside of T? In particular, can't T prove that G and $\neg G$ are not provable?

Surely not: $\vdash_T \neg Pr(\ulcorner G \urcorner) \xrightarrow{(DL)} \vdash_T G \rightarrow \vdash_T 0 = \overline{1}.$

But why not?

Answer: Because T cannot 'know' that it is consistent. *We* know that T is consistent, so we can conclude from the implication $\vdash_T G \rightarrow \vdash_T 0 = \overline{1}$ that $\nvdash_T G$. But T cannot conclude this.

2.3. Gödel's Second Incompleteness Theorem

Thm 2.8 (Gödel's Second Incompleteness Theorem). If T is a consistent extension of Q with a provability predicate Pr, then $\nvdash_T \neg Pr(\ulcorner 0 = \overline{1} \urcorner)$.

Proof of Thm 2.8.

$\vdash_T G \equiv \neg Pr(\ulcorner G \urcorner)$ [DL]
$\vdash_T Pr(\ulcorner G \urcorner) \supset \neg G$ [contraposition]
$\vdash_T Pr(\ulcorner Pr(\ulcorner G \urcorner) \urcorner) \supset Pr(\ulcorner \neg G \urcorner)$ [P1 and P2]
$\vdash_T Pr(\ulcorner G \urcorner) \supset Pr(\ulcorner Pr(\ulcorner G \urcorner) \urcorner)$ [P3]
$\vdash_T Pr(\ulcorner G \urcorner) \supset Pr(\ulcorner \neg G \urcorner)$ [transitivity]
$\vdash_T Pr(\ulcorner G \urcorner) \supset (Pr(\ulcorner \neg G \urcorner) \supset Pr(\ulcorner 0 = \overline{1} \urcorner))$ [P1 and P2]
$\vdash_T Pr(\ulcorner G \urcorner) \supset Pr(\ulcorner 0 = \overline{1} \urcorner)$ [modus ponens]
$\vdash_T \neg Pr(\ulcorner 0 = \overline{1} \urcorner) \supset \neg Pr(\ulcorner G \urcorner)$ [contraposition]

Thus, if $\vdash_T \neg Pr(\ulcorner 0 = \overline{1} \urcorner)$ then $\vdash_T \neg Pr(\ulcorner G \urcorner)$ and T is inconsistent.

2.3. Gödel's Second Incompleteness Theorem

This is truly a beautiful proof. Like us, T 'knows' that if G is provable then $\neg G$ is provable, so T 'knows' that if G is provable then T is inconsistent. But this is where things stop. We can finish the proof of Gödel's First Incompleteness Theorem, but T must remain ignorant of its own consistency.

2.4. Nonfirstorderizability

Let us close with a nice application of Gödel's First Incompleteness Theorem.

Recall that the following sentence is nonfirstorderizable:

(1) Some critics admire only one another. (Frege-Geach)

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

Boolos [1984] discusses an elegant proof by Kaplan that (1) has no first-order equivalent.

2.4. Nonfirstorderizability

Def 2.19. *Peano Arithmetic* PA has the following axioms:

(PA1) $\forall x(s(x) \neq 0)$
(PA2) $\forall x \forall y(s(x) = s(y) \supset x = y)$
(PA3) $\forall x(x + 0 = x)$
(PA4) $\forall x \forall y(x + s(y) = s(x + y))$
(PA5) $\forall x(x \times 0 = 0)$
(PA6) $\forall x \forall y(x \times s(y) = x \times y + x)$

Together with this induction schema where $\varphi(x_1, ..., x_{n+1})$ is a formula in $\mathcal{L}_{AR}$ containing the variables $x_1, ..., x_{n+1}$ free:

(PA7) $\forall x_1 ... x_n((\varphi(x_1, ..., x_n, 0) \wedge \forall y(\varphi(x_1, ..., x_n, y) \supset \varphi(x_1, ..., x_n, s(y)))) \supset \forall y(\varphi(x_1, ..., x_n, y)))$

PA is a consistent theory: (PA1)-(PA7) are all true in the standard model of arithmetic.

2.4. Nonfirstorderizability

But Gödel's First Incompleteness Theorem implies that (PA1)-(PA7) are also true in *nonstandard models*.

Since $\nvdash_{\mathrm{PA}} G$ and $\nvdash_{\mathrm{PA}} \neg G$ for some $G \in S_{\mathcal{L}_{AR}}$, both $\mathrm{PA} \cup \{G\}$ and $\mathrm{PA} \cup \{\neg G\}$ are consistent, so both sets of expressions have models.

G is true in the standard natural number structure.

G is false in nonstandard models of PA that consist of the natural number structure followed by copies of the integers.

2.4. Nonfirstorderizability

Back to nonfirstorderizability:

(1) Some critics admire only one another. (Frege-Geach)

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

Let $Axy : x = 0 \vee x = y + 1$

$\exists X(\exists x Xx \wedge \forall x \forall y((Xx \wedge (x = 0 \vee x = y + 1)) \supset (x \neq y \wedge Xy)))$

Note that $0 \notin X$ and X must include the immediate predecessor of each of its members.

This sentence is true in all nonstandard models.

This sentence is false in the standard model.

Thus, if (1) has a first-order equivalent, we can add its negation to PA to axiomatize T_{AR}, contradicting Gödel's First Incompleteness Theorem.

NASSLLI 2014

Modal Logic and Its Applications

Justin Bledin, Johns Hopkins University

3.1. Propositional Modal Logic

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...
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 on. 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.

3.1. Propositional Modal Logic

Def 3.1. The *basic sentential modal language* $\mathcal{L}_{SENT_\Diamond}$ 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 $\neg$.

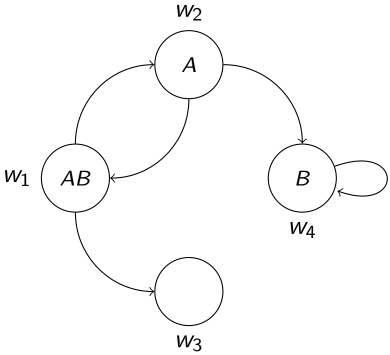
$At_{\mathcal{L}_{SENT_\Diamond}} = \{A, B, ...\}$ is the set of atoms in $\mathcal{L}_{SENT_\Diamond}$.
$S_{\mathcal{L}_{SENT_\Diamond}}$ is the set of well-formed sentences in $\mathcal{L}_{SENT_\Diamond}$.

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

3.1. Propositional Modal Logic

Def 3.2. 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 $\mathcal{V} : At_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \mapsto \{T, F\}$ mapping each sentence letter $p \in At_{\mathcal{L}_{SENT_\Diamond}}$ and world $w \in \mathcal{W}$ to a truth value.

3.1. Propositional Modal Logic



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

3.1. Propositional Modal Logic

Def 3.3. 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:

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

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

Given the previous model, $[\![B]\!]^{w_2}_{\mathcal{M}} = F$, $[\![\Diamond(A \wedge B)]\!]^{w_2}_{\mathcal{M}} = T$, and $[\![\Box B]\!]^{w_2}_{\mathcal{M}} = T$.

3.1. Propositional Modal Logic

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

Def 3.4. The argument from premises $\varphi_1, ..., \varphi_n$ to conclusion ψ is *logically valid*, $\{\varphi_1, ..., \varphi_n\} \models \psi$, just in case there is no pointed model $\mathcal{M}, w$ such that $[\![\varphi_1]\!]^w_{\mathcal{M}} = ... = [\![\varphi_n]\!]^w_{\mathcal{M}} = T$ and $[\![\psi]\!]^w_{\mathcal{M}} = F$.

Def 3.5. The sentence φ is a *logical validity*, $\models \varphi$, just in case there is no pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = F$.

Def 3.6. 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.

3.1. Propositional Modal Logic

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

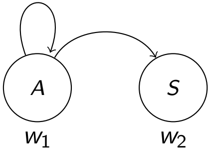
By the semantic clause for $\Box$, $[\![\Box(A \wedge S)]\!]^w_{\mathcal{M}} = T$.

