The Logic of Provability

Notes by R.J. Buehler
Based on *The Logic of Provability* by George Boolos

September 16, 2014

Contents

Preface v

1 GL and Modal Logic **1**
- 1.1 Introduction . . . 1
- 1.2 Natural Deduction . . . 2
  - 1.2.1 $\Box$ . . . 2
  - 1.2.2 $\Diamond$ . . . 2
- 1.3 Definitions and Terms . . . 3
- 1.4 Normality . . . 4
- 1.5 Refining The System . . . 6

2 Peano Arithmetic **9**
- 2.1 Introduction . . . 9
- 2.2 Basic Model Theory . . . 10
- 2.3 The Theorems of PA . . . 10
  - 2.3.1 P-Terms . . . 10
  - 2.3.2 Σ, Π, and Δ Formulas . . . 11
  - 2.3.3 Gödel Numbering . . . 12
  - 2.3.4 Bew(x) . . . 12
- 2.4 On the Choice of PA . . . 13

3 The Box as Bew(x) **15**
- 3.1 Realizations and Translations . . . 15
- 3.2 The Generalized Diagonal Lemma . . . 16
- 3.3 Löb's Theorem . . . 17
- 3.4 K4LR . . . 19
- 3.5 The Box as Provability . . . 20
- 3.6 GLS . . . 21

4 Model Theory for GL **23**

5 Completeness & Decidability of GL **25**

6 Canonical Models **27**

7 On GL **29**
- 7.1 Normal Form for Letterless Sentences . . . 29
  - 7.1.1 Constant Sentences . . . 30
- 7.2 Rank and Trace . . . 31
- 7.3 Reflection and Iterated Consistency . . . 33
- 7.4 Nice Normal Forms are Unusual . . . 33

7.5 Incompactness 34

8 The Fixed Point Theorem **35**

Preface

What follows are my personal notes on George Boolos' *The Logic of Provability.* Most of the ideas presented in this document are not my own, but rather Boolos' and should be treated accordingly. This text is not meant for reproduction or as a replacement for Boolos' book, but rather as a convenient reference and summary, suitable for use as lecture notes or a review and little more. For a complete presentation of the thoughts and arguments presented, please see the full text of *The Logic of Provability.*

Chapter 1

GL and Modal Logic

1.1 Introduction

Throughout this text, we study the system GL, named for the logicians Gödel and Löb. GL is a normal modal logic like the systems K, T, S4, S5, and others, meaning that it is at least as strong as the logic K.

Definition: K
The logic generated by the following axioms,

- All tautologies of propositional logic (PL), allowing modal substitutions
- The K axiom: $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

And allowing the following rules of inference,

- Necessitation: If $\vdash_K \phi$, then $\vdash_K \Box\phi$
- Modus Ponens: If $\vdash_K \phi \to \psi$ and $\vdash_K \phi$, then $\vdash_K \psi$

along with substitution.

The syntax of GL is precisely the same as those systems outlined above, and so it is omitted here. The theorems of GL differ greatly from that of the other modal logics listed because of the addition of a new axiom:

Definition: GL
The logic generated by the following axioms,

- All tautologies of propositional logic (PL), allowing modal substitutions
- The K axiom: $\Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$
- Löb's axiom: $\Box(\Box\phi \to \phi) \to \Box\phi$

And allowing the following rules of inference,

- Necessitation: If $\vdash_{GL} \phi$, then $\vdash_{GL} \Box\phi$
- Modus Ponens: If $\vdash_{GL} \phi \to \psi$ and $\vdash_{GL} \phi$, then $\vdash_{GL} \psi$

along with substitution.

It is worth noting that, if GL is to be consistent, the schema $\Box\phi \rightarrow \phi$–known as the T axiom– must not hold generally.

1.2 Natural Deduction

Throughout these notes, a hybrid natural deduction and axiomatization style is used; in particular, we take all of the natural deduction rules outlined in Barwise and Etchemendy's *Language, Proof, and Logic* while also introducing a new subproof form and several new rules to account for the behavior of $\Box$ and $\Diamond$ under K, finally adding the GL-axiom as invokable at any line of a proof to obtain a proof system for GL.

1.2.1 $\Box$

For the moment, we omit recapping the standard introduction and elimination rules of propositional logic and move immediately into introducing the new modal rules. Taking $\Box$ first, the structure shown below is a $\Box$-subproof,

$$\begin{array}{r|l} 1 & \Box \\ \hline 2 & \vdots \end{array}$$

A $\Box$-subproof may be invoked at any point and represents the only means of introducing a $\Box$ operator; the $\Box$-subproof differs, rather crucially, from the standard subproof in that only sentences with '$\Box$' as their outermost operation can be re-iterated into the subproof and–when this is done–this initial $\Box$ is stripped from the sentence; the rule encapsulating these constraints is known as $\Box$-reit and is shown below:

$$\begin{array}{r|l|l} 1 & \Box\phi & \\ 2 & \quad\Box & \\ 3 & \quad\phi & \Box\text{-reit } 1 \end{array}$$

A $\Box$-subproof can be ended at any time, presuming no subproofs within it are open. The rules that does so is dubbed $\Box$-intro and prepends a '$\Box$' to the last line of the subproof as below:

$$\begin{array}{r|l|l} 1 & \Box\phi & \\ 2 & \quad\Box & \\ 3 & \quad\phi & \Box\text{-reit } 1 \\ 4 & \Box\phi & \Box\text{-intro } 2\text{-}3 \end{array}$$

1.2.2 $\Diamond$

Just as with $\Box$, $\Diamond$ also gives rise to a new subproof:

$$\begin{array}{r|l} 1 & \Diamond \\ \hline 2 & \vdots \end{array}$$

Unlike $\Box$, however, the $\Diamond$ subproof is quite restrictive. In particular, the reiteration rule for this new subproof is only allowed to be invoked once and is simultaneous with the opening of the subproof; this means the subproof cannot be opened without immediately bringing a formula inside it! The $\Diamond$-reit rule, then, always starts a $\Diamond$-subproof, moves only a single, initial formula with '$\Diamond$' as its outermost formula inside, and–as with $\Box$-reit–strips the preceding $\Diamond$ in the process.

$$\begin{array}{r|lll}
1 & \Diamond\phi & & \\
2 & \quad \mid \underline{\;\Diamond\;} & & \\
3 & \quad \mid \;\phi & & \Diamond\text{-reit } 1
\end{array}$$

As earlier, a $\Diamond$-subproof may be ended at any time, so long as no open subproofs exist inside whereupon the last line of the subproof is perpended with a '$\Diamond$', a rule known as $\Diamond$-introduction:

$$\begin{array}{r|lll}
1 & \Diamond\phi & & \\
2 & \quad \mid \underline{\;\Diamond\;} & & \\
3 & \quad \mid \;\phi & & \Diamond\text{-reit } 1 \\
4 & \Diamond\phi & & \Diamond\text{-intro } 2\text{-}3
\end{array}$$

Additionally, however, we also allow $\Box$-reiteration to function as normal; that is, formulas with $\Box$ as their outermost operation can be re-iterated into a $\Diamond$-subproof, stripping the outermost $\Box$ as usual of course. The reverse process, however, is not admittable–$\Diamond$-reit is used only when opening a $\Diamond$-subproof.

$$\begin{array}{r|lll}
1 & \Diamond\phi & & \\
2 & \Box\psi & & \\
3 & \Box\xi & & \\
4 & \quad \mid \underline{\;\Diamond\;} & & \\
5 & \quad \mid \;\phi & & \Diamond\text{-reit } 1 \\
6 & \quad \mid \;\psi & & \Box\text{-reit } 2 \\
7 & \quad \mid \;\xi & & \Box\text{-reit } 3 \\
8 & \Diamond\psi & & \Diamond\text{-intro } 2\text{-}7
\end{array}$$

1.3 Definitions and Terms

Definition: Modal Sentence
A well-formed formula of GL with no unbound variables; also called a *sentence*.

Example 1. *Modal sentences.*

$$p$$

$$p \wedge q$$

$$p \to \bot$$

$$\Box(p \to q) \to \bot$$

Definition: Letterless Sentence
A well-formed formula of GL using only '$($', '$)$', '$\bot$', '$\to$', and '$\Box$'.

Example 2. *Letterless sentences.*

$$\bot$$

$$\Box\bot$$

$$\bot \land \neg\bot$$

$$\Box(\bot \to \bot) \to \bot$$

Definition: Substitution Instance
Let F be a modal sentence. We denote the substitution of some formula ϕ for a sentence letter p as $F_p(\phi)$ and call $F_p(\phi)$ a *substitution instance*.

Less formally, a substitution into a formula is defined precisely as it is elsewhere; we simply formalize the notion here.

1.4 Normality

We now show that GL proves, rather expectedly, a number of theorems associated with normal modal logics; that is, logics which have K as a sublogic.

▶ Theorem 1.
If $GL \vdash \phi \to \psi$, then $GL \vdash \Box\phi \to \Box\psi$

Proof Sketch. *Note that the proof of $\phi \to \psi$ can simply be done within a $\Box$ subproof, and so $\Box(\phi \to \psi)$ is obtained; the result follows by the K axiom.*

▶ Theorem 2.
If $GL \vdash \phi \leftrightarrow \psi$, then $GL \vdash \Box\phi \leftrightarrow \Box\psi$

Proof Sketch. *Utilize the theorem above.*

▶ Theorem 3.
$GL \vdash \Box(\phi \land \psi) \leftrightarrow \Box\phi \land \Box\psi$

Proof Sketch. *The left to right direction is shown below; the converse follows similarly.*

Proof.

