MMathPhil Mathematics and Philosophy

Part C: CCD Dissertation

# Axiomatising Provability

Candidate Number: 189111

*Hilary Term, 2016*

# Contents

# 1 Introduction

This dissertation will deal with the question of which principles of modal propositional logic hold when the symbol $\Box$ is interpreted as provability in Peano Arithemic (**PA**, hereafter).

For any formula $\phi$ of the language of **PA** we will see that a formula $Pr(\overline{\ulcorner \phi \urcorner})$ may be given that expresses "$\phi$ is provable in **PA**". Prima facie, $Pr(\overline{\ulcorner \urcorner})$ is similar in syntax to that of the box symbol, $\Box$, used in modal logic. Just as in philosophy where $\Box$ is interpreted as predicates such as "it is known that..." and "it used to be that...", likewise we might interpret $\Box$ as a "it is provable that..." predicate in a formal system of mathematics. Likewise as in the philosophical cases, we will ask what sentences of modal logic are derivable when given the provability interpretation. For example, in the case where $\Box$ is given the interpretation "it is known that...": if one holds that if $S$ is known, then $S$ is true, $\Box S \rightarrow S$ would generally be considered a correct inference.

Roughly, a translation from modal logic to **PA** is a function that assigns formulae of **PA** to each sentence of modal logic that respects the formation rules of sentences and translates $\Box$ as $Pr(\overline{\ulcorner \urcorner})$. We will be considering those modal sentences that are always provable in **PA** under any such translation. The primary objective of this dissertation is to give an exposition of the proof that the class of always provable modal sentences is finitely axiomatisable, and thus show exactly what **PA** can propositionally prove about its own pravability predicate.

Martin Löb showed in [4] that the following three conditions are true of the provability predicate in **PA**, For any formulae $\phi$ and $\psi$:

1. If $\vdash \phi$, then $\vdash Pr(\overline{\ulcorner \phi \urcorner})$
2. $\vdash Pr(\overline{\ulcorner \phi \rightarrow \psi \urcorner}) \rightarrow (Pr(\overline{\ulcorner \phi \urcorner}) \rightarrow Pr(\overline{\ulcorner \psi \urcorner}))$
3. $\vdash Pr(\overline{\ulcorner \phi \urcorner}) \rightarrow Pr(\overline{\ulcorner Pr(\overline{\ulcorner \phi \urcorner}) \urcorner})$

These were derived from conditions formulated by Hilbert and Bernays, we will refer to them as the (Hilbert-Bernays-Löb) derivability conditions. Their modal counterparts can be immediately recognised: what is commonly called the necessitation rule[1], the K-axiom schema, and the 4-axiom schema, respectively. We may deduce from this that at the very least provability in **PA** is sound with respect to the modal system **K4**, whose modal axioms are the K-axiom and 4-axiom schemas[2].

In addition to the above conditions, Löb proved a result about **PA**'s provability predicate that proved to be one of remarkable interest - Löb's Theorem:

If $\mathbf{PA} \vdash Pr(\overline{\ulcorner \phi \urcorner}) \rightarrow \phi$, then $\mathbf{PA} \vdash \phi$

The interest of this result is wide. Firstly, from it one can derive Gödel's Second Incompleteness Theorem for **PA**, which states that if **PA** is consistent it cannot prove its own consistency, in a neat one line proof:
If $\mathbf{PA} \nvdash \bot$, then by Löb's Theorem $\mathbf{PA} \nvdash Pr(\overline{\ulcorner \bot \urcorner}) \rightarrow \bot$, i.e. $\mathbf{PA} \nvdash \neg Pr(\overline{\ulcorner \bot \urcorner})$.

It also provides the answer to a question posed by Henkin[3]: can the formulae $\phi$ of **PA** for which $\mathbf{PA} \vdash \phi \leftrightarrow Pr(\overline{\ulcorner \phi \urcorner})$ themselves be proven in **PA**? We know by the Diagonal Lemma that any predicate $P(x)$ expressible in **PA** has these fixed points relative to a system of Gödel numbering. Löb's [4] was a response to Henkin's question, and Löb's Theorem clearly gives an affirmative answer.

Löb's Theorem has great interest for provability logic. Adding to the modal translations of the derivability conditions the modal translations of the arithmetisation of Löb's Theorem, $\Box(\Box S \rightarrow S) \rightarrow \Box S$, gives the modal system that is the provability logic of **PA**. This modal system is known as **GL** (after Gödel and Löb). The main result of this paper, Solovay's Arithmetical Completeness Theorem[4], *proves* that **GL** is the provability logic of **PA**.

[1] We will refer to this as the "box rule" to avoid any modal connotations.
[2] $\Box(S \rightarrow T) \rightarrow (\Box S \rightarrow \Box T)$ and $\Box S \rightarrow \Box\Box S$
[3] [2]
[4] First proved by Robert Solovay in [6].

In section 2, our focus will be on giving the syntax of our modal logic, and proving important results about **GL**: its soundness and completeness with respect to a certain class of Kripke frames (all those that are transitive and converse wellfounded), and its decidability. Section 3 will develop the language and deductive system of Peano Arithmetic, and give an outline of how we may construct the provability predicate, and prove some results on **PA** that will be needed in the proof of Solovay's Theorem. Section 4 and 5 will be utilise the results provided in the previous sections to prove the following:

For a modal sentence $S$, for every translation $*$, $\mathbf{PA} \vdash S^*$ if and only if $\mathbf{GL} \vdash S$.

The forward direction of which is Solovay's Theorem. For both completeness results, constructive examples are provided to shed light on the proofs. Furthermore, in section 5 we examine the mechanics of Solovay's Theorem to recreate a result of De Jongh, Jumelet, and Montagna[5] concerning an upper bound minimal set of properties required to prove Solovay's Theorem. This paper states that Solovay's Theorem is provable for all extensions of $I\Delta_0$ $(=I\Sigma_0)$ that have provable $\Sigma_1$-completeness, for example all extensions of $I\Delta_0 + EXP$.

I owe the majority of the material in this project to George Boolos's Logic of Provability, [1]. I would like to thank Alex Wilkie for his helpful comments.

# 2 The Modal Deductive System GL

In this section we develop the modal deductive system **GL** and the corresponding modal semantics.

## 2.1 Defining GL

**Definition 2.1.** *Our language $\mathcal{L}$ contains the following symbols:*
*Propositional variables $p_i$ for all $i \in \mathbb{N}$, $\rightarrow$, $\bot$, $\Box$, ( and ).*

[5] [3]

*S is an $\mathcal{L}$-sentence (or modal sentence) if:*

1. *$S$ is a propositional variable or $S = \bot$*

2. *$S = (T \to U)$ or $S = \Box T$ for $\mathcal{L}$-sentences $T$ and $U$.*

Throughout, we will use uppercase Roman letters as metavariables for modal sentences. We will use the following short-hands, writing $\neg S$ for $(S \to \bot)$, $(S \wedge T)$ for $((S \to (T \to \bot)) \to \bot)$, and $S \leftrightarrow T$ for $(S \to T) \wedge (T \to S)$. In many cases we will use these short-hands in proofs, with the understanding that a full proof can be written.

**Definition 2.2.** *The deductive system $\mathcal{L}_0$ has the following axioms and rules of deduction: For any $S$, $T$, $U$ $\mathcal{L}$-sentences:*

*A1.* $S \to (T \to S)$

*A2.* $(S \to (T \to U)) \to ((S \to T) \to (S \to U))$

*A3.* $((S \to \bot) \to (T \to \bot)) \to (T \to S)$

*There is one rule of deduction in $\mathcal{L}_0$, modus ponens:*

*MP: If $\vdash S \to T$ and $\vdash S$, then $\vdash T$.*

Rule A3 written in our shorthand is $(\neg S \to \neg T) \to (T \to S)$. For any formal deductive system $\mathbf{Q}$ we write $\mathbf{Q} \vdash S$ if $S$ is provable in that axiomatic deduction system.

**Definition 2.3.** *For a deductive system $\boldsymbol{Q}$ and a sentence $S$ of the language $\mathcal{L}$, $\boldsymbol{Q} \vdash S$ if there exists a finite sequence of $\mathcal{L}$-sentences $S_1, \ldots, S_n = S$ such that for each $S_i$, either $S_i$ is an axiom, or follows from $S_{j_1}, \ldots, S_{j_m}$ where each $j_k < i$ by a deduction rule of $\boldsymbol{Q}$.*

**Definition 2.4.** *A **truth-functional theorem** of a deductive system extending $\mathcal{L}_0$, $\boldsymbol{Q}$, is a theorem provable using only axioms A1-A3 and MP.*

In general, we will omit the proofs of truth-functional theorems.