3.1. Propositional Modal Logic

(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 invalid. Countermodel:



3.1. Propositional Modal Logic

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 3.7. 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.

3.1. Propositional Modal Logic

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

Def 3.8. 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 $p \in At_{\mathcal{L}_{SENT_\Diamond}}$, $\mathcal{V}^{\mathcal{M}}(p, x) = \mathcal{V}^{\mathcal{N}}(p, y)$.

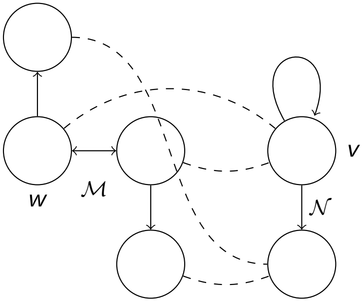
(zig) If $x\mathcal{R}^{\mathcal{M}}z$, then there exists $z' \in \mathcal{W}^{\mathcal{N}}$ such that $y\mathcal{R}^{\mathcal{N}}z'$ and $z\mathcal{Z}z'$.

(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*: $\mathcal{M}, w \leftrightarroweq \mathcal{N}, v$.

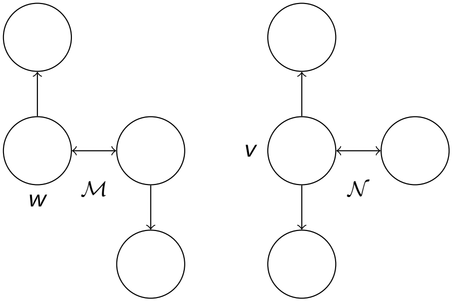
3.1. Propositional Modal Logic

These pointed models are bisimilar:



3.1. Propositional Modal Logic

These pointed models are not:



3.1. Propositional Modal Logic

Lem 3.1 (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 $[\![\varphi]\!]^w_{\mathcal{M}} \neq [\![\varphi]\!]^v_{\mathcal{N}}$.

For the previous non-bisimilar pointed models, $[\![\Box(\Box\bot \vee \Diamond\Box\bot)]\!]^w_{\mathcal{M}} = T$ but $[\![\Box(\Box\bot \vee \Diamond\Box\bot)]\!]^v_{\mathcal{N}} = F$.

3.1. Propositional Modal Logic

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 3.2. For finite pointed models $\mathcal{M}, w$ and $\mathcal{N}, v$, $\mathcal{M}, w \leftrightsquigarrow \mathcal{N}, v$ only if $\mathcal{M}, w \leftrightarrows \mathcal{N}, v$.

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

3.1. Propositional Modal Logic

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.

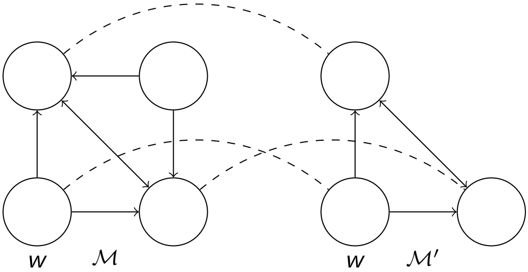
Def 3.9. Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$, $\mathcal{M}' = \langle \mathcal{W}', \mathcal{R}', \mathcal{V}' \rangle$ is a *submodel* of $\mathcal{M}$ if and only if:

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

Def 3.10. Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ and $w \in \mathcal{W}$, the *submodel 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 3.3. 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.

3.1. Propositional Modal Logic



3.1. Propositional Modal Logic

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

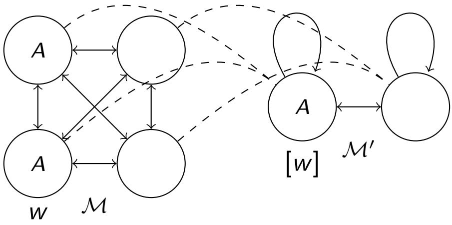
3.1. Propositional Modal Logic

Def 3.11. 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 3.4. If $\mathcal{M}'$ is the bisimulation contraction of $\mathcal{M}$, $\mathcal{M}, w \leftrightarroweq \mathcal{M}', [w]$. The relation sending $x \in \mathcal{W}$ to $[x] \in \mathcal{W}'$ is a bisimulation.

3.1. Propositional Modal Logic



In this figure, the submodel of $\mathcal{M}$ generated from w is just $\mathcal{M}$ itself. So the two kinds of bisimilar submodels that we have been considering differ.

3.1. Propositional Modal Logic

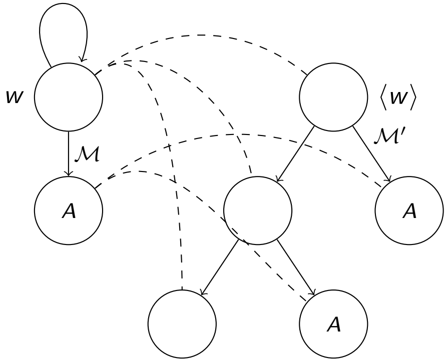
It is also sometimes useful to transform a pointed model into a larger bisimilar pointed model.

Def 3.12. 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:

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

Lem 3.5. 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.

3.1. Propositional Modal Logic



3.1. Propositional Modal Logic

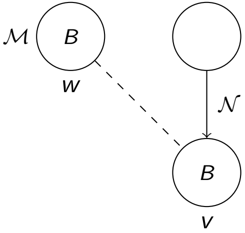
A nice application of the Invariance Lemma is that we can use it to prove that certain operators are undefinable in $\mathcal{L}_{SENT_\Diamond}$.

Lem 3.6. The following universal operator $\mathcal{A}$ is undefinable in $\mathcal{L}_{SENT_\Diamond}$:

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

3.1. Propositional Modal Logic

Proof of Lem 3.6. Suppose that $\mathcal{A}$ is definable in $\mathcal{L}_{SENT_\Diamond}$; that is, 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 $\mathcal{M}, w \leftrightarrows \mathcal{N}, v$, $\mathcal{M}, w \leftrightsquigarrow \mathcal{N}, v$, so $[\![\alpha(B)]\!]^v_\mathcal{N} = [\![\mathcal{A}B]\!]^v_\mathcal{N} = T$.

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

3.1. Propositional Modal Logic

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

Def 3.5. The sentence φ is a *logical validity*, $\models \varphi$, 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 3.13. The sentence φ is *satisfiable* just in case there is a pointed model $\mathcal{M}, w$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = T$.

φ 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, validity in $S_{\mathcal{L}_{SENT_\Diamond}}$ is *decidable*.

3.1. Propositional Modal Logic

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 3.1. 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 3.2. 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 φ.

3.1. Propositional Modal Logic

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, φ is invalid. If not, φ is valid.

3.1. Propositional Modal Logic

Def 3.14. 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 $|w|^{\sim_{\varphi}} = \{v \in \mathcal{W} : w \sim_{\varphi} v\}$.

3.1. Propositional Modal Logic

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

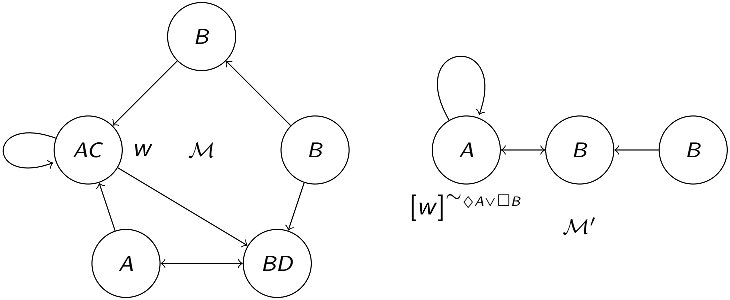
Def 3.15. Given $\mathcal{M} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$, the *filtration* of $\mathcal{M}$ through φ is the model $\mathcal{M}' = \langle \mathcal{W}', \mathcal{R}', \mathcal{V}' \rangle$ where:

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

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

3.1. Propositional Modal Logic

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



3.1. Propositional Modal Logic

Thm 3.3. If $\mathcal{M}'$ is the filtration of $\mathcal{M}$ through φ, then for each $w \in \mathcal{W}$ and $\psi \in sub(\varphi)$, $[\![\psi]\!]^{w}_{\mathcal{M}} = [\![\psi]\!]^{[w]^{\sim_\varphi}}_{\mathcal{M}'}$.