1	$\Box(\phi \wedge \psi)$	
2	$\quad \Box$	
3	$\quad \phi \wedge \psi$	$\Box$-reit 1
4	$\quad \phi$	$\wedge$-elim 3
5	$\quad \psi$	$\wedge$-elim 3
6	$\Box\phi$	$\Box$-intro 2-4
7	$\Box\psi$	$\Box$-intro 2-5
8	$\Box\phi \wedge \Box\psi$	$\wedge$-intro 6,7
9	$\Box(\phi \wedge \psi) \rightarrow \Box\phi \wedge \Box\psi$	$\rightarrow$-intro 1-8

□

► **Theorem 4.**
GL $\vdash \Box(\phi_1 \wedge \cdots \wedge \phi_n) \leftrightarrow \Box\phi \wedge \cdots \wedge \Box\phi_n$

Proof Sketch. *Argue by induction using the theorem above.*

► **Theorem 5.**
If GL $\vdash (\phi_1 \wedge \cdots \wedge \phi_n) \rightarrow \psi$, *then GL* $\vdash (\Box\phi_1 \wedge \cdots \wedge \Box\phi_n) \rightarrow \Box\psi$

Proof Sketch. *Utilize the earlier theorems.*

► **Theorem 6.**
If GL $\vdash \phi \rightarrow \psi$, *then GL* $\vdash \Diamond\phi \rightarrow \Diamond\psi$

Proof Sketch. *Assume GL* $\vdash \phi \rightarrow \psi$. *Then, given* ϕ, *there is a proof for* ψ. *Simply assume* $\Diamond\phi$, *start a* $\Diamond$*-subproof, pull in* $\Diamond\phi$, *and then past the proof for* ψ. *Upon exiting the subproof,* $\Diamond\psi$ *is established, and so the desired result follows from* $\rightarrow$*-introduction.*

► **Theorem 7.**
If GL $\vdash \phi \leftrightarrow \psi$, *then GL* $\vdash \Diamond\phi \leftrightarrow \Diamond\psi$

Proof Sketch. *Use the theorem above.*

► **Theorem 8.**
GL $\vdash \Diamond\phi \wedge \Box\psi \rightarrow \Diamond(\phi \wedge \psi)$

Proof Sketch. *Simply leverage* $\Diamond$*-subproof and* $\Box$*-reit.*

Now that the basics of GL have been established, we prove the first of our non-trivial theorems:

► **Theorem 9** (K4 $\subseteq$ GL)**.**
K4 is a sublogic of GL

Proof.

1	$\mid \Box\phi$	
2	$\mid \quad \mid \Box$	
3	$\mid \quad \mid \quad \mid \Box(\Box\phi \wedge \phi)$	
4	$\mid \quad \mid \quad \mid \phi$	$\Box$-reit 2
5	$\mid \quad \mid \quad \mid \quad \mid \Box$	
6	$\mid \quad \mid \quad \mid \quad \mid \Box\phi \wedge \phi$	$\Box$-reit 3
7	$\mid \quad \mid \quad \mid \quad \mid \phi$	$\wedge$-elim 7
8	$\mid \quad \mid \quad \mid \Box\phi$	$\Box$-intro 6-8
9	$\mid \quad \mid \quad \mid \Box\phi \wedge \phi$	$\wedge$-intro 9
10	$\mid \quad \mid \Box(\Box\phi \wedge \phi) \rightarrow (\Box\phi \wedge \phi)$	$\rightarrow$-intro 3-9
11	$\mid \Box[\Box(\Box\phi \wedge \phi) \rightarrow (\Box\phi \wedge \phi)]$	$\Box$-intro 2-10
12	$\mid \Box[\Box(\Box\phi \wedge \phi) \rightarrow (\Box\phi \wedge \phi)] \rightarrow \Box(\Box\phi \wedge \phi)$	GL
13	$\mid \Box(\Box\phi \wedge \phi)$	MP 11,12
14	$\mid \quad \mid \Box$	
15	$\mid \quad \mid \Box\phi \wedge \phi$	$\Box$-reit 13
16	$\mid \quad \mid \Box\phi$	$\wedge$-elim 15
17	$\mid \Box\Box\phi$	$\Box$-intro 16
18	$\Box\phi \rightarrow \Box\Box\phi$	$\rightarrow$-intro 1-17

$\Box$

1.5 Refining The System

As the proofs in the previous sections have shown, for those comfortable with the given rules, our current proof system is, while sufficient, a bit too clunky to be in good taste. We streamline it, therefore, by simplifying the notation for $\Box$ and $\Diamond$ subproofs and allowing the invocation of both the K and 4 axioms. Rather than the subproof-esque notation used earlier for $\Box$ and $\Diamond$ subproofs, we now use a double line to signify a modal subproof and leave it to the reader to determine whether a $\Box$ or $\Diamond$ subproof is being invoked (hint: a $\Diamond$ subproof must start with $\Diamond$-reit, while a $\Box$ subproof cannot). The proof that $K4 \subseteq GL$, then, is written as:

Proof.

1	$\Box\phi \rightarrow \Box\Box\phi$	4

$\Box$

If we wished to show it without the 4 axiom, however, we might do it as follows:

Proof.

1	$\Box\phi$	Assump.
2	$\quad\quad \Box(\Box\phi \wedge \phi)$	Assump.
3	$\quad\quad \phi$	$\Box$-reit 1
4	$\quad\quad \Box\Box\phi \wedge \Box\phi$	K 2
5	$\quad\quad \Box\phi$	$\wedge$-elim 4
6	$\quad\quad \Box\phi \wedge \phi$	$\wedge$-intro 3,5
7	$\quad \Box(\Box\phi \wedge \phi) \to (\Box\phi \wedge \phi)$	$\to$-intro 3-9
8	$\Box[\Box(\Box\phi \wedge \phi) \to (\Box\phi \wedge \phi)]$	$\Box$-intro 2-10
9	$\Box[\Box(\Box\phi \wedge \phi) \to (\Box\phi \wedge \phi)] \to \Box(\Box\phi \wedge \phi)$	GL
10	$\Box(\Box\phi \wedge \phi)$	MP 11,12
11	$\Box\Box\phi \wedge \Box\phi$	K 10
12	$\Box\Box\phi$	$\wedge$-elim 11
13	$\Box\phi \to \Box\Box\phi$	$\to$-intro 1-12

$\Box$

Which conveys, one hopes, the same information as the original, but at a cost of only $\frac{2}{3}$ the space.

Chapter 2

Peano Arithmetic

2.1 Introduction

Our subject matter for the next chapter is the mathematical theory known as 'Peano Arithmetic' or 'PA'. Peano arithmetic is the theory of arithmetic with the following axiomatization:

Definition: Peano Arithmetic
Classical first-order arithmetic with induction; also called arithmetic or PA. More formally, we take the signature of PA to have '0' as a constant and '+', '$\cdot$', and '$<$' as binary function symbols; PA is then the theory axiomatized by the following:

- $\forall x(sx \neq 0)$
- $\forall x, y(sx = sy \rightarrow x = y)$
- For every first-order formula $\phi(x, \bar{z})$, $\forall \bar{z}(\phi(0, \bar{z}) \wedge \forall x(\phi(x, \bar{z}) \rightarrow \phi(sx, \bar{z})) \rightarrow \forall x(\phi(x, \bar{z})))$
- $\forall x(x + 0 = x)$; $\forall x, y(s(x + y) = x + (sy))$
- $\forall x(x \cdot 0 = 0)$; $\forall x, y(x \cdot (sy) = x \cdot y + x)$
- $\forall x \neg(x < 0)$; $\forall x, y(x < (sy) \leftrightarrow x < y \vee x = y)$

Throughout this text we will make use of notations used by Kurt Gödel in *On Formally Undecidable Propositions of Principia Mathematica and Related Systems*. To begin, we adopt the standard symbol '$\vdash$' for provability using universal generalization and modus ponens; within this chapter, '$\vdash$' will always refer to provability within PA. If at any point the system in use is ambiguous, it will be subscripted below the turnstile.

Definition: Numeral
For a given natural number n, the numeral $\mathbf{n}$ for n is the term of P given by n occurrences of the successor relation and then 0, i.e. '$ss \ldots s0$'.

We adopt so-called Gödel numbering in order to facilitate self-reference in PA; in particular, for a sentence of PA, S, the numeral of the Gödel number for S, that is the numeral corresponding to the encoding of 'S', is denoted $\ulcorner S \urcorner$. In addition, we will make heavy use of the predicate $\mathrm{Bew}(x)$; short for the German *beweisbar* or 'provable'. The intended interpretation of $\mathrm{Bew}(x)$ is, of course, the natural language predicate 'is provable in PA'; much of the work of this chapter is devoted to showing that interpreting $\mathrm{Bew}(x)$ as intended is reasonable. It should never be forgotten, however, that $\mathrm{Bew}(x)$ itself is simply another formula of PA, a statement in the object language. It does not hold the same status as $\vdash$ or any other metalanguage symbol/concept–it is irrevocably trapped in the world of PA and simply happens (albeit by design) to be true of sentences which–under our interpretation of PA–are provable.

2.2 Basic Model Theory

Some basic model theory terms with their application to the current work are listed below; this is by no means a complete introduction, and so the reader is encouraged to pursue a more systematic exposition on model theory should what little is provided here not be sufficient.

Definition: Term
The finite sequences of symbols defined by the following:

- Every variable is a term
- '0' is a term
- For terms t_1 and t_2, 'st_1', '$t_1 = t_2$', '$(t_1 + t_2)$', '$(t_1 \cdot t_2)$' are all terms

Definition: Atomic Formula
A finite sequence of symbols α is an atomic formula if and only if α is either '$\bot$' or for some terms t_1 and t_2, α is '$t_1 = t_2$'.

To begin, the signature for PA given in the introduction–$(0; s, +, \cdot, \leq)$–is expanded into a language by taking all the atomic formulas of the signature given and adding the logical operators $\wedge, \vee, \neg, \rightarrow, \bot, \forall, \exists$ with the standard logical axioms. A *formula* of PA is any string of symbols achievable by combining atomic formulas of the language of PA in the standard logical ways, allowing free variables. A formula is *closed* or a *sentence* just in case it has no free variables.

2.3 The Theorems of PA

It is not at all clear at this point that much of interest can be formulated within PA; nevertheless, PA proves a surprisingly strong system. Strong enough, in fact, to develop a theory of its own syntax. This derivation is a well-established fact at this point, and so a full presentation is omitted here. Boolos' text gives a mostly complete presentation, and the interested reader should refer there or to my notes on Gödel's incompleteness theorems.

2.3.1 P-Terms

As both a brief aside into the strength of PA and an opportunity to introduce an important piece of terminology, note that PA, bearing only the constant 0 and functions $s, +, \cdot$ fails to produce any term for many common place functions, e.g. 2^x. Nevertheless, PA can still talk about such functions–albeit in a slightly roundabout way.

Definition: P-Term
A formula $F(\bar{x}, y)$ of PA is a p-term (short for pseudo-term) with respect to y if and only if the formula $\exists! y\ F(\bar{x}, y)$ is provable in PA.