We can now give the definition of two modal systems:

**Definition 2.5.** *The modal system* ***K*** *is that obtains by adding the following axiom schema and rule to those of* $\mathcal{L}_0$*:*

$\Box(S \to T) \to (\Box S \to \Box T)$ *(**K Schema**)*

*If* $\vdash S$*, then* $\vdash \Box S$ *(**Box Rule**)*

**Definition 2.6.** *The modal system* ***GL*** *is that obtains by adding the following axiom schema to* ***K****:*

$\Box(\Box S \to S) \to \Box S$ *(**GL Schema**)*

## 2.2 Modal Semantics

We will in this subsection provide a corresponding semantic theory for **GL**.

**Definition 2.7.** *A* ***frame*** *is an ordered pair* $\langle W, R\rangle$ *where* $W$ *is a non-empty set called the domain and* $R$ *is a binary relation on* $W$ *called the accessibility relation. Members of* $W$ *are usually known as worlds.*

Frames are just directed graphs, and are hence easy to draw and visualise with worlds as nodes, and the relation of elements as arrows, for example:

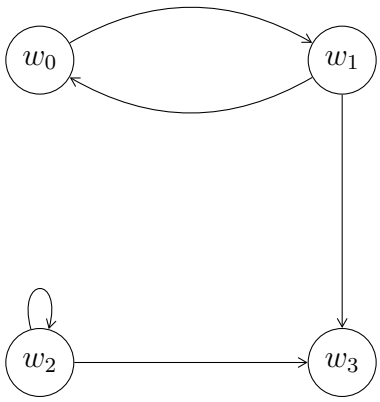

Figure 1

**Definition 2.8.** *A **model** is an ordered triple $\langle W, R, I\rangle$ where $W$ and $R$ are as in Definition 2.7, and $I$ is a function from propositional-variable-world pairs to the set $\{1, 0\}$, known as the interpretation function.*

For a frame $\mathcal{M} = \langle W, R\rangle$, we will write $\mathcal{M}_I$ for the model $\langle W, R, I\rangle$. The interpretation function induces a valuation on all the sentences of modal logic for a given frame $\mathcal{M}$:

**Definition 2.9.** *The **valuation** $V_{\mathcal{M},I}$ for a frame $\mathcal{M} = \langle W, R\rangle$ and interpretation $I$ is a function from sentence-world pairs to the set $\{1, 0\}$ defined inductively as follows:*

1. *For all $i \in \mathbb{N}$, all $w \in W$, $V_{\mathcal{M},I}(p_i, w) = I(p_i, w)$.*
2. *For all $w \in W$, $V_{\mathcal{M},I}(\bot, w) = 0$*
3. *$V_{\mathcal{M},I}(S \to T, w) = 1$ if and only if $V_{\mathcal{M},I}(S, w) = 0$ or $V_{\mathcal{M},I}(T, w) = 1$*
4. *$V_{\mathcal{M},I}(\Box S, w) = 1$ if and only if for all $v \in W$ such that $Rwv$, $V_{\mathcal{M},I}(S, v) = 1$ .*

**Definition 2.10.** *If $\mathcal{M}_I = \langle W, R, I\rangle$, we say that a modal sentence $S$ **is true at** $w \in W$ if $V_{\mathcal{M},I}(S, w) = 1$.*

**Definition 2.11.** *a modal sentence $S$ is **valid** in a frame $\mathcal{M} = \langle W, R\rangle$ if for all interpretation functions $I$, for all $w \in W$, $S$ is true at $w$.*
*We write this as $\mathcal{M} \models S$.*

All and only the theorems of K are valid in all frames. By putting restrictions on the accessibility relation $R$, we can limit the set of sentences that are valid in a frame, and thus produce the various corresponding semantic theories to each modal proof system. The frames that correspond to the class of theorems of a modal system are the frames in which all and only those theorems are valid. For example, a modal sentence $P$ is a theorem of

the **T** system of modal logic[6] if and only if $P$ is valid in all frames $\langle W, R\rangle$ in which $R$ is reflexive on $W$. In this case we say that the frame is reflexive. The following definitions are of particular interest to our discussion:

**Definition 2.12.** *A binary relation $R$ on a set $W$ is **transitive** if for all $u$, $v$, $w \in W$, if $Ruv$ and $Rvw$ then $Ruw$.*

**Definition 2.13.** *A binary relation $R$ on a set $W$ is **converse wellfounded** if for every non-empty set $X \subseteq W$ there is an element $w \in X$ such that for no $v \in X$ does $Rwv$ hold.*

**Definition 2.14.** *A binary relation $R$ is **irreflexive** on a set $W$ if for no $w \in W$ does $Rww$ hold.*

Note the following useful correspondence on finite frames:

**Lemma 2.15.** *A finite, transitive frame $\mathcal{M} = \langle W, R\rangle$ is irreflexive if and only if it is converse wellfounded.*

*Proof.* Let $\mathcal{M} = \langle W, R\rangle$ be finite and transitive.

Assume that $R$ is not converse wellfounded. By definition, for some non-empty $X \subseteq W$, for each $x \in X$ there exists some $y \in X$ such that $Rxy$. In other words, there is an infinite sequence $x_1, x_2, \ldots$ of elements of $X$ such that $Rx_ix_{i+1}$ for all $i \in \mathbb{N}$. $X$ is finite, and thus there are $i, j \in \mathbb{N}$ with $i < j$ such that $x_i = x_j$. Thus by transitivity of $R$, $Rx_ix_i$, which contradicts irreflexivity of $R$. Hence if $\mathcal{M}$ is finite, transitive and irreflexive, then it is converse wellfounded.

Now assume that $\mathcal{M}$ is not irreflexive, so there exists some $w \in W$ such that $Rww$. But then $\{w\}$ is a non-empty subset of $W$ and for every $v \in \{w\}$, $Rwv$ holds, so $\mathcal{M}$ is not converse wellfounded.

Therefore for finite transitive frames $\mathcal{M}$, $\mathcal{M}$ is irreflexive if and only if it is converse wellfounded. ■

[6] defined by adding the **T**-Schema, $\Box S \to S$, to **K**.

We will make use of this lemma in the proof of the completeness theorem for **GL**. Figure 2 shows an example of a frame that is finite transitive and irreflexive, (and hence converse wellfounded).

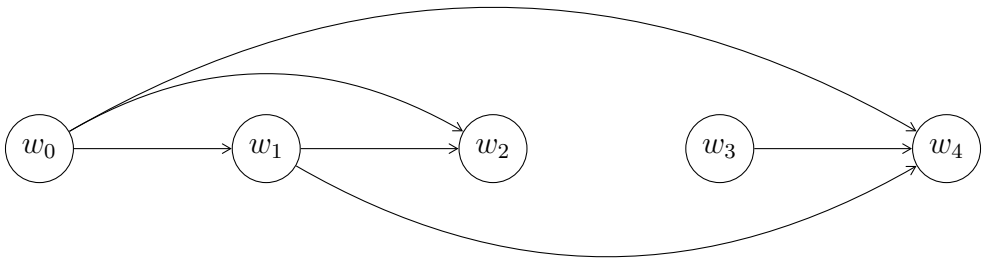


Figure 2

## 2.3 Soundness Theorem for GL

Here we will prove that **GL** is sound with respect to transitive, converse wellfounded frames; if $\mathbf{GL} \vdash S$, then $S$ is valid in all transitive converse wellfounded frames.[7]

**Lemma 2.16.** *For any modal deductive system* $\boldsymbol{Q}$ *extending* $\boldsymbol{K}$, $\boldsymbol{Q} \vdash \Box(A_1 \wedge \cdots \wedge A_n) \leftrightarrow (\Box A_1 \wedge \cdots \wedge \Box A_n)$

*Proof.* By induction on $n$.

Base case:

If $n = 1$, then by propositional logic $\mathbf{Q} \vdash \Box A_1 \leftrightarrow \Box A_1$.