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

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

3.2. The Modal Zoo

Def 3.16. The *minimal modal logic* **K** has the following rules and axioms:

(PL)	All (substitutions of) tautologies are axioms
(MP)	From φ and $\varphi \supset \psi$ infer ψ
(Nec)	From φ infer $\Box\varphi$
(K)	For any φ, ψ, $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$ is an axiom
(Duality)	Expressions involving $\Box$ and $\Diamond$ are interchangeable according to the duality $\Box \equiv \neg\Diamond\neg$

3.2. The Modal Zoo

Def 3.17. 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:

φ_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 φ_i according to Duality)

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

3.2. The Modal Zoo

$\vdash_{\mathbf{K}} \Box(\varphi \wedge \psi) \supset (\Box\varphi \wedge \Box\psi)$

1.	$(\varphi \wedge \psi) \supset \varphi$	PL
2.	$\Box((\varphi \wedge \psi) \supset \varphi)$	Nec 1
3.	$\Box((\varphi \wedge \psi) \supset \varphi) \supset (\Box(\varphi \wedge \psi) \supset \Box\varphi)$	K
4.	$\underline{\Box(\varphi \wedge \psi) \supset \Box\varphi}_{A}$	MP 3,2
5.	$(\varphi \wedge \psi) \supset \psi$	PL
6.	$\Box((\varphi \wedge \psi) \supset \psi)$	Nec 5
7.	$\Box((\varphi \wedge \psi) \supset \psi) \supset (\Box(\varphi \wedge \psi) \supset \Box\psi)$	K
8.	$\underline{\Box(\varphi \wedge \psi) \supset \Box\psi}_{B}$	MP 7,6
9.	$A \supset (B \supset (A \wedge B))$	PL
10.	$B \supset (A \wedge B))$	MP 9,4
11.	$(\underline{\Box(\varphi \wedge \psi)}_{C} \supset \underline{\Box\varphi}_{D}) \wedge (\underline{\Box(\varphi \wedge \psi)}_{C} \supset \underline{\Box\psi}_{E})$	MP 10,8
12.	$((C \supset D) \wedge (C \supset E)) \supset (C \supset (D \wedge E))$	PL
13.	$\Box(\varphi \wedge \psi) \supset (\Box\varphi \wedge \Box\psi)$	MP 12,11

3.2. The Modal Zoo

Thm 3.4 (Soundness Theorem for K). $\vdash_{\mathbf{K}} \varphi$ only if $\models \varphi$.

Thm 3.5 (Completeness Theorem for K). $\models \varphi$ only if $\vdash_{\mathbf{K}} \varphi$.

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

Def 3.18. $\Gamma \vdash_{\mathbf{K}} \psi$ iff there are sentences $\varphi_1, ..., \varphi_n \in \Gamma$ such that $\vdash_{\mathbf{K}} (\varphi_1 \wedge ... \wedge \varphi_n) \supset \psi$.

Thm 3.6 (Strong Soundness/Completeness Theorem for K). $\Gamma \vdash_{\mathbf{K}} \varphi$ iff $\Gamma \models \varphi$.

N.B. Γ can be infinite (Def 3.4 must be extended).

3.2. The Modal Zoo

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	$\varphi \supset \Box\Diamond\varphi$

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.

3.2. The Modal Zoo

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 3.19. 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 3.20. 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 $\mathcal{M}, w$ with $\mathcal{M}$ based on $\mathcal{F}$ such that $[\![\varphi]\!]^w_{\mathcal{M}} = F$.

3.2. The Modal Zoo

The various proof systems extending **K** are sound and complete with respect to validity on different kinds of frames.

Thm 3.7 (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*—that is, $\forall w(w\mathcal{R}w)$.

Thm 3.8 (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*—that is, $\forall w \exists v(w\mathcal{R}v)$.

Thm 3.9 (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*—that is, $\forall w, v, u((w\mathcal{R}v \wedge v\mathcal{R}u) \supset w\mathcal{R}u)$.

3.2. The Modal Zoo

Thm 3.10 (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*—that is, $\forall w, v, u((w\mathcal{R}v \wedge w\mathcal{R}u) \supset v\mathcal{R}u)$.

Thm 3.11 (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*—that is, $\forall w, v(w\mathcal{R}v \supset v\mathcal{R}w)$.

Thm 3.12 (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 3.13 (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.

3.2. The Modal Zoo

How are these logics related?

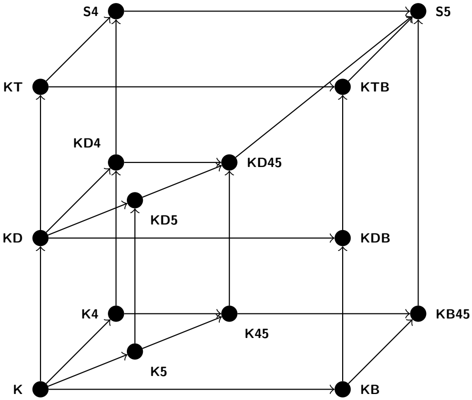
The following partial ordering is based on proof-theoretic strength:

Def 3.21. 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 $\varphi \in S_{\mathcal{L}}$. 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, **KD** $<$ **KT**.

3.2. The Modal Zoo

The full landscape looks like this:



3.2. The Modal Zoo

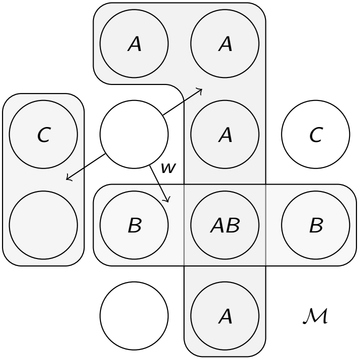
There are also many *non-normal* modal logics that lie below **K**.

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 3.22. 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 $\mathcal{V} : At_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \mapsto \{T, F\}$ 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.

3.2. The Modal Zoo



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

3.2. The Modal Zoo

Def 3.23. $[\varphi]_{\mathcal{M}} = \{w : [\![\varphi]\!]^w_{\mathcal{M}} = T\}$ is the *proposition expressed* by φ in $\mathcal{M}$.

Def 3.24. The following recursive definition of truth lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_\Diamond}} \times \mathcal{W} \mapsto \{T, F\}$ for $\mathcal{L}_{SENT_\Diamond}$:

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

Given the previous model, $[\![\Box A]\!]^w_{\mathcal{M}} = T$, $[\![\Box B]\!]^w_{\mathcal{M}} = T$, but $[\![\Box(A \wedge B)]\!]^w_{\mathcal{M}} = F$.

3.2. The Modal Zoo

Def 3.25. The *minimal non-normal modal logic* **E** has the following rules and axioms:

(PL)	All (substitutions of) tautologies are axioms
(MP)	From φ and $\varphi \supset \psi$ infer ψ
(RE)	From $\varphi \equiv \psi$ infer $\Box\varphi \equiv \Box\psi$
(Duality)	Expressions involving $\Box$ and $\Diamond$ are interchangeable according to the duality $\Box \equiv \neg\Diamond\neg$

Thm 3.14 (Soundness and Completeness Theorem for E).
$\vdash_{\mathbf{E}} \varphi$ iff $\models_{\mathcal{F}} \varphi$ for each neighborhood frame $\mathcal{F} = \langle \mathcal{W}, \mathcal{N} \rangle$.

3.2. The Modal Zoo

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	$\Box\neg\bot$

3.2. The Modal Zoo

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

Thm 3.15 (Soundness and Completeness Theorem for EM).
$\vdash_{\textbf{EM}} \varphi$ 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, $(X \in \mathcal{N}(w) \wedge X \subseteq Y) \supset Y \in \mathcal{N}(w)$.