Note that it's not unreasonable to think of $F(\bar{x}, y)$ as an n-ary ($|\bar{x}| = n$) function which takes $\bar{x}$ as input and returns y. This isn't, of course, what's actually happening, but–in PA–it's provable that no trouble arises from pretending it is. Indeed, it's common to make this interpretation concrete and write $f(\bar{x})$ to represent the unique y such that $F(\bar{x}, y)$ holds in PA. Thus, while PA lacks an actual term for functions like 2^x, it can construct p-terms for them that agree on all values and behave as though they were a term.

2.3.2 Σ, Π, and Δ Formulas

In talking about PA we will make heavy use of three particular classes of formulas. The most basic has already been introduced, but we review it here.

> **Definition: Σ Formulas**
> A formula is a Σ *formula* if and only if it is provably equivalent in PA to a prenex formula representing the existential generalization of a formula built using only the truth-functional connectives, atomic formulas, and bounded quantification.

A key feature of Σ sentences (Σ formulas without free variables) is the following:

> ▶ **Theorem 10.**
> *If ϕ is a true Σ sentence, then PA $\vdash \phi$*

Proof Sketch. *First, prove that if t is a closed term and t denotes i, then $\vdash t = \mathbf{i}$ using induction on the construction of t. Next, prove that if t, t' are closed terms and $t = t'$ is true, then $\vdash t = t'$ by leveraging the previous result. Finally, run an induction on formulas to obtain the desired result.*

Crucially, truths to the effect that a certain sentence is provable in some particular formal system or that a certain computational device eventually halts can all be expressed as Σ sentences in PA–and thus, by above, are provable in PA. This isn't, of course, to say that all sentences are provable in PA. In particular, the class of negations of Σ sentences–the so called Π sentences–contains some truths which PA does not prove.

What kinds of constraints are there on what PA can represent? This can't be EVERY logic can it?

> **Definition: Π-Formulas**
> A formula is a Π *formula* if and only if it is provably equivalent in PA to a prenex formula representing the universalization of a formula built using only the truth-functional connectives, atomic formulas, and bounded quantification.

Finally, the last class of formulas we consider is the Δ formulas:

> **Definition: Δ Formulas**
> A formula ϕ is a Δ *formula* if and only if both ϕ and $\neg\phi$ are Σ formulas

The class of Δ formulas is an critical one with an important and fundamental characteristic; first, however, we consider some properties of the class. Every atomic formula of PA, $t = t'$, is a Δ formula. Similarly, $t < t'$ is Δ for any terms t, t'. Note further that the class of Δ formulas is closed under negation and conjunction: the negation of a Δ formula is always a Δ formula and the conjunction of two Δ formulas is itself a Δ formula. With these in hand, it follows easily that the Δ formulas are closed under all of the boolean connectives. Note next that since the Σ formulas are closed under bounded quantification, so to are the Δ formulas!

Finally, recall the p-terms from the previous subsection. If $F(\bar{x}, y)$ is Σ and a p-term, then it is Δ (consider what it means to be a p-term). Let $A(y)$ be Δ and $F(\bar{x}, y)$ a Δ p-term. Then it's not hard to show that $A(f(\bar{x}))$ is also Δ. In sum, then, the Δ functions are closed under boolean operations, bounded quantification, and substitution of Σ p-terms.

Invoking the theorem of this section, however, gives an even more surprising result. If $\phi(\bar{x})$ is a Δ-formula and $\bar{n}$ a tuple of natural numbers, then either PA $\vdash \phi(\bar{\mathbf{n}})$ or PA $\vdash \neg\phi(\bar{\mathbf{n}})$. Thus, all Δ-formulas are decidable or, to use Gödel's term, *entscheidungsdefinit*.

2.3.3 Gödel Numbering

Gödel numbering is an encoding scheme; PA cannot, of course, talk directly about its own syntax. It has only the constant 0, functions $s, +, \cdot$, and standard logical machinery to work with–the former which we require to behave as though they were 0, successor, addition, and multiplication in the natural numbers. Nonetheless, if we cleverly encode the symbols of PA as numbers, and sequences of such symbols (formulas) as numbers, and sequences of sequences of such symbols (proofs) as numbers, we may obtain operations and predicates which–interpreting our encoding as intended–are talking about the syntax of PA. Moreover, it's provable that these operations and predicates–while actually only talking about natural numbers–under our encoding interpretation have precisely the characteristics which are associated with their intended metalanguage counterparts. It is in this roundabout sense, then, that PA 'talks' about its own syntax.

An important distinction in what follows is the difference between the Gödel number of some expression, the numeral for some (Gödel) number, and an expression itself. In brief, a Gödel number is a number representing some encoded expression; it is a number in the standard sense which, under our encoding scheme, actually stands for some collection of symbols in PA–possibly a single symbol, possibly a string of symbols, possibly several strings of symbols.

Example 3.
Without going into the specifics of creating an encoding scheme, the Gödel number of '$sss0$' could be 6, while the Gödel number for '$\forall x((sss0 + ssssss0) = x)$' could be 143.

In particular, notice that Gödel numbers are not numerals–they are not, themselves, expressions in PA. To denote the numeral of the Gödel number of some expression, corner brackets are placed around the expression to be encoded (which may include corner brackets itself).

Example 4.
Using the examples from earlier, $\ulcorner sss0 \urcorner$ is **6** *which is 'ssssss0', while $\ulcorner \forall x((sss0+ssssss0) = x) \urcorner$ is* **143** *which is one-hundred and forty three 's' symbols followed by a '0' in the language of PA. Note that if we iterate corner quotes the resulting numeral is different each time: $\ulcorner\ulcorner sss0 \urcorner\urcorner$ is $\ulcorner \mathbf{6} \urcorner$ or $\ulcorner ssssss0 \urcorner$ which is not* **6**.

2.3.4 Bew(x)

Of primary interest in all of this is the PA formula $\text{Bew}(x)$. The association of $\text{Bew}(x)$ with 'x is provable in PA' is owed in large part to the following five properties of $\text{Bew}(x)$:

For all sentences ϕ, ψ of Peano Arithmetic,

(i) If PA $\vdash \phi$, then PA $\vdash \text{Bew}(\ulcorner \phi \urcorner)$

(ii) PA $\vdash \text{Bew}(\ulcorner (\phi \rightarrow \psi) \urcorner) \rightarrow (\text{Bew}(\ulcorner \phi \urcorner) \rightarrow \text{Bew}(\ulcorner \psi \urcorner))$

(iii) PA $\vdash \text{Bew}(\ulcorner \phi \urcorner) \rightarrow \text{Bew}(\ulcorner \text{Bew}(\ulcorner \phi \urcorner) \urcorner)$

(iv) $\text{Bew}(\ulcorner \phi \urcorner)$ is a Σ Sentence

(v) If ϕ is a Σ Sentence, then PA $\vdash \phi \rightarrow \text{Bew}(\ulcorner \phi \urcorner)$

The properties (i), (ii), and (iii) are together known as the Hilbert-Bernays-Löb derivability conditions (for $\text{Bew}(x)$ and PA), so named because, together, they are sufficient to allow, for an arbitrary formula $B(x)$ and theory Z meeting them, the derivation of Gödel's Second Incompleteness Theorem with $B(x)$ in place of $\text{Bew}(x)$ and Z in place of PA (assuming Z satisfies a few facts about the natural numbers).

In this section, the justification for each of (i)-(v) above is sketched–starting, rather appropriately, with (i). Inspecting the construction of $\text{Bew}(x)$ through the standard encoding scheme, it's easy to show that $\text{Bew}(x)$ is, in fact, a Σ formula. Assume, then, that $\text{PA} \vdash \phi$ for some sentence ϕ of PA. Then, $\text{Bew}(\ulcorner\phi\urcorner)$ is also Σ since $\ulcorner\phi\urcorner$ is simply some term (?), and–by our theorem from earlier in the chapter–$\text{PA} \vdash \text{Bew}(\ulcorner\phi\urcorner)$. Note that, en route to proving (i), we also established (iv)! Unfortunately, $\text{Bew}(x)$ cannot be a Δ formula so long as PA is consistent. We turn now to (ii), which takes enough effort to qualify as a theorem:

It's non-obvious to me that substituting any sentence into a sigma formula doesn't break the sigma property...it's not actually a formula...it's a term...

▶ Theorem 11.
If ϕ, ψ are sentences of PA, then $PA \vdash Bew(\ulcorner(\phi \to \psi)\urcorner) \to (Bew(\ulcorner\phi\urcorner) \to Bew(\ulcorner\psi\urcorner))$

Proof Sketch. *A quick perusal of the derivation of $Bew(x)$ shows that PA has a formula that is true if and only if its first parameter is the Gödel number for a proof of its second parameter; moreover, $Bew(x)$ itself is the existential generalization of this formula with respect to the first parameter. It follows that taking the proofs for $\phi \to \psi$ and ϕ and pasting them together (again, definable in PA), gives a proof for ψ. Since all of these operations of are definable with respect to our encoding in PA, it must be that the result is a proof for ψ, and so $Bew(\ulcorner\psi\urcorner)$. Note that every relevant notion has a formalization within PA, and so our result is provable in PA itself.*

Next is (iii) and (v); the justification for these properties relies heavily on the formalization of PA's syntax within PA, and so we confine ourselves to an intuitive justification and sketching the necessary steps. First, we construct the means to represent $\text{Bew}[\phi(\bar{x})]$ for an arbitrary formula $\phi(\bar{x})$ of PA. For some tuple $\bar{a}$, $\text{Bew}[\phi(\bar{a})]$ is true if and only if $\text{PA} \vdash \phi(\bar{a})$ or–equivalently–$\text{PA} \vdash \text{Bew}(\ulcorner\phi(\bar{a})\urcorner)$. This notation in hand, we may now prove using the tools of our formalization of PA,

▶ Lemma 1.
For any formulas $\phi(\bar{x})$ and $\psi(\bar{y})$ of PA, $PA \vdash Bew[(\phi(\bar{x}) \to \psi(\bar{y}))] \to (Bew[\phi(\bar{x})] \to Bew[\psi(\bar{y})])$.

The above, along with the techniques which establish it, can also be used to establish a variation on (i):

How is PA proving a formula?