Inductive case:

Assume that the Lemma holds for $1 \leq n$ and let $A = A_1 \wedge \cdots \wedge A_n$ and $B = A_{n+1}$:

$$\text{By prop. logic, } \vdash (A \wedge B) \rightarrow A \tag{1}$$

$$\text{and } \vdash (A \wedge B) \rightarrow B \tag{2}$$

[7]The material in this subsection is adapted from [1], Chapters 1 and 4.

$$\text{By the Box Rule and K-Schema, } \vdash \Box(A \wedge B) \to \Box A \tag{3}$$

$$\text{and } \vdash \Box(A \wedge B) \to \Box B \tag{4}$$

$$\text{Hence, } \vdash \Box(A \wedge B) \to (\Box A \wedge \Box B) \tag{5}$$

$$\text{Also by prop. logic, } \vdash A \to (B \to (A \wedge B)) \tag{6}$$

$$\text{Box Rule, K-Schema, prop. logic, } \vdash \Box A \to (\Box B \to \Box(A \wedge B)) \tag{7}$$

$$\text{Which is equivalent to } \vdash (\Box A \wedge \Box B) \to \Box(A \wedge B) \tag{8}$$

$$\text{By (5) and (8), } \vdash \Box(A \wedge B) \leftrightarrow (\Box A \wedge \Box B) \tag{9}$$

$$\text{By the inductive hypothesis, } \vdash \Box A \leftrightarrow (\Box A_1 \wedge \cdots \wedge \Box A_n) \tag{10}$$

$$\text{Hence by substitution, } \vdash \Box(A \wedge B) \leftrightarrow (\Box A_1 \wedge \cdots \wedge \Box A_n \wedge \Box A_{n+1}) \tag{11}$$

■

**Lemma 2.17.** *For any modal sentence A,* $\boldsymbol{GL} \vdash \Box A \to \Box\Box A$.

*Proof.* Letting $B = \Box A \wedge A$ where appropriate for readability:

$$\text{By prop. logic, } \vdash A \to ((\Box\Box A \wedge \Box A) \to (\Box A \wedge A)) \tag{1}$$

$$\text{By Lemma 2.16 } \vdash \Box(\Box A \wedge A) \leftrightarrow (\Box\Box A \wedge \Box A) \tag{2}$$

$$\text{From (1) and (2) by prop. logic, } \vdash A \to (\Box(\Box A \wedge A) \to (\Box A \wedge A)) \tag{3}$$

$$\text{By the Box rule and K-schema, } \vdash \Box A \to \Box(\Box B \to B) \tag{4}$$

$$\text{By the } \mathbf{GL}\text{-schema, } \vdash \Box(\Box B \to B) \to \Box B \tag{5}$$

$$\text{Hence from (4) and (5) } \vdash \Box A \to \Box B \tag{6}$$

$$\text{From (2), } \vdash \Box B \to (\Box\Box A \wedge \Box A) \tag{7}$$

$$\text{From (6) and (7) by prop. logic, } \vdash \Box A \to \Box\Box A \tag{8}$$

■

**Lemma 2.18.** *$\mathcal{M} \models \Box A \to \Box\Box A$ if and only if $\mathcal{M}$ is transitive.*

*Proof.* Suppose $\mathcal{M} \models \Box p_1 \to \Box\Box p_1$, and for some $u$, $v$, $w \in W$, $Ruv$ and $Rvw$ hold. Define interpretation $I$ as follows: $I(p_1, x) = 1$ if and only if $Rux$. Then $V_{\mathcal{M},I}(\Box p_1, u) = 1$, and as $\Box p_1 \to \Box\Box p_1$ is valid in $\mathcal{M}$, $V_{\mathcal{M},I}(\Box\Box p_1, u) = 1$. Hence, as $Ruv$, $V_{\mathcal{M},I}(\Box p_1, v) = 1$, and further as $Rvw$ $V_{\mathcal{M},I}(p_1, w) = 1$. By definition of $I$, we have that $Ruw$; in other words, $R$ is transitive.

Now, suppose $R$ is transitive, and let $I$ be any interpretation. If $V_{\mathcal{M},I}(\Box p_1, u) = 1$ and $Ruv$, then for any $w \in W$ such that $Rvw$, $Ruw$ holds by transitivity, and thus $V_{\mathcal{M},I}(p_1, w) = 1$. Hence for any $v$ such that $Ruv$, $V_{\mathcal{M},I}(\Box p_1, v) = 1$ and so $V_{\mathcal{M},I}(\Box\Box p_1, u) = 1$. So $V_{\mathcal{M},I}(\Box p_1 \to \Box\Box p_1, u) = 1$, for any $I$ and any $u \in W$ and hence $\mathcal{M} \models \Box p_1 \to \Box\Box p_1$.

This result holds for any modal sentence $A$ in place of $p_1$. ■

**Lemma 2.19.** *Let $\boldsymbol{Q}$ be a modal deductive system extending $\mathcal{L}_0$ with only deductive rules MP and the Box Rule, and let $\mathcal{M}$ be a frame in which all the axioms of $\boldsymbol{Q}$ are valid. Then, all theorems of $\boldsymbol{Q}$ are valid in $\mathcal{M}$.*

*Proof.* By induction on the length of proofs. Let $S_1, \ldots, S_n = S$ be a proof of $S$ in $\mathbf{Q}$, and let $\mathcal{M} = \langle W, R\rangle$ be a frame in which all the axioms of $\mathbf{Q}$ are valid.

Base Case:

Suppose $n = 1$. Then $S_1 = S$ is an axiom of $\mathbf{Q}$, and hence is valid in $\mathcal{M}$ by assumption.

Inductive cases:

Suppose the hypothesis holds for all proofs of length $\leq n$, and consider a proof of $S$ of length $n + 1$.

Case 1: $S$ follows from $S_i$ and $S_j = S_i \to S$ for $i, j < n$, by MP. By the inductive hypothesis, for all interpretations $I$ over $\mathcal{M}$ and all $w \in W$, $V_{\mathcal{M},I}(S_i, w) = 1$ and $V_{\mathcal{M},I}(S_i \to S, w) = 1$, thus by our modal semantics, $V_{\mathcal{M},I}(S, w) = 1$.

Case 2: $S$ follows from $S_i$ by the Box Rule, so $S = \Box S_i$. By the inductive hypothesis, for all interpretations $I$ over $\mathcal{M}$, and for any $w \in W$, $V_{\mathcal{M},I}(S_i, w) = 1$. Then, a fortiori, for all $v \in W$ such that $Rwv$, $V_{\mathcal{M},I}(S_i, v) = 1$, so $V_{\mathcal{M},I}(\Box S_i, w) = 1$.

Hence by induction, any theorem of **Q** is valid in $\mathcal{M}$. ■

**Lemma 2.20.** *$\mathcal{M} \vDash \Box(\Box A \to A) \to \Box A$ if and only if $\mathcal{M}$ is transitive and converse wellfounded.*

*Proof.* Suppose $P = \Box(\Box p_1 \to p_1) \to \Box p_1$ is valid in some frame $\mathcal{M} = \langle W, R\rangle$. Then any instance of the GL-schema is also valid in $\mathcal{M}$, and hence by Lemma 2.19 and the fact that any instance of the K-schema is valid in any frame, all theorems of **GL** are valid in $\mathcal{M}$.

Transitivity:

By lemmas 2.17 and 2.18 above, $\mathcal{M} \models \Box p_1 \to \Box\Box p_1$, so $\mathcal{M}$ is transitive.

Converse wellfounded:

Suppose $R$ is no converse wellfounded, so there is non-empty $X \subseteq W$ such that for all $x \in X$ there is $y \in X$ such that $Rxy$. Let $w \in X$ and let $I$ be an interpretation function defined by $I(p_1, x) = 1$ if and only if $x \notin X$.

By definition of $X$, there is $u \in X$ such that $Rwu$. Then by definition of $I$, for any $u \in X$ such that $Rwu$, $V_{\mathcal{M},I}(p_1, u) = 0$ and hence $V_{\mathcal{M},I}(\Box p_1, w) = 0$ by the semantics of $\Box$. Similarly, for any such $u$, $V_{\mathcal{M},I}(\Box p_1, u) = 0$, and hence $V_{\mathcal{M},I}(\Box p_1 \to p_1, u) = 1$. Further, for any $v \notin X$ such that $Rwv$ holds, by definition of $I$, $V_{\mathcal{M},I}(p_1, v) = 1$ and hence $V_{\mathcal{M},I}(\Box p_1 \to p_1, v) = 1$ in all those worlds too. Thus it follows that $V_{\mathcal{M},I}(\Box(\Box p_1 \to p_1), w) = 1$ and because $V_{\mathcal{M},I}(\Box p_1, w) = 0$, $V_{\mathcal{M},I}(\Box(\Box p_1 \to p_1) \to \Box p_1, w) = 0$.