Thm 3.16 (Soundness and Completeness Theorem for EC).
$\vdash_{\textbf{EC}} \varphi$ 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, $X, Y \in \mathcal{N}(w) \supset X \cap Y \in \mathcal{N}(w)$.

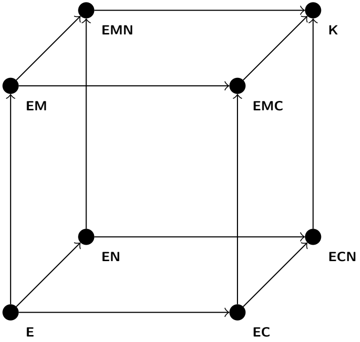
3.2. The Modal Zoo

Thm 3.17 (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, $\mathcal{W} \in \mathcal{N}(w)$.

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**.

3.2. The Modal Zoo

The landscape looks like this:



3.3. Temporal Logic

Def 3.26. 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 φ' and $\mathsf{H}\varphi$ as 'Hitherto φ'.

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 3.27. The *mirror image* of φ is the sentence obtained from φ by switching G and H operators.

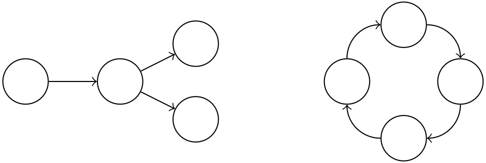
$(\mathsf{F}\varphi \vee \mathsf{H}\varphi) \wedge (\mathsf{G}\varphi \vee \mathsf{P}\varphi)$ is the mirror image of $(\mathsf{P}\varphi \vee \mathsf{G}\varphi) \wedge (\mathsf{H}\varphi \vee \mathsf{F}\varphi)$.

3.3. Temporal Logic

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

Def 3.28. 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).



3.3. Temporal Logic

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 3.29. 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 $\mathcal{L}_t$:

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

3.3. Temporal Logic

Def 3.30. The *minimal temporal logic* $\mathbf{K}_t$ has the following rules and axioms:

(PL)	All (substitutions of) tautologies are axioms
(MP)	From φ and $\varphi \supset \psi$ infer ψ
(TG)	From φ infer Gφ From φ infer Hφ
(DB)	For any φ, ψ, G$(\varphi \supset \psi) \supset (G\varphi \supset$ G$\psi)$ is an axiom For any φ, ψ, H$(\varphi \supset \psi) \supset (H\varphi \supset$ H$\psi)$ is an axiom
(4)	For any φ, G$\varphi \supset$ GGφ is an axiom
(CV)	For any φ, $(\varphi \supset$ GP$\varphi) \wedge (\varphi \supset$ HF$\varphi)$ is an axiom

Thm 3.18. $\mathbf{K}_t$ is sound and complete with respect to the class of all flows of time.

3.3. Temporal Logic

Def 3.31. 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 3.19. **Lin** is sound and complete with respect to the class of all linear flows of time.

3.3. Temporal Logic

Def 3.32. The logic **Lin.**$\mathbb{N}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

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

Thm 3.20. **Lin.**$\mathbb{N}$ is sound and complete with respect to $\langle \mathbb{N}, < \rangle$.

3.3. Temporal Logic

Def 3.33. The logic **Lin.**$\mathbb{Z}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

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

Thm 3.21. **Lin.**$\mathbb{Z}$ is sound and complete with respect to $\langle \mathbb{Z}, < \rangle$.

3.3. Temporal Logic

Def 3.34. The logic **Lin.**$\mathbb{Q}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

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

Thm 3.22. Lin.$\mathbb{Q}$ is sound and complete with respect to $\langle \mathbb{Q}, < \rangle$.

3.3. Temporal Logic

Def 3.35. The logic **Lin.**$\mathbb{R}$ is obtained from **Lin** by adding these correspondents (within the class of linear frames):

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

Thm 3.23. **Lin.**$\mathbb{R}$ is sound and complete with respect to $\langle \mathbb{R}, < \rangle$.

3.4. Counterfactuals

Our next application is *counterfactuals*. The possible worlds framework that we have been working with affords a powerful semantic analysis of natural language conditionals such as these:

(1) If kangaroos had no tails, they would topple over.
(2) If Oswald had not killed Kennedy, then someone else would have.

The basic idea is that a counterfactual conditional is true just in case its consequent holds at all of the *closest* worlds in which its antecedent holds.

3.4. Counterfactuals

Def 3.36. The *language of counterfactuals* $\mathcal{L}_{cf}$ extends the basic sentential language with two counterfactual conditional operators:

$$p \mid \bot \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \mathrel{\Box\!\!\rightarrow} \varphi) \mid (\varphi \mathrel{\Diamond\!\!\rightarrow} \varphi)$$

Read $\varphi \mathrel{\Box\!\!\rightarrow} \psi$ as 'If it were the case that φ then it would be the case that ψ' and $\varphi \mathrel{\Diamond\!\!\rightarrow} \psi$ as 'If it were the case that φ then it might be the case that ψ'.

Note that this language is redundant since $\Box\!\!\rightarrow$ and $\Diamond\!\!\rightarrow$ are interdefinable: $\varphi \mathrel{\Box\!\!\rightarrow} \psi \equiv \neg(\varphi \mathrel{\Diamond\!\!\rightarrow} \neg\psi)$ and $\varphi \mathrel{\Diamond\!\!\rightarrow} \psi \equiv \neg(\varphi \mathrel{\Box\!\!\rightarrow} \neg\psi)$.

3.4. Counterfactuals

To evaluate counterfactuals, we will use models that include a relation of comparative similarity between worlds:

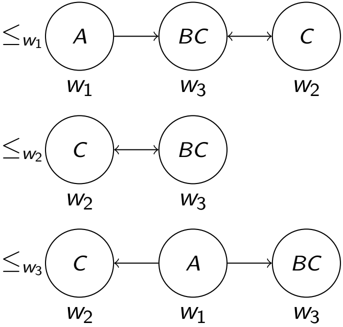
Def 3.37. A *Kripke world-ordering model* $\mathcal{M} = \langle \mathcal{W}, \{\leq_w\}_{w \in \mathcal{W}}, \mathcal{V} \rangle$ for $\mathcal{L}_{cf}$ consists of a nonempty set $\mathcal{W}$ of possible worlds, a valuation function $\mathcal{V} : At_{\mathcal{L}_{cf}} \times \mathcal{W} \mapsto \{T, F\}$, and a set of orders $\{\leq_w\}_{w \in \mathcal{W}}$ (one for each world) where $v \leq_w u$ just in case v is at least as similar to w as u.

It is required that each $\leq_w$ is transitive and reflexive.

$v <_w u$ abbreviates $v \leq_w u \wedge u \nleq_w v$.

3.4. Counterfactuals

Here is an example. If an arrow extends from v to u in the row for $\leq_w$, then $v \leq_w u$ (the arrows for transitivity and reflexivity are omitted).



3.4. Counterfactuals

This world-ordering model has some strange features.

First, $\leq_{w_2}$ is not defined over all of $\mathcal{W}$. Letting $\mathcal{W}_w$ designate the set of worlds over which $\leq_w$ is defined, $\mathcal{W}_{w_2} \neq \mathcal{W}$.

Second, w_3 is as similar to w_2 as w_2 itself.

Third, w_1 is *more* similar to w_3 than w_3 itself.

Fourth, w_2 are w_3 are incomparable according to $\leq_{w_3}$.

3.4. Counterfactuals

To block these oddities, we might impose some additional constraints.