▶ Lemma 2.
For any formula $\phi(\bar{x})$ of PA, if $PA \vdash \phi(\bar{x})$, then $PA \vdash Bew[\phi(\bar{x})]$

Finally, we establish the following theorem, of which both (iii) and (v) are simply special cases:

▶ Theorem 12.
For any Σ formula $\phi(\bar{x})$ of PA, $PA \vdash \phi(\bar{x}) \to Bew[\phi(\bar{x})]$

Proof Sketch. *Induction on formulas, noting that the two lemmas above allow us to assume that F is a strict Σ sentence (if it isn't it's equivalent to one and we may simply route through that). Here to the details of the formalization of PA come into play; the reader is referred to pg. 47-49 for a fuller sketch.*

2.4 On the Choice of PA

The choice of PA as the formal system in which we work is due largely to the simplicity of describing it and its familiarity. It is neither the weakest system able to formalize its own syntax, nor the weakest capable of proving the generalized diagonal lemma encountered in the next chapter.

Chapter 3

The Box as Bew(x)

The most basic conception underlying provability logic is the idea that the $\Box$ operator of a modal logic could be made to stand for provability in some formal system. In this chapter, we explore this idea and consider the principles such a $\Box$ operator should and should not validate.

3.1 Realizations and Translations

Before beginning, however, two notions will make it easier to formalize our goal.

> **Definition: Realization**
> A function which assigns to each sentence letter a sentence of the language of PA

It's standard to use '$*$' as a variable over realizations; herein, '$\#$' is used as well.

> **Definition: Translation**
> Given a modal sentence ϕ and realization $*$, a translation ϕ^* is given by the following inductive definition:
>
> - $\bot^* = \bot$
> - $p^* = *(p)$ for sentence letters p
> - $(\psi \to \xi)^* = (\psi^* \to \xi^*)$
> - $(\Box\psi)^* = \mathrm{Bew}(\ulcorner \psi^* \urcorner)$
>
> On an informal level, a translation is simply the replacement of sentence letters by the PA language statements which they stand for and $\Box$ by Bew.

■ **Fact.** *If $*$ and $\#$ are realizations such that $*(p) = \#(p)$ for every sentence letter p in a modal sentence ϕ, then $\phi^* = \phi^\#$.*

These definitions in hand, the topic for this chapter can finally be explicitly stated as two related questions; for which modal sentences ϕ is ϕ^* true for every realization $*$ in the standard interpretation of PA? For which modal sentences ϕ is ϕ^* provable in PA for every realization $*$? Our first step in this direction concerns not GL, but rather K4:

Should this really have 'in the standard interpretation'?

▶ Theorem 13.
If K4 $\vdash \phi$, then for every realization $*$, $PA \vdash \phi^*$

Proof. Let $*$ be an arbitrary realization. We show that every axiom schema of K4 is also valid in PA and that the inference rules of K4–modus ponens and necessitation–also hold for all translations in PA. If ϕ is a tautological combination of modal sentences, then so is ϕ^*, and thus PA $\vdash \phi^*$. Assume ϕ is $\Box(\psi \to \xi) \to (\Box\psi \to \Box\xi)$ for some sentences ψ, ξ. Then, ϕ^* is $\text{Bew}(\ulcorner(\psi^* \to \xi^*)\urcorner) \to (\text{Bew}(\ulcorner\psi^*\urcorner) \to \text{Bew}(\ulcorner\xi^*\urcorner))$–but this is simply an instance of (ii) from the last chapter, and so PA $\vdash \phi^*$. PA therefore validates all translations of instances of the K axiom. Assume ϕ is $\Box\psi \to \Box\Box\psi$ for some sentence ψ. Then, ϕ^* is $\text{Bew}(\ulcorner\psi^*\urcorner) \to \text{Bew}(\ulcorner\text{Bew}(\ulcorner\psi^*\urcorner)\urcorner)$–but this is simply an instance of (iii) from the last chapter, and so PA $\vdash \phi^*$. PA therefore validates all translations of instances of the 4 axiom, and thus PA validates all translations of instances of any of the axioms of K4.

We turn, then, to the inference rules of K4. Assume $\text{PA} \vdash \phi^*$ and $\text{PA} \vdash (\phi \to \psi)^*$, for some arbitrary formulas ϕ, ψ of K4. Then, $\text{PA} \vdash (\phi^* \to \psi^*)$ by the definition of $*$, and so $\text{PA} \vdash \psi^*$. It follows that PA validates the application of modus ponens in K4 whenever it validates both the formulas used. Assume $\text{PA} \vdash \phi^*$. Then, by (i) from the previous section, $\text{PA} \vdash \text{Bew}(\ulcorner\phi^*\urcorner)$. But, then, by the definition of $\Box$, $\text{PA} \vdash (\Box\phi)^*$. PA therefore validates necessitation with a sentence of K4 whenever it validates that sentence. Since we've established that PA validates all axioms of K4 and all inference rules of K4 when it validates the sentences used, it follows that, for any sentence ϕ of K4, if $\text{K4} \vdash \phi$, then $\text{PA} \vdash \phi^*$. Note finally that the realization used was arbitrary, and so the theorem follows. □

3.2 The Generalized Diagonal Lemma

In order to strengthen the result above to GL, we first establish a fundamental theorem about PA and other formalized theories:

▶ Theorem 14 (The Generalized Diagonal Lemma)**.**
Suppose that $\bar{y} = y_0, \ldots, y_n$ and $\bar{z} = z_0, \ldots, z_m$ are sequences of distinct variables and that $P_0(\bar{y}, \bar{z}), \ldots, P_n(\bar{y}, \bar{z})$ are formulas of the language of PA. Then, there exists formulas $S_0(\bar{z}), \ldots, S_n(\bar{z})$ of the language of PA such that

$$PA \vdash S_0(\bar{z}) \leftrightarrow P_0(\ulcorner S_0(\bar{z})\urcorner, \ldots, \ulcorner S_n(\bar{z})\urcorner, \bar{z})$$

$$\vdots$$

$$PA \vdash S_n(\bar{z}) \leftrightarrow P_n(\ulcorner S_0(\bar{z})\urcorner, \ldots, \ulcorner S_n(\bar{z})\urcorner, \bar{z})$$

Proof Sketch.
Using our formalization from the last chapter, define $su(w, \bar{x}, y)$ as a Σ p-term for the $n+2$-place substitution function whose value at $a, b_0, \ldots, b_n$ is the numeral of the Gödel number of the result of substituting the numerals $\mathbf{b_0}, \ldots, \mathbf{b_n}$ for the variables $x_0, \ldots, x_n$ in the formula with Gödel number a.

For each $i \leq n$, let $k_i(\bar{x}, \bar{z})$ be $P_i(su(x_0, \bar{x}), \ldots, su(x_n, \bar{x}), \bar{z})$. Interpreting this loosely, '$su(x_i, \bar{x})$' says to take a tuple $\bar{x}$ of numbers representing formulas, and substitute those numbers into the ith formula for its free variables. $k_i(\bar{x}, \bar{z})$, then, is asking whether P_i holds of the n numbers which encode the substitution above for each i and the tuple $\bar{z}$. Next, define $S_i(\bar{z})$ as $P_i(su(\mathbf{k_0}, \bar{\mathbf{k}}), \ldots, su(\mathbf{k_n}, \bar{\mathbf{k}}), \bar{z})$. Less formally, for each k_i defined above, we take the Gödel number of the result of substituting the tuple of Gödel numbers $\bar{\mathbf{k}}$ in for its variables.

Something is weird about the definitions they're giving...

▶ **Corollary.**
Suppose that $\bar{y} = y_0, \ldots, y_n$ is a sequence of distinct variables and that $P_0(\bar{y}), \ldots, P_n(\bar{y})$ are formulas of the language of PA. Then, there exists sentences $S_0, \ldots, S_n$ of the language of PA such that

$$PA \vdash S_0 \leftrightarrow P_0(\ulcorner S_0 \urcorner, \ldots, \ulcorner S_n \urcorner)$$
$$\vdots$$
$$PA \vdash S_n \leftrightarrow P_n(\ulcorner S_0 \urcorner, \ldots, \ulcorner S_n \urcorner)$$

▶ **Corollary** (The Diagonal Lemma)**.**
Suppose that y is a distinct variable and that $P(y)$ is a formula of the language of PA. Then, there exists a sentence S of the language of PA such that

$$PA \vdash S \leftrightarrow P(\ulcorner S \urcorner)$$

3.3 Löb's Theorem

In 1953, the logician Leon Henkin questioned whether, taking $P(y)$ to be $\mathrm{Bew}(x)$ and obtaining $PA \vdash S \leftrightarrow \mathrm{Bew}(\ulcorner S \urcorner)$ for some sentence S, the sentence S is itself provable in PA. Only a year later, M.H. Löb showed that for all sentences S, if $PA \vdash \mathrm{Bew}(\ulcorner S \urcorner) \rightarrow S$, then PA $\vdash S$, a result known as Löb's theorem. This result was not, however, an entirely expected one; there seems to be no *a priori* reason why PA shouldn't claim to be sound with respect to a proposition which it cannot actually prove; indeed, it even seems natural that $\mathrm{Bew}(\ulcorner S \urcorner) \rightarrow S$ should be true for any S. As Löb's theorem shows, however, PA is incredibly humble in this respect; it never claims to be sound with respect to a proposition unless it must, unless it can actually prove the proposition.

Before actually moving to the proof of Löb's theorem, we briefly present Curry's paradox which bears an eerie resemblance to our eventual proof.

Definition: Curry's Paradox
Let SC denote 'Santa Claus exists'. Define $c = \{x : x \in x \rightarrow SC\}$. Assume that $c \in c$. Then, by the definition of c, $c \in c \rightarrow SC$. Thus, by modus ponens, SC. It follows, then, that $c \in c \rightarrow SC$, and so $c \in c$. By modus ponens, then, SC.

A reformulated version of the paradox using Tarski's truth schema:

'ϕ' is true if and only if ϕ

is owed to Henkin and often referred to as Henkin's paradox:

Definition: Henkin's Paradox
Let SC denote 'Santa Claus exists' and S denote the sentence 'if S is true, then SC'. Assume S is true; then, by definition, 'if S is true, then SC' is true. Thus, by modus ponens, SC. It follows, then, that 'if S is true, then SC' is true–which is to say, S is true. By modus ponens, it follows that SC.

The cyclically self-referential nature of both Curry and Henkin's paradox arises yet again in Löb's theorem.