For the converse, let $\mathcal{M} = \langle W, R\rangle$ be a transitive converse wellfounded frame and assume that for some interpretation function $I$, and for some $w \in W$, $V_{\mathcal{M},I}(\Box(\Box p_1 \to p_1) \to \Box p_1, w) = 0$. Then:

(i) $V_{\mathcal{M},I}(\Box(\Box p_1 \to p_1), w) = 1$

(ii) $V_{\mathcal{M},I}(\Box p_1, w) = 0$

Define $Y := \{x \in W : Rwx \text{ and } V_{\mathcal{M},I}(p_1, x) = 0\}$. $Y$ is not empty as by (ii) there is some $u \in W$ such that $Rwu$, and $V_{\mathcal{M},I}(p_1, u) = 0$. As $\mathcal{M}$ is converse wellfounded, there is some element of $v \in Y$, such that $Rvu$ for no $u \in X$. By definition of $Y$, $Rwv$ and $V_{M,I}(p_1, v) = 0$ and all $u$ such that $Rvu$ are in $W \setminus Y$. If for some $u$ such that $Rvu$, by transitivity $Rwu$. We must have $V_{\mathcal{M},I}(p_1, u) = 1$ as $u \notin Y$. Thus $V_{\mathcal{M},I}(\Box p_1, v) = 1$. But as $v \in Y$, it follows that $V_{\mathcal{M},I}(\Box p_1 \to p_1, v) = 0$, which contradicts (i). Therefore it must be the case that $V_{\mathcal{M},I}(\Box(\Box p_1 \to p_1) \to \Box p_1, w) = 1$, and hence that $\mathcal{M} \models \Box(\Box p_1 \to p_1) \to \Box p_1$. ■

**Theorem 2.21** (Soundness Theorem for **GL**). *If* $\boldsymbol{GL} \vdash S$ *then* $S$ *is valid in all transitive converse wellfounded frames.*

*Proof.* We know that the K-schema, and all truth-functional theorems are valid in all frames, so a fortiori they are valid in all transitive converse wellfounded frames. So, by Lemma 2.20, all axioms of **GL** are valid in all transitive converse wellfounded frames, and hence by Lemma 2.19, all theorems of **GL** are too. ■

## 2.4 Completeness Theorem for GL

In this subsection we prove the Completeness Theorem[8] for **GL**. We will show that **GL** is complete with respect to converse wellfounded frames; if for all transitive converse wellfounded frames $\mathcal{M}$, $\mathcal{M} \models S$, then $\mathbf{GL} \vdash S$.

Throughout this subsection, we will take $D$ to be a modal sentence such that $\mathbf{GL} \nvdash D$ and show that we can construct a transitive converse wellfounded model $\mathcal{M}$ such that $\mathcal{M} \not\models D$. We will require the following terminology:

**Definition 2.22.** *Let* $T$ *be a modal sentence.* $S$ *is a* $T$***-formula*** *if it is either a subsentence of* $T$*, or the negation of a subsentence of* $T$*.*

[8]The proof of Completeness is due to [1], Chapter 5.

**Definition 2.23.** *For a set of modal sentences $X$, $X$ is* ***GL-consistent*** *if* $\mathbf{GL} \nvdash \neg \bigwedge X$.

$\bigwedge X$ is the sentence that is the conjunction of all the elements of $X$. As we are working in **GL** throughout we shall say consistency instead of **GL**-consistency.

**Definition 2.24.** *For a modal sentence $T$, a set of $T$-formulae, $X$, is* ***$T$-maximal-consistent*** *if $X$ is consistent and for each subsentence $S$ of $T$ either $S \in X$ or $\neg S \in X$.*

Let us prove two key facts about $T$-maximal-consistent sets:

**Lemma 2.25.** *Let $T$ be a modal sentence. Let $X$ be a $T$-maximal-consistent set, and $S_1, \ldots, S_n \in X$. Then:*

1. *$S_i \in X$ if and only if $\neg S_i \notin X$*
2. *For a subsentence $S_{n+1}$ of $T$, if $\mathbf{GL} \vdash \bigwedge_{i \leq n} \{S_i\} \rightarrow S_{n+1}$, then $S_{n+1} \in X$*

*Proof.* For (1), if $S_i \in X$ and $\neg S_i \in X$, then as $\mathbf{GL} \nvdash \neg(S_i \wedge \neg S_i)$, it follows that $\mathbf{GL} \nvdash \bigwedge X$ which contradicts the consistency of $X$: so $S_i \in X$ if and only if $\neg S_i \notin X$. For (2), if $S_{n+1} \notin X$ then by $T$-maximal-consistency $\neg S_{n+1} \in X$. But $\mathbf{GL} \vdash \bigwedge_{i \leq n} \{S_i\} \rightarrow S_{n+1}$ implies that $\mathbf{GL} \vdash \neg(\bigwedge_{i \leq n} \{S_i\} \wedge \neg S_{n+1})$ which further implies $\mathbf{GL} \vdash \neg \bigwedge X$ which contradicts consistency; so $S_{n+1} \in X$. ■

**Lemma 2.26.** *Let $T$ be a modal sentence. If $X$ is a consistent set of $T$-formulae, then it is contained in some $T$-maximal-consistent set.*

*Proof.* Let $X$ be a consistent set of $T$-formulae, so $\mathbf{GL} \nvdash \neg \bigwedge X$. Let $\mathbb{T}$ be the set of $T$-formulae and let $\{Y_i : i \in I\}$ (for some indexing set $I$) be the set of maximal sets of $T$-formulae: $\{Z \subset \mathbb{T} :$ for any T-formula S, $S \in Z \Leftrightarrow \neg S \notin Z\}$ By propositional logic, $\bigwedge X$ is equivalent to $\bigvee_{i \in I} \{\bigwedge X \wedge \bigwedge Y_i\}$.

At least one of the disjuncts $(\bigwedge X \wedge \bigwedge Y_k)$ is consistent, otherwise $\mathbf{GL} \vdash \neg \bigvee_{i \in I} (\bigwedge X \wedge \bigwedge Y_i)$ and hence $\mathbf{GL} \vdash \neg \bigwedge X$. Then the set $X \cup Y_k$ is $T$-maximal-consistent by construction. ■

For the modal sentence $D$, since $D$ is a subsentence of itself, and $\mathbf{GL} \nvdash D = \neg\neg D$, then $\{\neg D\}$ is a consistent set. By Lemma 2.26, this is contained in a $D$-maximal-consistent set and hence $\mathbb{D}$, the set of $D$-maximal-consistent sets, is non-empty.

Recall that we are trying to find a frame $\mathcal{M}$ that is transitive converese wellfounded and such that there is some interpretation and world such that $D$ is not true at that world. Let us take $\mathbb{D}$ to be the set of worlds, and define the interpretation $I$ as follows:

- For all $i \in \mathbb{N}$, all $w \in \mathbb{D}$ $I(p_i, w) = 1 \Leftrightarrow p_i \in w$.

In addition to the accessibility relation $R$ be converse wellfounded, we will require that the following is satisfied:

- For every subsentence $\Box S$ of $D$, $\Box S \in w \Leftrightarrow$ for all $v$ such that $Rwv$, $S \in v$.

Prima facie, these two conditions on $I$ and $R$ together ensure that for subsentences of $D$, membership at a world is equivalent to truth at that world, which we prove in the following lemma:

**Lemma 2.27.** *Let $\mathcal{M}_I = \langle \mathbb{D}, R, I \rangle$ be a model such that:*

1. *$I(p_i, w) = 1 \Leftrightarrow p_i \in w$*
2. *For every subsentence $\Box S$ of $D$, $\Box S \in w \Leftrightarrow$ for all $v$ such that $Rwv$, $S \in v$*

*Then, for every subsentence $S$ of $D$ and every $D$-maximal-consistent set $w \in \mathbb{D}$, $S \in w$ if and only if $V_{\mathcal{M},I}(S, w) = 1$.*

*Proof.* Let $S$ be a subsentence of $D$, and let $\mathcal{M}_I$ be a model as described. We will prove the lemma by induction on the length of sentences. We have two base cases to consider:

(i) $S = \bot$

(ii) $S = p_i$ for some $i \in \mathbb{N}$

For (i), $\bot \notin w$ as $w$ is consistent and $\mathbf{GL} \vdash \neg \bot$. Also by definition of interpretations, $V_{\mathcal{M},I}(\bot, w) = 0$. (ii) follows directly from the definition of $I$.

We have two inductive cases to consider:

(iii) $S = T \to U$

(iv) $S = \Box T$

Assume $T$ and $U$ are subsentences of $D$ such that the hypothesis holds. For (iii), assume that $S \notin w$. Since $S = T \to U$, by propositional logic, $\mathbf{GL} \vdash \neg S \to T$, $\mathbf{GL} \vdash \neg S \to \neg U$ and $\mathbf{GL} \vdash (T \wedge \neg U) \to \neg S$. Then by Lemma 2.25, $S \notin w$ if and only if $\neg S \in w$, and further, $\neg S \in w$ if and only if $T \in w$ and $\neg U \in w$. By the inductive hypothesis, this is if and only if $V_{\mathcal{M},I}(T, w) = 1$ and $V_{\mathcal{M},I}(U, w) = 0$, if and only if $V_{\mathcal{M},I}(S, w) = 0$.