Def 3.38. For any proposition $X \subseteq \mathcal{W}$ and world $w \in \mathcal{W}$, the set of *closest* X-worlds to w is this:

$\text{Min}_{\leq_w}(X) = \{v \in X \cap \mathcal{W}_w : \neg\exists u \in X(u <_w v)\}$

Centering. $\text{Min}_{\leq_w}(\mathcal{W}) = \{w\}$

Weak Centering. $w \in \text{Min}_{\leq_w}(\mathcal{W})$

Totality. $\forall u, v \in \mathcal{W}_w(v \leq_w u \vee u \leq_w v)$

Well-Foundedness. $\forall X \subseteq \mathcal{W}_w(X \neq \emptyset \supset \text{Min}_{\leq_w}(X) \neq \emptyset)$

In the previous model, $\leq_{w_2}$ and $\leq_{w_3}$ violate Centering, $\leq_{w_3}$ violates Weak Centering and Totality, but all of the orderings satisfy Well-Foundedness.

3.4. Counterfactuals

Assuming Well-Foundedness, the semantics is fairly straightforward:

Def 3.39. The following recursive clauses lift $\mathcal{V}$ to the complete interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{cf}} \times \mathcal{W} \mapsto \{T, F\}$ for $\mathcal{L}_{cf}$:

$$
\begin{array}{rcl}
[\![p]\!]^w_{\mathcal{M}} = T & \text{iff} & \mathcal{V}(p, w) = T \\
[\![\bot]\!]^w_{\mathcal{M}} = T & \text{iff} & 0 = 1 \\
[\![\neg\varphi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = F \\
[\![\varphi \wedge \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ and } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \vee \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & [\![\varphi]\!]^w_{\mathcal{M}} = T \text{ or } [\![\psi]\!]^w_{\mathcal{M}} = T \\
[\![\varphi \mathrel{\Box\!\!\rightarrow} \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & \text{Min}_{\leq_w}([\varphi]_{\mathcal{M}}) \subseteq [\psi]_{\mathcal{M}} \\
[\![\varphi \mathrel{\Diamond\!\!\rightarrow} \psi]\!]^w_{\mathcal{M}} = T & \text{iff} & \text{Min}_{\leq_w}([\varphi]_{\mathcal{M}}) \cap [\psi]_{\mathcal{M}} \neq \emptyset
\end{array}
$$

If we drop Well-Foundedness, then the clauses for the counterfactual conditional operators are more complex. See Lewis [1973] for details.

In the previous model, $[\![C \mathrel{\Box\!\!\rightarrow} B]\!]^{w_1}_{\mathcal{M}} = F$, $[\![C \mathrel{\Diamond\!\!\rightarrow} B]\!]^{w_2}_{\mathcal{M}} = T$, and $[\![B \mathrel{\Box\!\!\rightarrow} C]\!]^{w_3}_{\mathcal{M}} = T$.

3.4. Counterfactuals

Assuming Centering, it is easy to show that the following inference forms are valid for the counterfactual conditional:

Modus Ponens/Tollens.

$$\frac{\varphi \mathrel{\Box\!\!\rightarrow} \psi \qquad \varphi}{\psi} \qquad \frac{\varphi \mathrel{\Box\!\!\rightarrow} \psi \qquad \neg\psi}{\neg\varphi}$$

However, other inference forms that are valid for the material conditional are (correctly) invalidated by the counterfactual semantics.

3.4. Counterfactuals

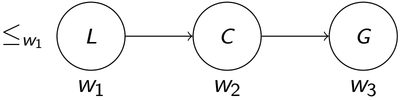
Strengthening the Antecedent.

$$\frac{\varphi \mathrel{\Box\!\!\rightarrow} \chi}{(\varphi \wedge \psi) \mathrel{\Box\!\!\rightarrow} \chi}$$

Counterexample:

(P1) If the Liberals had not won the last election, then the Conservatives would have won it.

(C) If the Liberals had not won the last election and the Greens had gotten ninety percent of the popular vote, then the Conservatives would have won it.



3.4. Counterfactuals

Transitivity.

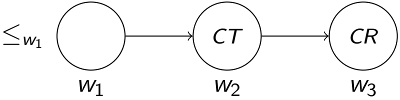
$$\frac{\varphi \mathrel{\Box\!\!\rightarrow} \psi \qquad \psi \mathrel{\Box\!\!\rightarrow} \chi}{\varphi \mathrel{\Box\!\!\rightarrow} \chi}$$

Counterexample:

(P1) If J. Edgar Hoover had been born a Russian, then he would have been a communist.

(P2) If J. Edgar Hoover had been a communist, then he would have been a traitor.

(C) If J. Edgar Hoover had been born a Russian, then he would have been a traitor.



3.4. Counterfactuals

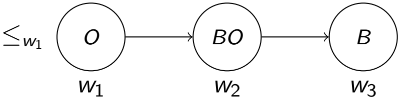
Contraposition.

$$\frac{\varphi \mathrel{\Box\!\!\rightarrow} \psi}{\neg\psi \mathrel{\Box\!\!\rightarrow} \neg\varphi}$$

Counterexample:

(P1) If Boris had gone to the party, Olga would still have gone.
(C) If Olga had not gone, then Boris would still not have gone.

(Background context: Boris wanted to attend the party but stayed away to avoid Olga who has been pursuing his heart)



3.4. Counterfactuals

So far so good. But what about the following inference pattern?

Simplification of Disjunctive Antecedents.

$$\frac{(\varphi \vee \psi) \mathrel{\Box\!\!\rightarrow} \chi}{(\varphi \mathrel{\Box\!\!\rightarrow} \chi) \wedge (\psi \mathrel{\Box\!\!\rightarrow} \chi)}$$

SDA is also invalidated by the semantics in §2.

But it seems good:

(P1) If either Oswald had not fired or Kennedy had been in a bulletproof car, then Kennedy would still be alive.

(C1) If Oswald had not fired, then Kennedy would still be alive.

(C2) If Kennedy had been in a bulletproof car, then he would still be alive.

3.5. Deontic Logic

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

Def 3.40. 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 φ' and $\mathsf{P}\varphi$ as 'It is permissible that φ'.

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

3.5. Deontic Logic

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**.

3.5. Deontic Logic

Def 3.41. *Standard Deontic Logic* (SDL) is the logic **KD**:

(PL)	All (substitutions of) tautologies are axioms
(MP)	From φ and $\varphi \supset \psi$ infer ψ
(Nec_d)	From φ infer $\mathrm{O}\varphi$
(K_d)	For any φ, ψ, $\mathrm{O}(\varphi \supset \psi) \supset (\mathrm{O}\varphi \supset \mathrm{O}\psi)$ is an axiom
(D_d)	For any φ, $\mathrm{O}\varphi \supset \mathrm{P}\varphi$ is an axiom
(Duality)	Expressions involving O and P are interchangeable according to the duality $\mathrm{O} \equiv \neg\mathrm{P}\neg$

3.5. Deontic Logic

Intuitively, $\mathsf{O}(\mathsf{O}\varphi \supset \varphi)$ 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 3.42. SDL^+ is the logic obtained by supplementing **KD** with the axiom schema $\mathsf{O}(\mathsf{O}\varphi \supset \varphi)$.

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

3.5. Deontic Logic

Conflicting Obligations. SDL rules out conflicting obligations:

1.	$(O\varphi \wedge O\neg\varphi) \supset O(\varphi \wedge \neg\varphi)$	C
2.	$\neg O(\varphi \wedge \neg\varphi)$	From D_d
3.	$\neg(O\varphi \wedge O\neg\varphi)$	PL 1,2

However, such conflicts arguably occur:

(P1) I ought to fight in the war (since I signed a contract to do so).
(P2) I ought not to fight in the war (since the war is unjust).

Response: Abandon **C** 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'.

3.5. Deontic Logic

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 $\vee$.

3.5. Deontic Logic

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.	$\mathrm{O}(H \wedge R)$	P1
2.	$\mathrm{O}R$	M, MP

Response: Deny that $\mathrm{O}(H \wedge R)$ is a good translation of P1.

Response: Abandon **M** and work with neighborhood semantics.

3.5. Deontic Logic

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.