▶ **Theorem 15** (Löb's Theorem)**.**
If $PA \vdash Bew(\ulcorner \phi \urcorner) \rightarrow \phi$, then $PA \vdash \phi$

Proof.
Applying the diagonal lemma to $(\text{Bew}(x) \rightarrow \phi)$, there is a sentence ψ such that

$$\text{PA} \vdash \psi \leftrightarrow (\text{Bew}(\ulcorner \psi \urcorner) \rightarrow \phi)$$

And so, by definition:

$$\text{PA} \vdash \psi \rightarrow (\text{Bew}(\ulcorner \psi \urcorner) \rightarrow \phi)$$

By property (i) of $\text{Bew}(x)$,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \rightarrow (\text{Bew}(\ulcorner \psi \urcorner) \rightarrow \phi) \urcorner)$$

By property (ii) of $\text{Bew}(x)$,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \urcorner) \rightarrow \text{Bew}(\ulcorner \text{Bew}(\ulcorner \psi \urcorner) \rightarrow \phi \urcorner)$$

By property (ii) of $\text{Bew}(x)$ again,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \urcorner) \rightarrow (\text{Bew}(\ulcorner \text{Bew}(\ulcorner \psi \urcorner) \urcorner) \rightarrow \text{Bew}(\ulcorner \phi \urcorner))$$

Taking the following instance of property (iii) of $\text{Bew}(x)$,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \urcorner) \rightarrow \text{Bew}(\ulcorner \text{Bew}(\ulcorner \psi \urcorner) \urcorner)$$

Combining the two previous formulas,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \urcorner) \rightarrow \text{Bew}(\ulcorner \phi \urcorner)$$

Assume that $\text{PA} \vdash \text{Bew}(\ulcorner \phi \urcorner) \rightarrow \phi$. Then, using the above,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \urcorner) \rightarrow \phi$$

But then, by our very first formula,

$$\text{PA} \vdash \psi$$

and by property (i) of $\text{Bew}(x)$,

$$\text{PA} \vdash \text{Bew}(\ulcorner \psi \urcorner)$$

Finally, by two formulas previous,

$$\text{PA} \vdash \phi$$

□

▶ **Corollary** (Gödel's Second Incompleteness Theorem).
If PA is consistent, then PA $\nvdash \neg \text{Bew}(\ulcorner \bot \urcorner)$

Proof.
Assume that PA is consistent and that $\text{PA} \vdash \neg\text{Bew}(\ulcorner \bot \urcorner)$. Then, trivially, $\text{PA} \vdash \text{Bew}(\ulcorner \bot \urcorner) \rightarrow \bot$. But then, by Löb's theorem, $\text{PA} \vdash \bot$–a contradiction. It must be, then, that $\text{PA} \nvdash \neg\text{Bew}(\ulcorner \bot \urcorner)$, and thus the theorem is established. □

3.4 K4LR

In analogy to Löb's theorem, define:

> **Definition: Löb's Rule**
> The modal logic rule of inference given by 'If $\vdash \Box A \to A$, then $\vdash A$'; abbreviated LR.

Using the standard naming scheme for modal logics, we therefore have K4LR as the result of adding Löb's rule as a rule of inference to K4. How does this new logic relate to PA and GL?

> ▶ **Theorem 16.**
> *If K4LR $\vdash \phi$, then PA $\vdash \phi^*$ for all realizations $*$*

Proof Sketch. *Using the first theorem of this chapter, if K4 $\vdash \phi$, then for all realizations $*$, PA $\vdash \phi^*$. Note further, however, that the only new inferences licensed in K4LR are those from $\Box\phi \to \phi$ to ϕ–but, by Löb's theorem, this schema is valid for all sentences of PA.*

> ▶ **Theorem 17.**
> *If K4LR $\vdash \phi$, then GL $\vdash \phi$*

Proof Sketch. *Note that it is sufficient to show that LR is valid in GL. Assume GL $\vdash \Box\phi \to \phi$. Then, by necessitation, $\Box(\Box\phi \to \phi)$. Noting that $\Box(\Box\phi \to \phi) \to \Box\phi$ is an instance of Löb's axiom, by modus ponens $\Box\phi$. But then, by modus ponens with our assumption, ϕ.*

Is KLR equivalent to GL? I don't use the 4 axiom...

> ▶ **Theorem 18.**
> *If GL $\vdash \phi$, then K4LR $\vdash \phi$*

Proof Sketch. *Note that it is sufficient to show that Löb's axiom is valid in K4LR. Without using any axioms associated with GL,*

Proof.

1	$\Box(\Box\phi \to \phi)$	Assump.
2	$\Box\phi \to \phi$	$\Box$-reit 1
3	ϕ	LR 2
4	$\Box\phi$	$\Box$-intro 3
5	$\Box(\Box\phi \to \phi) \to \Box\phi$	$\to$-intro 1-4

□

It follows, then, that K4LR and GL are, in truth, the same logic. Thus,

> ▶ **Corollary.**
> *If GL $\vdash \phi$, then PA $\vdash \phi^*$ for all realizations $*$*

It's natural to ask whether the converse of the above holds–and, in fact, it does. Showing that this is the case, however, is difficult and will have to wait several chapters. For the moment, we settle for naming the class of modal sentences with the desired property:

Definition: Always Provable
A modal sentence ϕ is *always provable* if and only if PA $\vdash \phi^*$ for every realization $*$

3.5 The Box as Provability

At long last there is sufficient evidence to not only interpret $\text{Bew}(x)$ as provability in PA, but also the $\Box$ of GL as $\text{Bew}(x)$. For the remainder of this chapter and, indeed, the book, we use this interpretation with abandon. In particular, suppose that ϕ, ψ are sentences of PA; then, the following are intended:

Assertion	Arithmetization
ϕ is provable (in PA)	$\text{Bew}(\ulcorner \phi \urcorner)$
ϕ is consistent (with PA)	$\neg\text{Bew}(\ulcorner \neg\phi \urcorner)$
ϕ is unprovable (in PA)	$\neg\text{Bew}(\ulcorner \phi \urcorner)$
ϕ is disprovable (in PA)	$\text{Bew}(\ulcorner \neg\phi \urcorner)$
ϕ is decideable (in PA)	$\text{Bew}(\ulcorner \phi \urcorner) \vee \text{Bew}(\ulcorner \neg\phi \urcorner)$
ϕ is undecideable (in PA)	$\neg\text{Bew}(\ulcorner \phi \urcorner) \wedge \neg\text{Bew}(\ulcorner \neg\phi \urcorner)$
ϕ is equivalent to ψ (in PA)	$\text{Bew}(\ulcorner (\phi \leftrightarrow \psi) \urcorner)$
ϕ implies ψ (in PA)	$\text{Bew}(\ulcorner (\phi \to \psi) \urcorner)$
PA is consistent	$\text{Bew}(\ulcorner \bot \urcorner)$
PA is inconsistent	$\neg\text{Bew}(\ulcorner \bot \urcorner)$

► Theorem 19.
PA proves the arithmetization of Gödel's Second incompleteness theorem for PA

Proof. Using the translations above, our target–in GL– is $\neg\Box\bot \to \neg\Box\neg\Box\bot$.

1	$\neg\Box\bot$	Assump.
2	$\Box\neg\Box\bot$	Assump.
3	$\neg\Box\bot$	$\Box$-reit 2
4	$\Box\bot$	Assump.
5	$\neg\Box\bot$	Reit 3
6	$\bot$	$\bot$-intro 4,5
7	$\Box\bot \to \bot$	$\to$-intro 4-6
8	$\Box(\Box\bot \to \bot)$	$\Box$-intro 3-7
9	$\Box(\Box\bot \to \bot)$	Assump.
10	$\Box(\Box\bot \to \bot) \to \Box\bot$	GL
11	$\Box\bot$	MP 9,10
12	$\neg\Box\bot$	Reit 1
13	$\neg\Box(\Box\bot \to \bot)$	Reductio 2-4
14	$\neg\Box\neg\Box\bot$	Reductio 2-13
15	$\neg\Box\bot \to \neg\Box\neg\Box\bot$	$\to$-intro 1-14

Noting now that PA proves all realizations of the theorems of GL, PA $\vdash \neg\mathrm{Bew}(\ulcorner\bot\urcorner) \to \neg\mathrm{Bew}(\ulcorner\neg\mathrm{Bew}(\ulcorner\bot\urcorner)\urcorner)$ as requested. □

A similar argument establishes the following theorem,

► **Theorem 20.**
PA proves the arithmetization of 'if the inconsistency of PA is not provable, then the consistency of PA is undecidable'.

Löb's theorem was proved earlier for PA, and the aware reader may have noticed an intimate connection to Löb's axiom $\Box(\Box\phi \to \phi) \to \Box\phi$ from our axiomatization of GL. Löb's axiom is the GL equivalent of *formalized* Löb's theorem: for any ϕ, PA $\vdash \mathrm{Bew}(\ulcorner\mathrm{Bew}(\ulcorner\phi\urcorner) \to \phi\urcorner) \to \mathrm{Bew}(\ulcorner\phi\urcorner)$. Given our previous results, it's easy to show using Löb's axiom that formalized Löb's theorem does indeed hold for PA. For ease of reference, a name is introduced for the kind of formulas used in Löb's theorem:

Why is this the '[]' instead of the Gödel numbering?

Definition: Reflection
Let ϕ be a sentence of PA. The reflection principle for ϕ or reflection for ϕ is the PA sentence $\mathrm{Bew}[\phi] \to \phi$

Löb's theorem therefore states that for any sentence ϕ of PA, if reflection for ϕ is provable in PA, then ϕ is provable.

3.6 GLS

Taking some time to review our previous work and thinking about the intended model of PA–the standard conception of the natural numbers, symbolized by '$\mathcal{N}$'–leads to the following realization: anything provable in PA is a theorem of PA, and thus true in $\mathcal{N}$. That is, for all realizations $*$, the translation of $(\Box\phi \to \phi)^*$ holds in the model $\mathcal{N}$. Taking this further, we may ask what logic corresponds to those formulas which are *always true* in $\mathcal{N}$, what is the logic of formulas which are true in $\mathcal{N}$ irrespective of realization. In this spirit, we introduce the system GLS, short for Gödel-Löb-Solovay.

Definition: GLS
The logic generated by the following axioms,

- All theorems of GL
- The T axiom: $\Box\phi \to \phi$

And allowing the following rules of inference,

- Modus Ponens: If $\vdash_{GLS} \phi \to \psi$ and $\vdash_{GLS} \phi$, then $\vdash_{GLS} \psi$

along with universal substitution.