For case (iv), assume that $S \in w$. By condition (1) on the accessibility relation, $S = \Box T \in w$ if and only if for all $v$ such that $Rwv$, $T \in v$. By the inductive hypothesis, this is if and only if $V_{\mathcal{M},I}(T, v) = 1$ for all such $v$, which is if and only if $V_{\mathcal{M},I}(S, w) = 1$ by the modal semantics of $\Box$.

Thus by induction we have shown that for any subsentence $S$ of $D$, and every $w \in \mathbb{D}$, $S \in \mathbb{D}$ if and only if $V_{\mathcal{M},I}(S, w) = 1$. ■

**Theorem 2.28** (Completeness Theorem for GL)**.** *If a modal sentence $S$ is valid in all transitive converse wellfounded frames, then $\boldsymbol{GL} \vdash S$.*

*Proof.* Let $D$ be a modal sentence such that $\mathbf{GL} \nvdash D$, and consider the model $\mathcal{M}_I = \langle \mathbb{D}, R, I \rangle$ with $R$ and $I$ defined as follows:

- $Rwv \Leftrightarrow$ for all $\Box T \in w$, both $\Box T$ and $T \in v$ and there exists some $\Box U \in v$ such that $\neg \Box U \in w$
- $I(p_i, w) = 1 \Leftrightarrow p_i \in w$

. Our aim is to show that $R$ is transitive converse wellfounded and that $R$ satisfies the condition of Lemma 2.27:

($*$) for every subsentence $\Box S$ of $D$, and every $w \in \mathbb{D}$, $\Box S \in w$ if and only if for all $v$ such that $Rwv$, $S \in v$.

Transitivity of $R$:

Suppose that $Rwv$ and $Rvu$. We want to show that $Rwu$. If $\Box T \in w$, then $\Box T \in v$ and hence $\Box T$ and $T \in u$. Also for some $\Box U \in v$, $\neg\Box U \in w$. But $\Box U \in u$ too as $Rvu$, so $Rwu$.

Converse wellfoundedness of $R$:

$R$ must be irreflexive, otherwise if $Rww$, then for some $\Box T \in w$, $\neg\Box T \in w$, which contradicts the consistency of $w$. So we know that $\mathcal{M}$ is finite, transitive, and irreflexive. Thus by Lemma 2.15, $R$ is converse wellfounded.

Property ($*$):

Assume that $\Box T \in w$ and that $Rwv$. Then by definition of $R$ we have that $T \in v$. Now assume that $\Box T \notin w$. We want to show that there exists an $x \in W$ such that $Rwx$ and $T \notin x$. Let $X = \{\neg T, \Box T\} \cup \{S, \Box S : \Box S \in w\}$.

If $X$ is inconsistent, then:

$$\vdash \neg(\neg T \wedge \Box T \wedge \bigwedge\{S_i \wedge \Box S_i\}) \tag{1}$$

Hence by prop. logic, $$\vdash \bigwedge\{S_i \wedge \Box S_i\} \to (\Box T \to T) \tag{2}$$

By the Box Rule and K-schema, $$\vdash \Box\bigwedge\{S_i \wedge \Box S_i\} \to \Box(\Box T \to T) \tag{3}$$

By Lemma 2.1, $$\vdash \bigwedge\{\Box S_i \wedge \Box\Box S_i\} \to \Box(\Box T \to T) \tag{4}$$

By the GL-Schema, $$\vdash \Box(\Box T \to T) \to \Box T \tag{5}$$

By Lemma 2.2, for each $S_i$, $$\vdash \Box S_i \to \Box\Box S_i \tag{6}$$

Hence by prop. logic, $$\vdash \bigwedge\{\Box S_i\} \to \Box T \tag{7}$$

As each $\Box S_i \in w$, this implies by lemma 2.25 that $\Box T \in w$, which contradicts our assumption, so we must have that $X$ is consistent. By Lemma 2.26 for some $D$-maximal-consistent set $x$, $X \subseteq x$. Since $\neg T \in X$, $\neg T \in x$ and hence $T \notin x$. To finish, we show that $Rwx$. If $\Box S \in w$, then by construction $\Box S$ and $S \in X \subseteq x$. Further, since by assumption $\Box T \notin w$, by maximality $\neg\Box T \in w$ and by construction $\Box T \in X \subseteq x$. Hence $Rwx$. Thus property $(*)$ is satisfied for $R$.

As $\neg D$ is contained in some $D$-maximal-consistent set, $w_0$, by Lemma 2.27 $V_{\mathcal{M},I}(\neg D, w_0) = 1$, and hence $V_{\mathcal{M},I}(D, w) = 0$; so $\mathcal{M} \not\models D$. Therefore by contraposition, if for all transitive converse wellfounded frames $\mathcal{M}$ and the modal sentence $D$, $\mathcal{M} \models D$, then $\mathbf{GL} \vdash D$. ■

### 2.4.1 Strengthening the Completeness Theorem for GL

It will be beneficial in the proof of Solovay's Theorem to be able to reduce and simplify the structure of these constructed models further, which we characterise using the following theorem:[9]

**Theorem 2.29.** *For a modal sentence $D$, if $\boldsymbol{GL} \nvdash D$, then there is a model $\mathcal{N}_J = \langle X, S, J\rangle$ and $w_0 \in X$ such that the following hold:*

1. $V_{\mathcal{N},J}(D, w_0) = 0$
2. *$S$ is transitive and converse wellfounded*
3. *If $y \in X$ and $y \neq w_0$ then $Sw_0y$*
4. *$X$ is finite*

*Proof.* Let $D$ be a modal sentence such that $\mathbf{GL} \nvdash D$, and let $\mathcal{M}_I = \langle W, R, I\rangle$ be a finite transitive and converse wellfounded (therefore irreflexive) model such that $V_{\mathcal{M},I}(D, w_0) = 0$. We construct a new model $\mathcal{N}_J = \langle X, S, J\rangle$ as follows:

[9] This proof differs to that found in [1], p84.

- $X = \{x \in W : Rw_0x \text{ or } x = w_0\}$
- $S = R|_X$
- $J(p_i, x) = I(p_i, x)$ for each $i \in \mathbb{N}$, $x \in X$

$S$ is clearly irreflexive as $R$ is. Further, if for $x, y, z \in X$, $Sxy$ and $Syz$, then $Rxy$ and $Ryz$, so $Rxz$ and hence $Sxz$. So $S$ is also transitive. Notice that if $x \in X$ and $Rxy$, then $Rw_0x$, by definition of $X$, and hence $y \in X$ by transitivity of $R$. This means that $Sxy$ if and only if $Rxy$, for any $x, y \in W$ (where $Sxy$ is false if $x$ or $y$ are not in $X$).

By induction of the construction of sentences, we will show that for all sentences $A$, $V_{\mathcal{M},I}(A, x) = 1$ if and only if $V_{\mathcal{N},J}(A, x) = 1$. We will write $V$ and $V'$ for $V_{\mathcal{M},I}$ and $V_{\mathcal{N},J}$ respectively.

Base case:
If $A = p_i$ for some $i \in \mathbb{N}$, then by definition of $J$, $V'(p_i, x) = V(p_i, x)$. If $A = \perp$, the result follows by definition of a valuation function.

Inductive case:
If $A = B \to C$, then $V'(A, x) = 1$ if and only if $V'(B, x) = 0$ or $V'(C, x) = 1$, which by the inductive hypothesis is if and only if $V(B, x) = 0$ or $V(C, x) = 1$, if and only if $V(A, x) = 1$. If $A = \Box B$, then $V'(A, x) = 1$ if and only if for all $y \in X$ such that $Sxy$, $V'(B, y) = 1$. This is if and only if for all $y \in W$ such that $Rxy$, $V(B, y) = 1$, if and only if $V(A, x) = 1$.

Hence by induction, for all modal sentences $A$, $V_{\mathcal{M},I}(A, x) = 1$ if and only if $V_{\mathcal{N},J}(A, x) = 1$, and in particular $V_{\mathcal{N},J}(D, w_0) = 0$, so $\mathcal{N}_J$ is in fact a countermodel for $D$, and clearly satisfies the 4 conditions by construction. ■

### 2.4.2 Constructing a Countermodel

We have shown in the above proof of the Completeness Theorem for **GL** that given a modal sentence that is not a theorem of **GL**, one can construct a finite transitive irreflexive model

that does not satisfy that sentence. As this is indeed a constructive procedure we have given, let us look at an example modal sentence that is not provable in **GL** and find the countermodel that the methods used to prove completeness produces.