3.5. Deontic Logic

However, these statements are inconsistent in SDL:

1.	$\mathrm{O}H$	P1
2.	$\mathrm{O}(H \supset T)$	P2
3.	$\neg H \supset \mathrm{O}\neg T$	P3
4.	$\neg H$	P4
5.	$\mathrm{O}H \supset \mathrm{O}T$	K_d, MP 2
6.	$\mathrm{O}T$	MP 5,1
7.	$\mathrm{P}T$	D_d, MP
8.	$\neg\mathrm{O}\neg T$	Duality 7
9.	$\mathrm{O}\neg T$	MP 3,4
10.	$\bot$	From 8,9

3.5. Deontic Logic

Response: Translate P2 and P3 as $H \supset OT$ and $\neg H \supset O\neg T$ respectively. But then P2 follows from P4 so the statements are not independent.

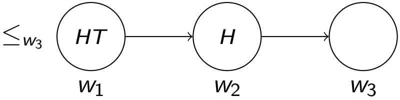
Response: Translate P2 and P3 as $O(H \supset T)$ and $O(\neg H \supset \neg T)$ respectively. But then P3 follows from P1 so the statements are not independent.

3.5. Deontic Logic

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 ψ *given that* φ' and $\mathsf{P}(\psi/\varphi)$ as 'It is permissible that ψ *given that* φ'.

The semantics for these operators is similar to the semantics for counterfactuals in using an ordering on worlds. But now $v \leq_w u$ 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:



3.5. Deontic Logic

Response: Keep the unary deontic operators but replace the material conditional with a more sophisticated intensional conditional.

3.6. Epistemic Logic

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 change when new information comes to light.

Def 3.43. 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 φ'.

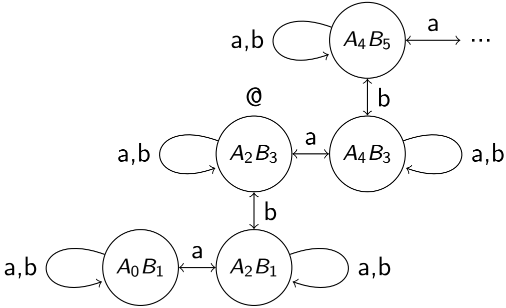
Defining the dual $\mathsf{K}_a^* = \neg\mathsf{K}_a\neg$, read $\mathsf{K}_a^*\varphi$ as 'It is compatible with what Agent *a* knows that φ'.

3.6. Epistemic Logic

Def 3.44. 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.)

For example, suppose that Agent a has a 2 written on her forehead and Agent b has a 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 $n \in \mathbb{N}$.

3.6. Epistemic Logic



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 an *a*-arrow or *b*-arrow reflects Agent *a*'s or Agent *b*'s *ignorance* respectively.

In this model $\mathcal{M}$, $[\![\mathsf{K}_a B_3 \wedge \mathsf{K}_b A_2]\!]^{@}_{\mathcal{M}} = T$, $[\![\mathsf{K}_a \neg B_5 \wedge \mathsf{K}_b \neg B_5]\!]^{@}_{\mathcal{M}} = T$, but $[\![\mathsf{K}_a \mathsf{K}_b \neg B_5]\!]^{@}_{\mathcal{M}} = F$.

3.6. Epistemic Logic

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**.

3.6. Epistemic Logic

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.	$\mathsf{K}\neg H_{501}$	MP 5,4

Repeating this reasoning, you can conclude $\mathsf{K}\neg H_{1000}$ and so $\neg H_{1000}$ by the **T** Axiom. This contradicts H_{1000}.

3.6. Epistemic Logic

We can also define some interesting notions of collective knowledge.

For instance, we might introduce the following *everyone in X knows* operator 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 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 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 $\mathsf{S}_X\varphi \equiv \bigvee_{a\in X}\mathsf{K}_a\varphi$.

3.6. Epistemic Logic

More interestingly, we might introduce the notions of *common knowledge* and *distributed knowledge*.

Def 3.45. 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:

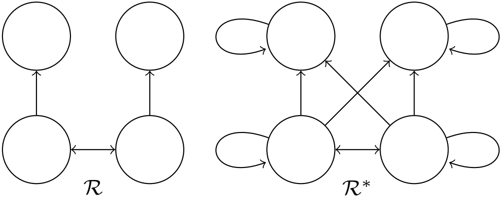
$$[\![\mathsf{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad [\![\mathsf{E}_X\varphi]\!]^w_{\mathcal{M}} = [\![\mathsf{E}_X\mathsf{E}_X\varphi]\!]^w_{\mathcal{M}} = ... = T$$

Our Kripke models afford a more elegant truth clause:

$$[\![\mathsf{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \begin{array}{l}\text{For all } v \in \mathcal{W}, \text{ if } v \text{ is reachable from } w \\ \text{in a finite number of steps along any } \mathcal{R}_a \\ \text{where } a \in X, \text{ then } [\![\varphi]\!]^v_{\mathcal{M}} = T\end{array}$$

3.6. Epistemic Logic

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.



Then the truth clause for C_X can be restated thus:

$$[\![\mathsf{C}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w(\textstyle\bigcup_{a\in X}\mathcal{R}_a)^* v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)$$

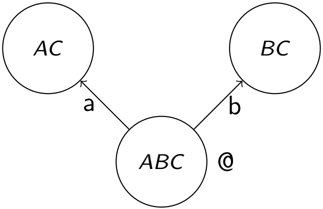
3.6. Epistemic Logic

Def 3.46. 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:

$$[\![\mathsf{D}_X\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w \bigcap_{a\in X} \mathcal{R}_a v\}([\![\varphi]\!]^v_{\mathcal{M}} = T)$$

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.

3.6. Epistemic Logic



In this model $\mathcal{M}$, $[\![\mathsf{E}_X(A \vee B)]\!]^{@}_{\mathcal{M}} = T$, $[\![\mathsf{S}_X A]\!]^{@}_{\mathcal{M}} = T$, $[\![\mathsf{C}_X C]\!]^{@}_{\mathcal{M}} = T$, and $[\![\mathsf{D}_X(A \wedge B)]\!]^{@}_{\mathcal{M}} = T$.

3.6. Epistemic Logic

Let us close with a nice application of epistemic logic: the Muddy Children Puzzle.

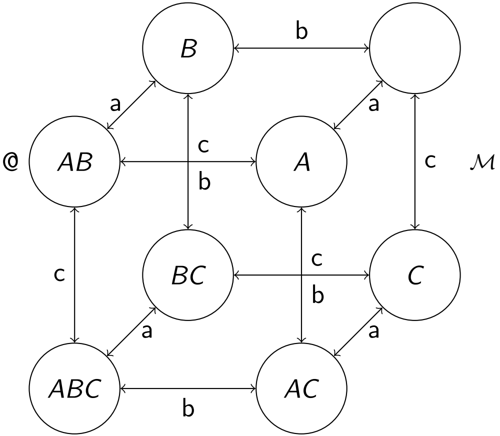
Three children a, b, and c have been playing outside in the mud.

Let A, B, and C designate that a, b, and c have mud on their forehead respectively.

In fact, A and B but $\neg C$.

3.6. Epistemic Logic

Here is the initial epistemic model:



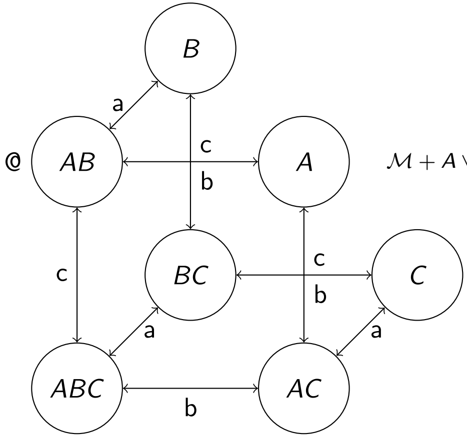
3.6. Epistemic Logic

Returning home, their mother says that at least one of the children has mud on their forehead. How does the model change?