The theorem below follows easily from the considerations which motivated the inspection of GLS in the first place.

► **Theorem 21.**
If GLS $\vdash \phi$, then, for every realization $*$, ϕ^* is true

GLS is, unlike GL, not a normal modal logic; in particular, it fails to validate necessitation. On the other hand, GLS validates some inference rules not allowed by GL. For instance, if GLS $\vdash \phi$ for any ϕ, then GLS $\vdash \Diamond\phi$. LOOK AT 67...what does that mean?

Chapter 4

Model Theory for GL

The semantics for GL are simply the Kripke semantics that have come to dominate normal modal logic model theory; a full presentation of basic definitions and results is omitted here. Rather, we move to discuss the particular case of GL:

> **Definition: Well-Founded**
> A relation R is *well-founded* if and only if for every non-empty set X there is $w \in X$ such that $\nexists y \in X$ where yRw.

> **Definition: Converse Well-Founded**
> A relation R is *converse well-founded* if and only if for every non-empty set X there is $w \in X$ such that $\nexists y \in X$ where wRy.

Less formally, a relation is converse well-founded if and only if there doesn't exist a infinite walk along the relation; every path eventually terminates. Note that converse well-founded relations must be irreflexive or else an infinite walk is guaranteed (simply loop at the same world).

When proving results over converse well-founded relations (say, R), it's often useful to make use of *induction on the converse of* R. This proof technique is simply to prove that an arbitrary element w of the total set W has the requested property ψ if every y such that wRy has the property ψ. Note that this suffices to show all of W has ψ since–by R converse well-founded–there must be a world ψ which doesn't relate to any others, and thus is ψ. But then, there must be world which related to, at most, only this ψ world (again, by R converse well-founded), and thus is also ψ. So on and so forth.

► Theorem 22.
The schema '$\Box(\Box\phi \to \phi) \to \Box\phi$' is valid on a frame $\langle W, R\rangle$ if and only if R is transitive and converse well-founded.

Proof.
$(\to)$
Assume that the schema '$\Box(\Box\phi \to \phi) \to \Box\phi$' is valid on a frame $\langle W, R\rangle$. Then, since GL combined only this schema with the axioms of K which are valid on all frames, all theorems of GL are valid. It follows, then, that the schema $\Box\phi \to \Box\Box\phi$ is valid on the frame, and so by a standard modal logic result, the frame must be transitive.

Assume that R is not converse well-founded. Then there exists some subset of W, say X, containing an infinite walk. Let p be an atomic sentence and take the valuation which makes p true at all and only $w \in W$ such that $w \notin X$. Let $x \in X$. Let y be such that xRy. Then, either $y \in X$ or not. Assume

$y \in X$. Then, by construction, $y \models \neg p$. Note further that y must also a world in the walk–which also models $\neg p$. So $y \models \Diamond\neg p$. But, by definition, this is $\neg\Box\neg\neg p$ or rather $\neg\Box p$. Thus, $y \models \neg\Box p$, and thus $y \models \Box p \to p$ by contraposition. Assume, then, that $y \notin X$. Then $y \models p$, and so $y \models \Box p \to p$. It follows, then, that in either case $y \models \Box p \to p$, and so $x \models \Box(\Box p \to p)$. Note, however, that $x \models \neg\Box p$ since it must access the next world in the infinite walk which is a $\neg p$ world. $\bot$.

$(\leftarrow)$
Assume R is transitive and converse well-founded. Let $w \in W$ and assume that every x such that wRx models $\Box(\Box\phi \to \phi) \to \Box\phi$. Assume $\Box(\Box\phi \to \phi)$ is true at w. Then, $\Box\phi \to \phi$ is true at every world accessible from w. Note further that, by transitivity, $\Box\phi \to \phi$ is true at every world accessible from every world accessible from w. Thus, $\Box(\Box\phi \to \phi)$ is true at every world accessible from w. Then, since all these worlds model $\Box(\Box\phi \to \phi) \to \Box\phi$, $\Box\phi$ is true at all of these worlds. Since these worlds all contain $\Box\phi \to \phi$ and $\Box\phi$, they must also all contain ϕ. It follows immediately that $w \models \Box(\Box\phi \to \phi) \to \Box\phi$. □

▶ Theorem 23.
Suppose $F = \langle W, R\rangle$ is finite and transitive. Then, F is converse well-founded if and only if F is irreflexive.

▶ Theorem 24.
If $GL \vdash \phi$, then ϕ is valid in not only all transitive and converse well-founded frames, but also all finite transitive and irreflexive frames.

Many of the frame properties which arise in a modal context are first-order, meaning they correspond to the frame validity of some first-order sentence of the language; unfortunately, the property of being converse well-founded is not a first-order property.

▶ Theorem 25.
The class of converse well-founded Kripke models is not first-order definable

EXCEPT GL IS NOT COMPACT

Proof.
Assume that the property of being converse well-founded is first-order; let ρ be the first-order sentence to which it corresponds and R the accessibility relation of the model. Note next that taking infinitely many constants $c_1, c_2, c_3, \ldots$, the theory $\{\rho\} \cup \{c_i R c_j : i < j\}$ is finitely satisfiable (take the obvious converse well-founded model). But, by the compactness theorem, there must then be a model of the entire theory–but such a model is not converse well-founded (simply move c_i to c_{i+1} each time) and models ρ! □

A nearly identical argument shows that the transitive and converse well-founded frames also fail to be equivalent to the frame validity of a first-order sentence.

Chapter 5

Completeness & Decidability of GL

In this chapter, we establish the completeness of GL with respect to the class of transitive, converse well-founded finite Kripke models; that is, we show that any sentence valid over the class of transitive, converse well-founded finite Kripke models is, in fact, a theorem of GL. We will eventually prove this by proving the contrapositive; that is, if ϕ is not a theorem of GL, then it is not valid over the class of transitive, converse well-founded finite Kripke models. Before doing so, however, it's necessary to establish several definitions and smaller results.

> **Definition: ϕ-Formula**
> For a given sentence ϕ, a modal sentence ψ is a ϕ-formula if and only if ψ is a subsentence of ϕ or ψ is the negation of a subsentence of ϕ

> **Definition: GL-Consistent**
> A set of ϕ-formulas Σ is GL-consistent if and only if GL $\nvdash \neg(\bigwedge \Sigma)$

> **Definition: Maximally GL-Consistent**
> A set of ϕ-formulas Σ is maximally GL-consistent with respect to ϕ if and only if for every subsentence ψ of ϕ, either ψ or $\neg\psi$ is in Σ and Σ is GL-consistent

It follows easily that if ψ is a subsentence of ϕ and X is a maximally GL-consistent set with respect to ϕ, then $\psi \in X$ if and only if $\neg\psi \notin X$. Note further that if $\psi_1, \ldots, \psi_n \in X$ and GL $\vdash \psi_1 \wedge \cdots \wedge \psi_n \to \chi$ with χ a subsentence of ϕ, then $\chi \in X$ as well. Finally, note that every consistent set is included in some maximally consistent set.

► Theorem 26.
If a sentence ϕ is valid over the class of transitive, converse well-founded finite Kripke models, then ϕ is a theorem of GL.

EXCEPT GL IS NOT COMPACT

Proof.
Let ϕ be an arbitrary sentence that is not a theorem of GL. We now proceed to construct a transitive, converse well-founded finite Kripke model which falsifies ϕ. Let W be the set of all distinct, maximally GL-consistent sets with respect to ϕ. Note that since ϕ is itself finite, there are only finitely many ϕ-formulas and so only finitely many distinct, maximally GL-consistent sets; W is thus finite, and it

follow easily from our assumptions that it is also non-empty. As our choice of variable has hinted, we will take W as the set of worlds in our model and each maximally consistent set as a world; the valuation V on the model is defined as follows: for every sentence letter p and every $w \in W$, let p be true at w if and only if p appears in ϕ and $p \in w$. Finally, we define the accessibility relation R on W. For two worlds $w, x \in W$, let wRx if and only if for every sentence of the form $\Box\psi$ in w, both $\Box\psi$, $\psi \in x$ and there is some sentence of the form $\Box\xi$ in x such that $\neg\Box\xi \in w$. R is clearly transitive and irreflexive, so by a theorem from last section, R is converse well-founded.

We now show that for every subsentence of ϕ of the form $\Box\psi$ and every $w \in W$, $\Box\psi \in w$ if and only if $\psi \in x$ for every x accessible from w–the defining characteristic of the $\Box$ operator. The forward direction follows readily from the construction of R so consider the converse. Let $w \in W$ and $\Box\psi$ a subsentence of ϕ. Assume that $\psi \in x$ for every x accessible from w. Let y be the set $\{\Box\xi, \xi : \Box\xi \in w\} \cup \{\neg\psi, \Box\psi\}$. If y is GL-consistent, then it can be extended to a maximally GL-consistent set with respect to ϕ (see last section), and–by construction–w accesses a $\neg\psi$ world, a contradiction. It must be, then, that y is inconsistent. By compactness, there must be some finite subset of the ξ and $\Box\xi$ along with $\neg\psi$ and $\Box\psi$ which generates the contradiction–that is for some $\xi_1, \ldots, \xi_k$ used, GL $\vdash \neg(\xi_1 \wedge \Box\xi_1 \wedge \cdots \wedge \xi_k \wedge \Box\xi_k \wedge \Box\psi \wedge \neg\psi)$. Which, by the standard propositional logic proof, is the same as GL $\vdash \xi_1 \wedge \Box\xi_1 \wedge \cdots \wedge \xi_k \wedge \Box\xi_k \rightarrow (\Box\psi \rightarrow \psi)$, and by necessitation and the K-axiom, GL $\vdash \Box\xi_1 \wedge \Box\Box\xi_1 \wedge \cdots \wedge \Box\xi_k \wedge \Box\Box\xi_k \rightarrow \Box(\Box\psi \rightarrow \psi)$. Noting the top-level consequent is the antecedent of an instance of the GL-axiom, GL $\vdash \Box\xi_1 \wedge \Box\Box\xi_1 \wedge \cdots \wedge \Box\xi_k \wedge \Box\Box\xi_k \rightarrow \Box\psi$. Noting finally that the 4-axiom is valid in GL, GL $\vdash \Box\xi_1 \wedge \cdots \wedge \Box\xi_k \rightarrow \Box\psi$. But then, by our discussion just before the current theorem, $\Box\psi \in w$.