We will take our modal sentence to be $A = p_0 \to \Box p_0$. We first find the set of maximal $A$-consistent sets, which we will denote $W_A$. Recall that a set of $A$-formulae, $T$ is maximal $A$-consistent if $\mathbf{GL} \nvdash \neg \bigwedge T$, and for each subsentence $S$ of $A$, either $S \in T$ or $\neg S \in T$. The subsentences of $A$ are $p_0$, $\Box p_0$, and $A$. Hence $W_A$ is a subset of:
$\{\{p_0, \Box p_0, A\}, \{\neg p_0, \Box p_0, A\}, \{p_0, \neg\Box p_0, A\}, \{p_0, \Box p_0, \neg A\}, \{\neg p_0, \neg\Box p_0, A\}, \{\neg p_0, \Box p_0, \neg A\},$
$\{p_0, \neg\Box p_0, \neg A\}, \{\neg p_0, \neg\Box p_0, \neg A\}\}$.

As $A$ is true if and only if $p_0$ is false or $\Box p_O$ is true, and **GL** is sound and complete with respect to propositional truth when restricted to propositional formulae, we can immediately see that $W_A = \{\{p_0, \Box p_0, A\}, \{\neg p_0, \Box p_0, A\}, \{\neg p_0, \neg\Box p_0, A\}, \{p_0, \neg\Box p_0, \neg A\}\}$.

We will label the elements of $W_A$ $w_0$, $w_1$, $w_2$, $w_3$, respectively. Next we define $R_A$, a binary relation, in line with how it was defined in the proof of Theorem 2.28; $R_A uv$ if and only if for all $\Box T \in u$, $\Box T, T \in v$, and there exists some $\Box U \in v$ such that $\neg\Box U \in u$. In terms of our particular set $W_A$, this translates as: $R_A uv$ if and only if [if $\Box p_0 \in u$, then $p_0 \in v$] and [$\neg\Box p_0 \in u$ and $\Box p_0 \in v$]. If $\Box p_0 \in u$, then by the second conjunct there is no $v$ such that $R_A uv$, so this condition reduces to: if and only if $\neg\Box p_0 \in u$ and $\Box p_0 \in v$. Hence, we have:
$R_A = \{\langle w_2, w_0\rangle, \langle w_2, w_1\rangle, \langle w_3, w_0\rangle, \langle w_3, w_1\rangle\}$

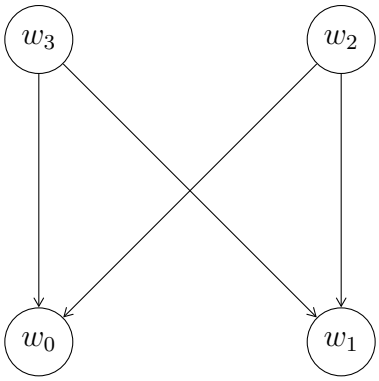


Figure 3

We can very clearly see that this is finite, transitive, and irreflexive, so $\langle W_A, R_A \rangle$ is a frame of the appropriate kind.

Finally we give the interpretation function $I_A$: for all $i \in \mathbb{N}$, for all $j \in \{0, 1, 2, 3\}$, $I_A(p_i, w_j) = 1$ if and only if $p_i$ occurs in $A$ and $p_i \in w_j$. For our frame $\langle W_A, R_A \rangle$, this translates to: $I_A(p_i, w_j) = 1$ if and only if $i = 0$ and $j = 0$ or $j = 3$.

Now $\neg A \in w_3$, so by Lemma 2.27 $V_{\langle W_A, R_A \rangle, I_A}(\neg A, w_3) = 1$, and hence $V_{\langle W_A, R_A \rangle, I_A}(A, w_3) = 0$, so this is indeed a countermodel for $A$.

If we now apply the procedure in Theorem 2.29 to the model $\langle W_A, R_A, I_A \rangle$, we recover a model $\langle X_A, S_A, J_A \rangle$ such that:

- $X_A = \{w_0, w_1, w_3\}$
- $S_A = \{\langle w_3, w_0 \rangle, \langle w_3, w_1 \rangle\}$
- $J_A(p_i, x) = I_A(p_i, x)$ for each $i \in \mathbb{N}$, $x \in X_A$

The valuation function $V$ given by $J_A$ is such that $V(p_0 \to \Box p_0, w_3) = 0$, as $V(p_0, w_3) = 1$ and there exists $y \in X_A$ (namely $w_1$) such that $S_A x_0 y$ and $V(p_0, y) = 0$.

While this construction guarantees an appropriate frame can be produced, in our particular example we can see that the following frame would suffice:


Figure 4

Where $I$ is given such that $I(p_0, w_0) = 1$ and $I(p_0, w_1) = 0$. Thus the construction does not ensure the most "efficient" of frames is procured.

## 2.5 Decidability of GL

The proof of the Completeness Theorem for **GL** gives us a method for deciding whether or not a modal sentence is a theorem of **GL**. Let $D$ be a modal sentence. If $D$ has $k$ subsentences, then there are at most $2^k$ maximal $D$-consistent sets, as no such set can contain both a sentence and its negation (as certainly $\mathbf{GL} \nvdash A \wedge \neg A$, for any sentence $A$). Hence, the proof of the Completeness and Soundness Theorems for **GL** together tell us the following: $\mathbf{GL} \nvdash D$ if and only if there is a model $\langle W, R, I\rangle$ satisfying the following:

1. $|W| \leq 2^k$
2. $R$ transitive and irreflexive (as $W$ is finite)
3. $I(p_i, w) = 1$ if and only if $p_i$ occurs in $D$
4. The valuation $V$ generated by $I$ has for some $w \in W$ $V(D, w) = 0$

Equivalently, $\mathbf{GL} \vdash D$ if and only if for all models $\mathcal{M}_I = \langle W, R, I\rangle$ such that $|W| \leq 2^k$, $R$ is transitive and irreflexive, and $I(p_i, w) = 1$ if and only if $p_i$ occurs in $D$, we have that for all $w \in W$, $V_{\mathcal{M},I}(D, w) = 1$.

The method used to proof the Completeness Theorem is not a decision procedure. It involves finding the maximal $D$-consistent sets, which involves deciding whether or not $\mathbf{GL} \nvdash \neg \bigwedge X$ for some set $X$ containing exactly one of $S$ or $\neg S$ for each subsentence of $D$, which is just determining the provability of more modal sentences. A procedure we can (theoretically)

complete is to construct all the transitive and irreflexive relations on a set of $n$ elements. A transitive, irreflexive relation on a set is also known as a strict partial order. Further we do not care how the elements of the set are labelled, so we want to produce all the strict partial orders on a set of unlabelled elements. It is easy to show that the models $\langle W, R, I\rangle$ and $\langle W \cup \{x\}, R, I'\rangle$ (where $I'$ is any function that extends $I$ to all $\langle p_i, x\rangle$ pairs in any way) prove exactly the same modal sentences at the worlds they share, so we have that: $D$ is valid in all **GL**-appropriate frames of size less than $n$ if and only if $D$ is valid in all **GL**-appropriate frames of size $n$.

Whence are equivalence reduces to: $\mathbf{GL} \vdash D$ if and only if for all models $\mathcal{M}_I = \langle W, R, I\rangle$ such that $|W| = 2^k$, $R$ is a strict partial order on $W$, and $I(p_i, w) = 1$ if and only if $p_i$ occurs in $D$, we have that for all $w \in W$, $V_{\mathcal{M},I}(D, w) = 1$.

A consequence of the inductive definitions of sentences mean that all sentences are uniquely readable. Thus given a sentence one can determine its immediate subsentences. This process can be repeated until one reaches the atomic sentences, and thus it is easy to determine the number of subsentences of a sentence. Also, given an unlabelled set of size $n$, it is possible to produce the set of all strict partial orders on that set. Further, there are decision procedures that decide whether a modal sentence is true at a world in a model, and hence there is a terminating procedure for deciding whether or not a modal sentence is provable or not in **GL**. Here is a sketch of how that procedure would go:

> Set a modal sentence $D$.
>
> Let $n =$ number of subsentences of $D$.
>
> Let $m =$ number of partial orders on unlabeled sets of size $2^n$.
>
> Label the frames $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_m$ (The interpretation for each is defined: $I(p_i, w) = 1$ if and only if $p_i$ occurs in $D$, which is decidable).
>
> For $i = 1$ to $i = m$;
>
> label the elements of $\mathcal{M}_i$, $1, \ldots, 2^n$,

For $j = 1$ to $j = 2^n$;

If $(V_{\mathcal{M}_i,I}(D, j) = 0)$, then end and return $\mathbf{GL} \nvdash D$;

If for all $i$, all $j$, $V_{\mathcal{M}_i,I}(D, j) = 1$ return $\mathbf{GL} \vdash D$.