Def 3.47. Given $\mathcal{M} = \langle \mathcal{W}, \{\mathcal{R}_a\}_{a \in \mathsf{Agt}}, \mathcal{V} \rangle$, the model *updated with* φ 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 $\mathcal{W} + \varphi$

3.6. Epistemic Logic



$\mathcal{M} + A \vee B \vee C$

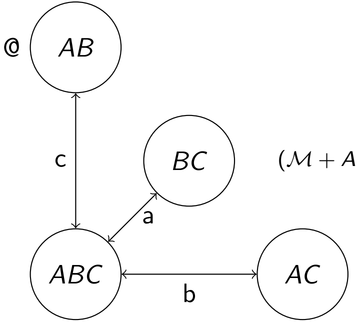
3.6. Epistemic Logic

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.

3.6. Epistemic Logic



$(\mathcal{M} + A \vee B \vee C) + \neg \mathsf{K}_a A \wedge \neg \mathsf{K}_a \neg A \ldots$

3.6. Epistemic Logic

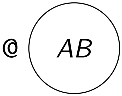
The mother again asks each of the children to step forward if they know whether they are dirty.

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.

3.6. Epistemic Logic



In fact, only @ remains open after this last update.

Next time the mother asks her question, c steps forward as well.

NASSLLI 2014

Logical Consequence: Against Truth Preservation

Justin Bledin, Johns Hopkins University

4.1. Intuitionistic Logic

Recall the standard truth preservation view of logic:

- 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 the logical form* of the sentences in the argument, and not due to the meaning of any non-logical symbols.

Should we maintain this view?

4.1. Intuitionistic Logic

Intuitionistic logicians think not.

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.

4.1. Intuitionistic Logic

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 φ into a proof of $\bot$
- A proof of $(\varphi \wedge \psi)$ is a proof of φ and a proof of ψ
- A proof of $(\varphi \vee \psi)$ is a proof of φ or a proof of ψ
- A proof of $(\varphi \sqsupset \psi)$ is a method of converting any proof of φ into a proof of ψ

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.

4.1. Intuitionistic Logic

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 $B \supset (A \vee B)$
- $(A \supset C) \supset ((B \supset C) \supset ((A \vee B) \supset C))$
- $\bot \supset A$
- $A \vee \neg A$

$\vdash_{\mathbf{CL}} \varphi$ designates that φ is provable in this classical system **CL**.

4.1. Intuitionistic Logic

We get intuitionistic sentential logic by eliminating the last axiom $A \vee \neg A$.

$\vdash_{\mathbf{IL}} \varphi$ designates that φ is provable in the resulting proof system **IL**.

Is there a semantics corresponding to **IL**?

4.1. Intuitionistic Logic

Def 4.1. An *intuitionistic model* for $\mathcal{L}_{ISENT}$ is a Kripke model where $\mathcal{R}$ is a *partial order* (reflexive, transitive, and antisymmetric) and the following *heredity* property holds:

For all $p \in At_{\mathcal{L}}$ and $w, v \in \mathcal{W}$,
if $\mathcal{V}(p, w) = T$ and $w\mathcal{R}v$ then $\mathcal{V}(p, v) = T$.

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.

4.1. Intuitionistic Logic

Def 4.2. A recursive specification lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{ISENT}} \times \mathcal{W} \mapsto \{T, F\}$ for $\mathcal{L}_{ISENT}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{ISENT}}$ and world $w \in \mathcal{W}$ to a value in $\{T, F\}$:

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

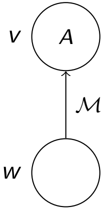
Since $\neg\varphi \equiv \varphi \sqsupset \bot$,

$$
[\![\neg\varphi]\!]^w_{\mathcal{M}} = T \quad \text{iff} \quad \forall v \in \{v : w\mathcal{R}v\}([\![\varphi]\!]^v_{\mathcal{M}} = F)
$$

When $[\![\varphi]\!]^w_{\mathcal{M}} = T$, we needn't say that φ is *true* at w in $\mathcal{M}$.

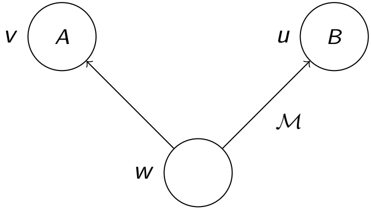
We can say that w *forces* φ in $\mathcal{M}$.

4.1. Intuitionistic Logic



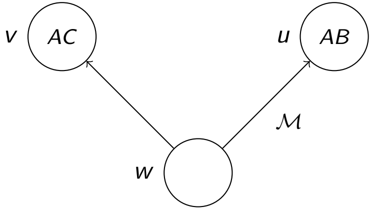
$$[\![A \vee \neg A]\!]^w_{\mathcal{M}} = [\![\neg\neg A \sqsupset A]\!]^w_{\mathcal{M}} = F$$

4.1. Intuitionistic Logic



$$[\![\neg(A \land B) \supset (\neg A \lor \neg B)]\!]^{w}_{\mathcal{M}} = F$$

4.1. Intuitionistic Logic



$$[\![(A \supset (B \vee C)) \supset ((A \supset B) \vee (A \supset C))]\!]^{w}_{\mathcal{M}} = F$$

4.1. Intuitionistic Logic

Back to our original question:

Thm 4.1 (Soundness and Completeness Theorem for IL).
$\vdash_{\mathbf{IL}} \varphi$ if and only if φ is valid over the class of *finite* intuitionistic Kripke models.

4.2. Field's Truth-Theoretic Argument

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.

4.2. Field's Truth-Theoretic Argument

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.

4.2. Field's Truth-Theoretic Argument

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.

4.2. Field's Truth-Theoretic Argument

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 $\varphi \in S_{\mathcal{L}}$.

Since T proves $\neg Tr(\ulcorner\lambda\urcorner) \equiv \lambda$ for the Liar sentence λ, 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.

4.2. Field's Truth-Theoretic Argument

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.

4.2. Field's Truth-Theoretic Argument

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	λ	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

4.2. Field's Truth-Theoretic Argument

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.

4.2. Field's Truth-Theoretic Argument

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 φ and $Tr(\ulcorner\varphi\urcorner)$ must be *intersubstitutable* in transparent (non-quotational, etc.) contexts.

But this intersubstitutivity combined with classical logic leads to trouble.

4.2. Field's Truth-Theoretic Argument

1	$\lambda \equiv \neg Tr(\ulcorner \lambda \urcorner)$	Liar Property
2	$\lambda \vee \neg \lambda$	Axiom
3	$\quad \lambda$	Assumption
4	$\quad \neg Tr(\ulcorner \lambda \urcorner)$	From 1,3
5	$\quad \neg \lambda$	From 4 (Sub *Tr*)
6	$\quad \bot$	From 3,5
7	$\quad \neg \lambda$	Assumption
8	$\quad Tr(\ulcorner \lambda \urcorner)$	From 1,7
9	$\quad \lambda$	From 8 (Sub *Tr*)
10	$\quad \bot$	From 7,9
11	$\bot$	From 2,3-6,7-10

4.2. Field's Truth-Theoretic Argument

Thus, theories that are not *externally dialethethic*—that reject the existence of 'true contradictions'—and hold on to the intersubstitutivity of φ and $Tr(\langle\varphi\rangle)$ and the classical elimination rules for $\vee$ and $\equiv$ must restrict the law of excluded middle.

Now, let γ be a Curry sentence, that is, a sentence such that the theories under consideration all prove $\gamma \leftrightarrow (Tr(\ulcorner\gamma\urcorner) \rightarrow \bot)$, and consider the argument from γ and $\gamma \rightarrow \bot$ to $\bot$.

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.

4.2. Field's Truth-Theoretic Argument

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	γ	From 4,6
8	$\bot$	From 6,7

4.2. Field's Truth-Theoretic Argument

But if logical validity shouldn't be defined in terms of truth preservation, then what is logic about?

Field's [2009a] answer is this:

"[I]f 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*.

4.2. Field's Truth-Theoretic Argument

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.