At long last, letting $\mathcal{M} = \langle W, R, V \rangle$ we have that $\mathcal{M}$ is a transitive, converse well-founded finite Kripke model. Note, however, that $\mathcal{M}$ also has the interesting property that, for every subsentence ψ of ϕ and $w \in W$, $\psi \in w$ if and only if $w \models \psi$ (recall our valuation!). Noting that $\neg\phi$ is a ϕ-formula and–by assumption–ϕ is not a theorem of GL, $\{\neg\phi\}$ is a GL-consistent set, and so can be extended to a maximally GL-consistent set. Given our definition of $\mathcal{M}$, this extension is a world $v \in W$ such that $v \models \neg\phi$. □

▶ Corollary.
Whether or not an arbitrary modal sentence is a theorem of GL is decidable

Proof.
Note that the proof above shows that if a sentence is not a theorem there is a finite model which has a world where it fails; more importantly, this finite model has a size that is upperbounded by 2^n where n is the number of subsentences of the sentence in question–again, by the construction above. It follows that a brute force search of all models of size $\leq 2^n$ is sufficient to determine whether or not a sentence is a theorem of GL. □

▶ Theorem 27.
A sentence is valid in all finite transitive and irreflexive frames if and only if it is valid in all finite transitive and irreflexive frames that are trees.

Proof and discussion omitted pending usefulness; pg. 84

Chapter 6

Canonical Models

Pending relevance to GL; pg. 85-91. Eventually to be included in my model theory and modal logic notes.

Chapter 7

On GL

Within this section we present a number of disparate results, some with applications to the study of provability and some that are simply interesting in their own right.

7.1 Normal Form for Letterless Sentences

> **Definition: Letterless Sentence**
> A well-formed formula of GL using only '$($', '$)$', '$\perp$', '$\rightarrow$', '$\Box$', '$\wedge$', '$\vee$', '$\neg$', and '$\perp$'; that is, without sentence letters.

For the sake of brevity, we adopt the following recursively-defined notation: '$\Box^0\phi$' is simply 'ϕ' while '$\Box^{i+1}\phi$' is '$\Box\Box^i\phi$'. By the *normal form for a letterless sentence* ϕ, we understand a truth functional combination of sentences of the form $\Box^i\perp$.

▶ Theorem 28.
For any sentence ϕ of GL, there is a normal form letterless sentence ψ such that $GL \vdash \phi \leftrightarrow \psi$

Proof.
It suffices to show how to construct a normal form letterless sentence ψ for $\Box\phi$ where ϕ is itself an arbitrary normal form letterless sentence (consider the obvious induction on formulas). To begin, place ϕ in conjunctive normal form (note that connectives only occur at the highest level, and so ϕ's modal nature never comes into play). Since GL validates distribution of the $\Box$ over $\wedge$ (GL $\vdash \Box(\xi_1 \wedge \cdots \wedge \xi_k) \leftrightarrow \Box\xi_1 \wedge \cdots \wedge \Box\xi_k$), we may distribute the $\Box$ of $\Box\phi$ over ϕ's CNF equivalent. It now suffices for us to construct a normal-form sentence equivalent–say ψ_i–to $\Box\phi_i$ (one of the clauses of our CNF formula ϕ, we may imagine) where, by assumption, ϕ_i is an arbitrary disjunction of $\Box^j\perp$ and $\neg\Box^{j'}\perp$ sentences. Separating the negated sentences from the non-negated, ϕ_i has the form,

$$\Box^{n_1}\perp \vee \cdots \vee \Box^{n_p}\perp \vee \neg\Box^{m_1}\perp \vee \cdots \vee \neg\Box^{m_q}\perp$$

If no disjunct of ϕ appears unnegated, let ϕ_i be $\perp \vee \phi_i$. Thus ϕ_i has at least one unnegated disjunct, $\Box^{n_r}\perp$. Since, leveraging the 4 axiom, GL $\vdash \Box^i\perp \rightarrow \Box^j\perp$ whenever $i \leq j$, consider instead of ϕ_i,

$$\Box^n\perp \vee \neg\Box^m\perp$$

where $n = \max\{n_1, \ldots, n_p\}$ and $m = \min\{m_1, \ldots, m_q\}$. We now show that $\Box\phi_i$ is equivalent to either $\Box^{n+1}\perp$ or $\top$–either of which is in normal form. If $\neg\Box^m\perp$ doesn't appear above (no negated disjuncts in ϕ_i), it's easy to show that GL $\vdash \Box^{n+1}\bot \leftrightarrow \Box\phi_i$. Assume, then, that $\neg\Box^m\bot$ does appear. Rewrite $\Box^n\bot \vee \neg\Box^m\bot$ as $\Box^m\bot \to \Box^n\bot$. If $m \leq n$, $\Box^m\bot \to \Box^n\bot$ is a theorem of GL, and it's easy to show that GL $\vdash \Box\phi_i \leftrightarrow \top$. Assume then that $n < m$, and so $n+1 \leq m$. As earlier,

$$GL \vdash (\Box^m\bot \to \Box^n\bot) \to (\Box^{n+1}\bot \to \Box^n\bot)$$

By necessitation and the K axiom,

$$GL \vdash \Box(\Box^m\bot \to \Box^n\bot) \to \Box(\Box^{n+1}\bot \to \Box^n\bot)$$

Noting that the consequent is the antecedent of an instance of Löb's axiom,

$$GL \vdash \Box(\Box^m\bot \to \Box^n\bot) \to \Box^{n+1}\bot$$

It's trivial to show the converse, and so we have:

$$GL \vdash \Box(\Box^m\bot \to \Box^n\bot) \leftrightarrow \Box^{n+1}\bot$$

From which it follows easily that

$$GL \vdash \Box\phi_i \leftrightarrow \Box^{n+1}\bot$$

$\Box$

7.1.1 Constant Sentences

Definition: Constant Sentence
A sentence ϕ of PA is called a constant sentence if and only if it is a member of the smallest class of sentences containing $\bot$ and, whenever it contains S and S', $(S \to S')$ and $\mathrm{Bew}(\ulcorner S \urcorner)$.

As a brief insight, note that the letterless sentences of GL are constant across all realizations and that constant sentences translate to letterless sentences.

Example 5. *Constant sentences.*

$$\bot$$

$$Bew(\ulcorner \bot \urcorner)$$

$$\neg Bew(\ulcorner \bot \urcorner)$$

$$\neg Bew(\ulcorner \neg Bew(\ulcorner \bot \urcorner) \urcorner)$$

Harvey Friedman introduced the notion of the constant sentences and questioned whether or not there exists an effective method for determining their truth; a brief perusal of the examples shows that the class of constant sentences, despite its simple definition, contains many sentences of importance, and so Friedman's question was far from unmotivated. The answer, it turns out, is actually 'yes'. Moreover, the algorithm for doing so is wonderfully concise:

For an arbitrary constant sentence ϕ, find a letterless ψ of GL such that $\psi^* = \phi$. Put ψ into normal form (which we just showed is always possible!); note further that $\Box^i\bot$ has the same truth value as $\bot$ for all $i \geq 0$. Delete, then, every $\Box$ from ψ and evaluate this formula a la propositional calculus. We obtain $\top$ if and only if ϕ is true! To go even further, we may decide whether a constant sentence ϕ is provable by finding a letterless ψ of GL such that $\psi^* = \phi$ and deciding the truth of $\Box\psi$!

7.2 Rank and Trace

Is this right??????!?!?!?!?!?!

Definition: Rank
Let $\langle W, R\rangle$ be a finite, transitive, and irreflexive frame. For any $w \in W$, the rank of w, formalized as $\rho_{\langle W,R\rangle}(w)$ or simply $\rho(w)$ is the integer corresponding to the number of steps in the longest possible walk through the frame (along R) starting from w.

Since the frame is finite, transitive, and irreflexive–and thus, converse well-founded–$\rho(w)$ is finite for all w. It's similarly easy to show that $\rho(w) < |W|$ and if wRx, $\rho(w) \geq \rho(x) + 1$.

► Lemma 3.
If $\rho(w) > i$, then for some x, wRx and $\rho(x) = i$.

Proof Sketch.
Assume $\rho(w) = n > i$. Then, by definition, there is some walk $w = w_n R w_{n-1} R \ldots R w_1 R w_0$ through the model. Considering an arbitrary w_i, wRw_i by transitivity; note the consequences if $\rho(w_i) \neq i$.

Definition: Trace
Let ϕ be a letterless sentence. The trace of ϕ, $[\![\phi]\!]$, is a set of natural numbers defined inductively as follows:

$$
\begin{aligned}
[\![\bot]\!] &= \emptyset \\
[\![\neg\psi]\!] &= \mathbb{N} - [\![\psi]\!] \\
[\![\psi \wedge \psi']\!] &= [\![\psi]\!] \cap [\![\psi']\!] \\
[\![\psi \vee \psi']\!] &= [\![\psi]\!] \cup [\![\psi']\!] \\
[\![\psi \to \psi']\!] &= (\mathbb{N} - [\![\psi]\!]) \cup [\![\psi']\!] \\
[\![\Box\psi]\!] &= \{n \in \mathbb{N} : \forall i < n, i \in [\![\psi]\!]\}
\end{aligned}
$$

A set X of natural numbers is said to be *cofinite* if and only if $\mathbb{N} - X$ is finite; thus, just as any subset of a finite set is finite, so is any superset of a cofinite set, itself cofinite.

► Lemma 4.
For every letterless ϕ, $[\![\phi]\!]$ is either finite or cofinite.

Proof Sketch.
Induction on formulas.

► Lemma 5.
Let $\mathcal{M}$ be a finite, transitive, and irreflexive model, $w \in W$, and ϕ letterless. Then, $\mathcal{M}, w \models \phi$ if and only if $\rho(w) \in [\![\phi]\!]$.

Proof Sketch.
Induction on formulas.

► Lemma 6.
If ϕ is letterless, then $GL \vdash \phi$ if and only if $[\![\phi]\!] = \mathbb{N}$

Proof Sketch.
In the forward direction, not that for every n there exists a finite, transitive, and irreflexive model $\mathcal{M}$ such that for some $w \in W$, $\rho(w) = n$. The backward direction is similarly simple.

Note that, from the lemma above and the definition of trace, we also have that if ϕ and ψ are letterless, $[\![\phi]\!] \subseteq [\![\psi]\!]$ if and only if GL $\vdash \phi \to \psi$ and $[\![\phi]\!] = [\![\psi]\!]$ if and only if GL $\vdash \phi \leftrightarrow \psi$.