This is not efficient. The number of strict partial orders on a set is the same as the number of partial orders on a set, and [7] tells us the number of partial orders on an unlabelled sets of sizes from $n = 1$ to 16 are: 1, 2, 5, 16, 63, 318, 2045, 16999, 183231, 2567284, 46749427, 1104891746, 33823827452, 1338193159771, 68275077901156, 4483130665195087. So even for a modal sentence with only 4 subsentences, the number of models to check is 4483130665195087, which is about $4.5 \times 10^{16}$.

The decidability of **GL** together with the main result of this paper – that **GL** is the provability logic **PA**– is of interest because **PA** is not itself decidable. This means that a hugely interesting portion of **PA** *is* decidable.

# 3 Peano Arithmetic

In this section we define the syntax and deductive system of **PA**, show how proofs and formulae may be coded into **PA** and prove the Generalised Diagonal Lemma.[10]

## 3.1 The language of PA

**Definition 3.1.** *The language of Peano Arithmetic is a first-order language $\mathcal{L}_{PA}$ consisting of the following symbols:*

1. *Logical symbols: variable symbol $v$, prime $'$, $\bot$, $\rightarrow$, $\forall$, $=$, ( and )*
2. *Non-logical symbols: constant symbol 0, one-place function symbol $\boldsymbol{s}$, and two-place function symbols $+$ and $\times$*

[10] The majority of the material in this section is taken from [1], chapters 2 and 3.

**Definition 3.2.** *We define the variables of Peano Arithmetic inductively as follows:*

1. *$v$ is a variable.*

2. *If $x$ is a variable, then the concatenation of $x$ with the subscript $\prime$, $x_\prime$ , is a variable.*

**Definition 3.3.** *We define the terms of Peano Arithmetic inductively as follows:*

1. *$0$ and all variables are terms.*

2. *if $t$ is a term, $\boldsymbol{s}t$ is a term.*

3. *if $t$ and $t'$ are terms, then $(t + t')$ and $(t \times t')$ are terms.*

**Definition 3.4.** *Atomic formulae of the language are defined as follows:*

1. *for terms $t$ and $t'$, $t = t'$ is an atomic formula*

2. *$\bot$ is an atomic formulae*

*The formulae of the language of Peano Arithmetic are defined inductively as follows:*

1. *All atomic formulae are formulae*

2. *if $\phi$ and $\psi$ are formulae, then $(\phi \to \psi)$ is a formula*

3. *if $\phi$ is a formula and $x$ a variable, then $\forall x\phi$ is a formula.*

We will use the greek letters $\phi$, $\psi$ and $\chi$ as our metavariables for formulae. We will continue to use the short-hands $\wedge$, $\vee$, $\neg$, $\leftrightarrow$ as given in Section 2.1, and further write $\exists x\phi$ for $\neg\forall x\neg\phi$. We will not be strict about the symbols we use as variables.

We will assume some familiarity with predicate logic, such as what it is for a variable to occur freely, in a formula, the recursive definitions of term substitution and a term being free for a variable in a formula.

**Definition 3.5.** *A formula is a sentence if it contains no free variables.*

## 3.2 Deductive System of Peano Arithmetic

Peano Arithmetic is a deduction system extending first-order predicate logic.

**Definition 3.6.** *Let $\phi$, $\psi$ and $\chi$ be $\mathcal{L}_{PA}$ formulae and $x$ and $y$ be variables. The logical axioms of **PA** are:*

*A1.* $\phi \to (\psi \to \phi)$

*A2.* $(\phi \to (\psi \to \chi)) \to ((\phi \to \psi) \to (\phi \to \chi))$

*A3.* $((\phi \to \bot) \to (\psi \to \bot)) \to (\psi \to \phi)$

*A4.* $\forall x\phi \to \phi[t/x]$ *if* $t$ *is a term free for* $x$ *in* $\phi$

*A5.* $\forall x(\phi \to \psi) \to (\phi \to \forall x\psi)$ *if* $x \notin Free(\phi)$

*A6.* $\forall x\, x = x$

*A7.* $x = y \to (\phi \to \phi')$ *where* $\phi$ *is atomic and* $\phi'$ *is obtained from* $\phi$ *by replacing some occurrences of* $x$ *with* $y$.

*The non-logical axioms of **PA** are:*

*B1.* $0 = sx \to \bot$

*B2.* $sx = sy \to x = y$

*B3.* $(x + 0) = x$

*B4.* $(x + \boldsymbol{s}y) = \boldsymbol{s}(x + y)$

*B5.* $(x \times 0) = 0$

*B6.* $(x \times \boldsymbol{s}y) = ((x \times y) + x)$

*And for all formulae $\phi$, variables $x$ and $y$ different from $x$ and not occurring in $\phi$, the induction axioms*[11] *of* ***PA****:*

$$(\forall x(x = 0 \to \phi) \wedge \forall y(\forall x(x = y \to \phi) \to \forall x(x = \boldsymbol{s}y \to \phi))) \to \phi$$

*The rules of deduction are MP (as in Definition 2.2) and Generalisation:*

*GEN: if* $\vdash \phi$*, then* $\vdash \forall x\phi$ *(for any variable* $x$*)*

As before, we write $\mathbf{PA} \vdash \phi$ (or simply $\vdash \phi$ if the context is clear) if $\phi$ is provable from the above axioms and rules, where provability is defined just as in Definition 2.3.

When we talk of a formula of **PA** being true, we will mean true in the standard model of arithmetic; the natural numbers $\mathbb{N}$ with the familiar successor, addition and multiplication functions. Axioms A1-A7 are such that they are sound and complete with respect to the semantics of first-order predicate logic (Gödel's Completeness Theorem). The axioms B1-B6 give a correct characterisation of their corresponding functions. The induction axioms characterise formally the notion of "proof by induction": if 0 has the property expressed by $\phi$ and for each $n$ that has that property, $n+1$ too has that property, then every number has the property expressed by $\phi$.

Many standard arithmetical definitions and results be given and derived in **PA**. For example the less-than relation: $x < y := \exists z\,(x + \mathbf{s}z) = y$, and the commutativity of addition: $(x + y) = (y + x)$.

In **PA** we do not have individual symbols to denote each of the natural numbers, only 0. In order to ease notation, for natural number $n$, we will write $\bar{n}$ to denote the term of **PA** that is $\underbrace{\mathbf{s}\ldots\mathbf{s}}_{n \text{ times}}0$.

**Definition 3.7.** *$\forall x$ is a bounded quantifier in $\forall x\phi$ if $\phi$ is equivalent to $x < y \to \psi$, where $y \neq x$, or $\phi$ is equivalent to $x < \bar{n} \to \psi$ for some $n \in \mathbb{N}$.*

[11] Using the short-hand $\wedge$ for readability here.

*$\exists x$ is a bounded quantifier in $\exists x\phi$ if $\phi$ is equivalent to $x < y \wedge \psi$, where $y \neq x$, or $\phi$ is equivalent to $x < \bar{n} \wedge \psi$ for some $n \in \mathbb{N}$*

We will write $\forall x < y\, \phi$ for $\forall x(x < y \to \phi)$ and $\exists x < y\, \phi$ for $\exists x(x < y \wedge \phi)$.

**Definition 3.8.** *Let $\boldsymbol{x} = (x_1, \ldots, x_n)$ and suppose $\phi(\boldsymbol{x}, y)$ is a formula of the language of **PA**, in which the variables $y$ and $x_i$ for all $0 \leq i \leq n$ are free. We say that $\phi(\boldsymbol{x}, y)$ is a **pterm** if **PA** $\vdash \exists y(\phi(\boldsymbol{x}, y) \wedge \forall z(\phi(\boldsymbol{x}, z) \to y = z))$*

The definition of pterm given above says that for each $n$-tuple $\boldsymbol{x}$, there is a unique $y$ such that $\phi(\boldsymbol{x}, y)$ holds; a formula $\phi$ is a pterm (p for pseudo) if it represents an $n$-place function. We will write $\exists y(\phi(\boldsymbol{x}, y) \wedge \forall z(\phi(\boldsymbol{x}, z) \to y = z))$ as $\exists! y\, \phi(\boldsymbol{x}, y)$. This is a useful representation to have, as it allows discussion within the language of **PA** of functions that are not denoted by a term of the language.