4.3. Informational View of Logic

There are reasons aside from the semantic paradoxes to think that truth preservation and good deductive argument come apart. Certain arguments involving informational modal operators and the indicative conditional suggest that logically valid arguments do not preserve truth.

We do well to make this argument in both categorical and hypothetical deliberative contexts (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.

But the inference from P1 and P2 to C arguably fails to preserve truth.

4.3. Informational View of Logic

Let us now add the indicative conditional $\Rightarrow$ to $\mathcal{L}_{SENT_\Rightarrow}$:

Def 4.3. The language $\mathcal{L}_{SENT_\Rightarrow}$ has the following syntax:

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

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

4.3. Informational View of Logic

Def 4.4. A *model* $\mathcal{M} = \langle \mathcal{W}, \mathcal{V} \rangle$ for $\mathcal{L}_{SENT_{\Rightarrow}}$ consists of a nonempty set $\mathcal{W}$ of possible worlds and a valuation $\mathcal{V} : At_{\mathcal{L}_{SENT_{\Rightarrow}}} \times \mathcal{W} \mapsto \{T, F\}$ mapping each sentence letter $p \in At_{\mathcal{L}_{SENT_{\Rightarrow}}}$ and world $w \in \mathcal{W}$ to a truth value.

Sentences in $S_{\mathcal{L}_{SENT_{\Rightarrow}}}$ 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).

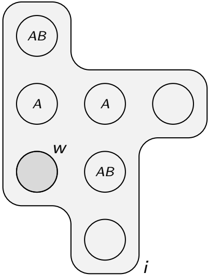
4.3. Informational View of Logic

Def 4.5. A recursive specification of *truth-in-a-model* lifts $\mathcal{V}$ to the full interpretation function $[\![\ ]\!]_{\mathcal{M}} : S_{\mathcal{L}_{SENT_{\Rightarrow}}} \times \mathcal{W} \times 2^{\mathcal{W}} \mapsto \{T, F\}$ for $\mathcal{L}_{SENT_{\Rightarrow}}$ mapping each sentence $\varphi \in S_{\mathcal{L}_{SENT_{\Rightarrow}}}$, world $w \in \mathcal{W}$, and information state $i \in 2^{\mathcal{W}}$ to a truth value:

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

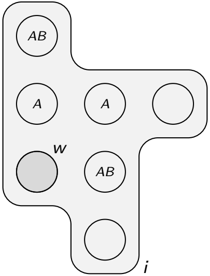
where $i + \varphi$ appearing in the clause for the indicative is the largest subset $i' \subseteq i$ such that $\forall w \in i'([\![\varphi]\!]^{w,i'}_{\mathcal{M}} = T)$.

4.3. Informational View of Logic



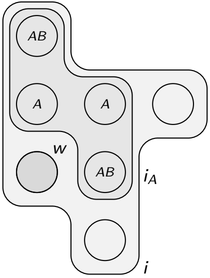
$$[\![\Box(A \vee \neg B)]\!]^{w,i}_{\mathcal{M}} = T$$

4.3. Informational View of Logic



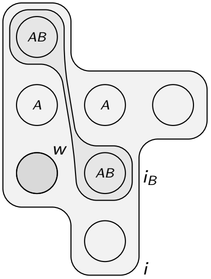
$$[\![\Diamond(A \wedge B)]\!]^{w,i}_{\mathcal{M}} = T$$

4.3. Informational View of Logic



$$[\![A \Rightarrow B]\!]_{\mathcal{M}}^{w,i} = F$$

4.3. Informational View of Logic



$$[\![B \Rightarrow A]\!]^{w,i}_{\mathcal{M}} = T$$

4.3. Informational View of Logic

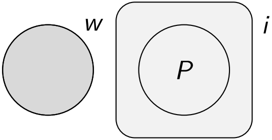
What about the formal consequence relation?

Running with the truth preservation view of logic, here is a natural first suggestion:

Def 4.6. $\{\varphi_1, ..., \varphi_n\} \models_0 \psi$ just in case there is no model $\mathcal{M}$ such that for some $w \in W$ and $i \in 2^W$, $[\![\varphi_1]\!]^{w,i}_{\mathcal{M}} = ... = [\![\varphi_n]\!]^{w,i}_{\mathcal{M}} = T$ and $[\![\psi]\!]^{w,i}_{\mathcal{M}} = F$.

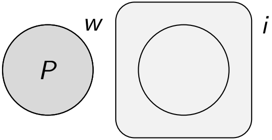
However, Def 4.6 is too demanding. $\models_0$ invalidates some good arguments.

4.3. Informational View of Logic



(P1) Mrs. Peacock must have done it.
(C) Mrs. Peacock did it.

4.3. Informational View of Logic



(P1) Mrs. Peacock did it.
(C) Mrs. Peacock might have done it.

4.3. Informational View of Logic

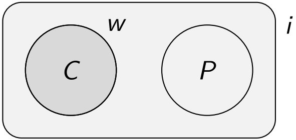
Kolodny and MacFarlane [2010] offer this fix:

Def 4.7. $\{\varphi_1, ..., \varphi_n\} \models_{Tr} \psi$ just in case there is no model $\mathcal{M}$ such that for some $i \in 2^W$ and $w \in i$, $[\![\varphi_1]\!]^{w,i}_{\mathcal{M}} = ... = [\![\varphi_n]\!]^{w,i}_{\mathcal{M}} = T$ and $[\![\psi]\!]^{w,i}_{\mathcal{M}} = F$.

$\{\Box A\} \models_{Tr} A$ and $\{A\} \models_{Tr} \Diamond A$, so $\models_{Tr}$ supports our judgments that these implications hold.

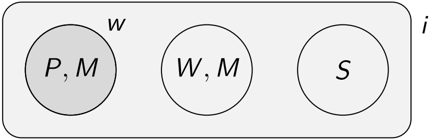
However, $\models_{Tr}$ also invalidates some intuitively good deductive arguments.

4.3. Informational View of Logic



(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.

4.3. Informational View of Logic



(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.

4.3. Informational View of Logic

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.

4.3. Informational View of Logic

We might follow Field and take logic to be essentially normative.

However, Field's line seems drastic. Not only must we give up the truth preservation view of logic, we must give up the more basic idea that logic is a descriptive science. Is there a better option?

4.3. Informational View of Logic

The formal semantics suggests another formal consequence relation (cf. Veltman [1996], Yalcin [2007], Kolodny and MacFarlane [2010]):

Def 4.8. $\{\varphi_1, ..., \varphi_n\} \models_I \psi$ just in case there is no $\mathcal{M}$ such that for some $i \in 2^W$, $\forall w \in i([\![\varphi_1]\!]^{w,i}_{\mathcal{M}} = ... = [\![\varphi_n]\!]^{w,i}_{\mathcal{M}} = T)$ and $\neg\forall w \in i([\![\psi]\!]^{w,i}_{\mathcal{M}} = T)$.

Unlike the relations in Def 4.6 and Def 4.7, 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 $Sent_{\mathcal{L}_{\Rightarrow}}$ just in case all information states in all models have a particular kind of *structure*.

4.3. Informational View of Logic

Each sentence $\varphi \in Sent_{\mathcal{L}_{\Rightarrow}}$ corresponds to a potential feature of information states:

Def 4.9. $i \rhd \varphi$ just in case $\forall w \in i(\llbracket \varphi \rrbracket^{w,i} = T)$.

Let us say that when $i \rhd \varphi$, φ is *incorporated* in i.

For instance:

$i \rhd A$ just in case $\forall w \in i(\llbracket A \rrbracket^{w,i} = T)$; that is, $i \rhd A$ just in case every world in i is an A-world.

$i \rhd \Diamond B$ just in case $\forall w \in i(\llbracket \Diamond B \rrbracket^{w,i} = T)$; that is, $i \rhd \Diamond B$ just in case some world in i is a B-world.

4.3. Informational View of Logic

Def 4 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.

4.3. Informational View of Logic

$\models_I$ 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.