► **Lemma 7.**
For every n, $[\![\Box^n \bot]\!] = \{m \in \mathbb{N} : m < n\}$

Proof Sketch.
Induction on n.

► **Corollary.**
For every n, $[\![\neg(\Box^{n+1}\bot \to \Box^n \bot)]\!] = \{n\}$

► **Lemma 8.**
Suppose $[\![\phi]\!]$ is finite. Let $\psi = \bigvee\{\neg(\Box^{n+1}\bot \to \Box^n\bot) : n \in [\![\phi]\!]\}$. Then, $GL \vdash \phi \leftrightarrow \psi$.

Proof Sketch.
Use lemma 4 and the corollary to lemma 6 to construct a string of biconditionals showing an arbitrary finite, transitive, and irreflexive model and arbitrary world in it validate $\phi \leftrightarrow \psi$; invoke completeness.

► **Lemma 9.**
Suppose $[\![\phi]\!]$ is cofinite. Let $\psi = \bigwedge\{\Box^{n+1}\bot \to \Box^n\bot : n \notin [\![\phi]\!]\}$. Then, $GL \vdash \phi \leftrightarrow \psi$.

Proof Sketch.
Mirror the previous lemma.

Combining lemmas 3, 7, and 8 gives another proof of the normal form theorem for letterless sentences established earlier. Of more note, combined with lemma 5, they give the letterless cases of Solovay's completeness for GL and GLS–our next two results.

► **Theorem 29.**
Let ϕ be letterless, $*$ arbitrary. Then, $GLS \vdash \phi$ if and only if $\mathcal{N} \models \phi^*$.

Proof.
The forward direction has already been established. Consider, then, the reverse direction. If $[\![\phi]\!]$ is finite, then by lemma 8, $\mathrm{GL} \vdash \phi \leftrightarrow \bigvee\{\neg(\Box^{n+1}\bot \to \Box^n\bot) : n \in [\![\phi]\!]\}$, and, since $(\Box^{i+1}\bot)^*$ is false in PA, ϕ^* must be false as well. If ϕ^* is true, it must be, then, that $[\![B]\!]$ is not finite–and thus, co-finite. Then, by lemma 9, $\mathrm{GL} \vdash \phi \leftrightarrow \bigwedge\{\Box^{n+1}\bot \to \Box^n\bot : n \notin [\![\phi]\!]\}$. But then, by definition, this a theorem of GLS as well. Moreover, note that the right-hand side contains a conjunction of theorems of GLS. Then, since GLS contains modus ponens, $\mathrm{GLS} \vdash \phi$. ☐

► **Theorem 30.**
Let ϕ be letterless, $*$ arbitrary. Then, $GL \vdash \phi$ if and only if $PA \vdash \phi^*$.

Proof.
The forward direction has already been established. Consider, then, the reverse direction. If $[\![\phi]\!]$ is finite, then by lemma 8, $GL \vdash \phi \leftrightarrow \bigvee\{\neg(\Box^{n+1}\bot \rightarrow \Box^{n}\bot) : n \in [\![\phi]\!]\}$, and, since $(\Box^{i+1}\bot)^*$ is false in PA, ϕ^* must be false–and thus unprovable in PA (assuming consistency). If $PA \vdash \phi^*$, it must be, then, that $[\![B]\!]$ is not finite–and thus, co-finite. Then, by lemma 9, $GL \vdash \phi \leftrightarrow \bigwedge\{\Box^{n+1}\bot \rightarrow \Box^{n}\bot : n \notin [\![\phi]\!]\}$. If $[\![\phi]\!] = \mathbb{N}$, then by lemma 6, $GL \vdash \phi$. Assume, then, that $[\![\phi]\!] = \mathbb{N}$. Then, there is an i such that $\Box^{i+1}\bot \rightarrow \Box^{i}\bot$ is a conjunct in the right side of the biconditional above. Assume that $PA \vdash \Box^{i+1}\bot \rightarrow \Box^{i}\bot$. Then, by Löb's theorem, $PA \vdash \Box^{i}\bot$– impossible given our consistency assumptions. It follows, then, that $PA \not\vdash \phi^*$–counter our assumption. It follows that the only relevant case has been established, and thus the theorem follows. □

What's the point of all the iterated consistency stuff? 97-98

7.3 Reflection and Iterated Consistency

> **▶ Theorem 31.**
> *Let $\mathcal{M}$ be a transitive model. Suppose that for some natural number n, $w_nRw_{n-1}R\ldots Rw_1Rw_0$ and $X = \{\Box\phi_i \rightarrow \phi_i : i < n\}$. Then, for some $j \leq n$, $w_j \models \bigwedge X$.*

Proof.
Assume not. Then, for all $j \leq n$ there is an $i < n$ such that $w_j \not\models \Box\phi_i \rightarrow \phi_i$. By the pigeonhole principle, for one of these sentences $\Box\phi_i \rightarrow \phi_i$ there is j and j' such that $w_j \not\models \Box\phi_i \rightarrow \phi_i$ and $w_{j'} \not\models \Box\phi_i \rightarrow \phi_i$. Either $j' < j$ or vice versa. Without loss of generality, assume the former. $w_j \not\models \Box\phi_i \rightarrow \phi_i$ implies $w_j \not\models \neg\Box\phi_i \vee \phi_i$, which is to say $w_j \models \neg(\neg\Box\phi_i \vee \phi_i)$. Then, $w_j \models \Box\phi_i \wedge \neg\phi_i$. A similar argument establishes that $w_{j'} \models \Box\phi_i \wedge \neg\phi_i$ as well. But, by transitivity, $w_{j'}$ accesses w_j, and so $w_{j'} \not\models \Box\phi_i$! Thus, the theorem follows by reductio. □

Restrictions on A_i and p here?

> **▶ Theorem 32** (Leivant)**.**
> *Let $X = \{\Box\phi_i \rightarrow \phi_i : i < n\}$. Suppose $GL \vdash \bigwedge X \rightarrow (\Box^k\psi \rightarrow \psi)$. Then, $k \leq n$.*

Proof.
□

Why do we care?

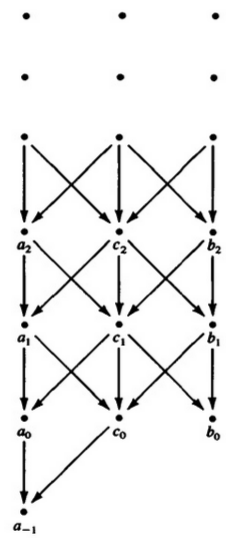
7.4 Nice Normal Forms are Unusual

Much of the work in this chapter was possible because of the normal form theorems established for letterless sentences; we now show that the letterless sentences are unusual in this regard. To begin, let H_0 denote the set of all formulas containing no sentence letter other than p and GL-equivalent to $p, \neg p, \bot,$ or $\top$. Define H_{n+1} to be the set of all truth-functional combinations of sentences with the form $\Diamond^r\phi$ where $r \geq 0$ and $\phi \in H_n$. It's easy to see than every sentence which doesn't contain a sentence letter beside p eventually appears in some H_m, all the letterless sentences appear in H_1, and that $H_n \subseteq H_{n+1}$.

> **▶ Theorem 33** (Solovay)**.**
> *For every n, $H_n \neq H_{n+1}$*

Proof Sketch.
Let ϕ_1 be $\Diamond p$. Let ϕ_{n+1} be defined by $\Diamond(p \wedge \phi_n)$. Note that $\phi_1 \in H_1$. Assume that $\phi_n \in H_n$. Then, by definition, $\phi_{n+1} \in H_{n+1}$. To establish the theorem, all we now need to show is that $\phi_{n+1} \notin H_n$.

Consider the model $\mathcal{M}$ depicted below. Note that the arrows which must exist by transitivity have been left off for the sake of readability. Define the valuation of $\mathcal{M}$ to make p true at all and only the a's and b's.



Show by induction on i that for all $\psi \in H_i$, $a_i \models \psi$ if and only if $b_i \models \psi$. Note finally that $a_i \models \phi_{i+1}$ (follow the left side), but b_i does not. It follows immediately that $\phi_{i+1} \notin H_i$.

7.5 Incompactness

Compactness is a versatile and powerful property. Unfortunately, we now show that GL doesn't have it:

Standard Translation...what happens?

> ▶ **Theorem 34** (Fine-Rautenberg).
> *GL does not have the compactness property.*

Proof Sketch.
Let $p_0, p_1, p_2, \ldots$ be an infinite sequence of distinct sentence letters. Let $T = \{\Diamond p_0\} \cup \{\Box(p_i \to \Diamond p_{i+1}) : i \in \mathbb{N}\}$. Show that any finite subset with p_n its greatest variable is satisfied by the model below (filling in transitivity arrows as appropriate) under the valuation which makes p_i true at only w_i.

$$w \to w_0 \to w_1 \to \cdots \to w_n$$

They do validity at a world. What happens with frame validity?

We now prove that no world in a converse well-founded and irreflexive frame can model T. Assume for reductio that $\mathcal{F}$ is a frame with a world, w, which models T. Define $X = \{x : wRx$ and, for some i, $x \models p_i\}$. Since $w \models \Diamond p_0$, X must be non-empty. Note, however, that since $w \models \Box(p_i \to \Diamond p_{i+1})$, we may generate an infinite walk of worlds which belong to X–each p_i world must connect to the p_{i+1}!

Goldfarb has shown that the infinite number of sentence letters used in Fine and Rautenberg's proof isn't necessary to the establishment of the incompactness of GL; see pg. 102-103 for more details.

Chapter 8

The Fixed Point Theorem

This is apparently really cool.

Definition: $\boxdot$
For any modal sentence ϕ, $\boxdot\phi$ is the sentence $(\Box\phi \wedge \phi)$.

Definition: Modalized
A modal sentence is *modalized in a sentence letter* p if and only if all occurrences of p occur within the scope of a $\Box$ operator.

► **Theorem** (Fixed Point Theorem).
For every sentence ϕ modalized in p, there is a sentence ψ containing only sentence letters in ϕ and not containing p such that $GL \vdash \boxdot(p \leftrightarrow \phi) \leftrightarrow \boxdot(p \leftrightarrow \psi)$

Definition: Fixed Point

Bibliography

[1] Boolos, George. *The Logic of Provability*