**Theorem 3.9.** *All terms of **PA** denote polynomial functions.*

*Proof.* By induction on the structure of terms:
Base case: 0 is a polynomial, and for variable $x$, $x$ is a polynomial (the identity map on $\mathbb{N}$).
Inductive case: Presume that for all terms of length $< k$ that they denote polynomial functions. Let $t$ be a term of length $k$, and let $t_0$ and $t_1$ have length $< k$ and denote polynomials $f(x_1, \ldots, x_n)$ and $g(y_1, \ldots, y_m)$ respectively. We have 3 cases to consider: $t = \mathbf{s}t_0$, $t = t_0 + t_1$, and $t = t_0 \times t_1$. In the 3 cases, $t$ denotes the functions $f + 1$, $f + g$ and $fg$ respectively, which are all equivalent to polynomials. Hence by induction all terms of **PA** denote polynomials. ■

We can deduce from this that functions that do not have a polynomial equivalent do not have a denotation in the terms of **PA**, thus showing the need for the use of pterms to define functions. An example of such a function is $y = 2^x$, which does not have a polynomial equivalent[12].

[12] $2^x$ majorizes every polynomial; for any polynomial $p(x)$, there exists some value $t$ such that $p(t) < 2^t$.

## 3.3 The Arithmetical Hierarchy

The Arithmetical Hierarchy classifies all formulae of our language with respect to the number of alternations between sequences for $\forall x$ and sequences of $\exists x$ quantifiers in that formula's Prenex Normal Form (PNF)[13].

**Definition 3.10.** *The Arithmetical Hierarchy of formulae is defined inductively as follows:*

1. *$\phi$ is a strict $\Sigma_0$ formula if it is a member of the smallest class of formulae containing all atomic formulae and closed under Boolean operations, and bounded quantification.*
2. *$\phi$ is $\Sigma_0 = \Pi_0 = \Delta_0$ if it is provably equivalent to a strict $\Sigma_0$ formula.*
3. *$\phi$ is a $\Sigma_{n+1}$ formula if $\phi$ is provably equivalent to $\exists y_1 \dots \exists y_k \psi$ for some $k$, where $\psi$ is $\Pi_n$*
4. *$\phi$ is a $\Pi_{n+1}$ formula if $\phi$ is provably equivalent to $\forall y_1 \dots \forall y_k \psi$ for some $k$, where $\psi$ is $\Sigma_n$*
5. *$\phi$ is a $\Delta_n$ formula if $\phi$ is provably equivalent to both a $\Sigma_n$ and a $\Pi_n$ formula*

We can see that every formula that is equivalent to some $Q_1 y_1 \dots Q_n y_n \psi$ where $\psi$ is $\Sigma_0$ receives a classification in this hierarchy.

The class $\Sigma_0$ is also closed under substitution of $\Sigma_1$ pterms.

**Definition 3.11.** *A formal system $S$* ***correctly decides*** *a formula $\phi$ if either $\phi$ is true and $S \vdash \phi$ or $\phi$ is false and $S \vdash \neg\phi$.*

The following is adapted from [5], p66-68.

**Lemma 3.12.** ***PA*** *correctly decides all $\Sigma_0$ sentences.*

[13] A formula $\phi$ is in PNF if $\phi = Q_1 x_1 \dots Q_n x_n \psi$ where each $Q_i \in \{\forall, \exists\}$ and $\psi$ is quantifier free. Every first-order formula is provably equivalent to one in PNF.

*Proof.* We will make the following reasonable assumptions about what can be proven in **PA**:

1. All true atomic sentences can be proven.
2. If $m \neq n$, then $\mathbf{PA} \vdash \neg \bar{m} = \bar{n}$
3. For any variable $x$ and number $n$, $\mathbf{PA} \vdash x < \bar{n} \leftrightarrow (x = 0 \vee \cdots \vee x = \overline{n-1})$

We then use induction on the construction of $\Sigma_0$ formulae:

Base case:

Let $\phi$ be a false atomic sentence, so $\phi$ is $t = t'$ for some closed terms $t$ and $t'$ denoting $i$ and $i'$, where $i \neq i'$. Then the atomic formulae $t = \bar{i}$ and $t' = \bar{i'}$ are both true, and hence by (1), provable. By (2), $\mathbf{PA} \vdash \neg \bar{i} = \bar{i'}$ so by predicate logic $\mathbf{PA} \vdash \neg t = t'$. Whence by this and (1), **PA** decides all atomic sentences.

Inductive case:

Suppose the formulae $\psi$ and $\chi$ are decidable by **PA**, and let $\phi = \psi \to \chi$. If $\phi$ true, $\psi$ is false or $\chi$ is true, so $\mathbf{PA} \vdash \neg\psi$ or $\mathbf{PA} \vdash \chi$, whence by logic $\mathbf{PA} \vdash \phi$. If $\phi$ is false, $\psi$ is true and $\chi$ is false, so $\mathbf{PA} \vdash \psi$ and $\mathbf{PA} \vdash \neg\chi$, whence by logic $\mathbf{PA} \vdash \neg\phi$ .

Suppose $\phi = \forall x < \bar{n}\, \psi(x)$ where $\psi(x)$ is $\Sigma_0$ and free in just the variable $x$.

If $\phi$ is true, then for all $i < n$, $\psi(\bar{i})$ is true, and hence by the inductive assumption, for all $i < n$, $\mathbf{PA} \vdash \psi(\bar{i})$. By predicate logic, for all $i < n$, $\mathbf{PA} \vdash x = \bar{i} \to \psi(x)$ whence by (3), and propositional logic, $\mathbf{PA} \vdash x < \bar{n} \to \psi(x)$. By GEN, $\mathbf{PA} \vdash \phi = \forall x < \bar{n}\, \psi(x)$.

If $\phi$ is false, there exists some $i < n$ such that $\psi(\bar{i})$ is false, and hence by assumption $\mathbf{PA} \vdash \neg\psi(\bar{i})$. By logic, $\mathbf{PA} \vdash \neg(\bar{i} = \bar{i} \to \psi(\bar{i}))$. By (3), $\mathbf{PA} \vdash \bar{i} = \bar{i} \leftrightarrow \bar{i} < \bar{n}$, and hence by logic $\mathbf{PA} \vdash \neg(\bar{i} < \bar{n} \to \psi(\bar{i}))$. The following holds by predicate logic: $\mathbf{PA} \vdash \neg(\bar{i} < \bar{n} \to \psi(\bar{i})) \to \neg\forall x < \bar{n}\, \psi(x)$ and hence by modus ponens, $\mathbf{PA} \vdash \neg\phi = \neg\forall x < \bar{n}\, \psi(x)$.

Hence by induction we have shown that **PA** correctly decides all $\Sigma_0$ sentences. ■

**Corollary 3.13** ($\Sigma_0$-Completeness)**.** *If $\phi$ is a true $\Sigma_0$ sentence, then $\boldsymbol{PA} \vdash \phi$.*

**Theorem 3.14** ($\Sigma_1$-Completeness)**.** *If $\phi$ is a true $\Sigma_1$ sentence, then $\boldsymbol{PA} \vdash \phi$.*

*Proof.* By induction on the construction of $\Sigma_1$ sentences.

If $\phi$ is a $\Sigma_0$ sentence, then $\mathbf{PA} \vdash \phi$ by $\Sigma_0$-completeness.

If $\phi = \exists x \psi(x)$ is true where $\psi(x)$ is $\Sigma_0$, then there is some $i$ such that $\psi(\bar{i})$ is true, and by $\Sigma_0$-completeness $\mathbf{PA} \vdash \psi(\bar{i})$. Therefore $\mathbf{PA} \vdash \exists x \psi(x)$.

If $\phi$ is provably equivalent to $\exists x \psi$, where $\psi$ is $\Sigma_0$, then $\phi$ is provable. So any true $\Sigma_1$ sentence is provable in **PA**. ■

## 3.4 Coding Formulae and Proofs in PA

Our goal in this section is to give an outline of how the formulae and proofs of **PA** may be coded within **PA**. Such a coding system gives us the tools to code the provability predicate $Pr(\overline{\ulcorner \urcorner})$ and allows us to reason about **PA**'s metatheory within **PA**.

**Definition 3.15.** *Let $\mathcal{S}(\boldsymbol{PA})$ be the class of symbols, terms, formulae, and sequence of formulae of the language of **PA**. A **Gödel Numbering** is an injective function $G : \mathcal{S}(\boldsymbol{PA}) \to \mathbb{N}$.*

Given a Gödel Numbering $G$, we write $\ulcorner \phi \urcorner$ for $G(\phi)$, and hence $\overline{\ulcorner \phi \urcorner}$ for the corresponding term of **PA**.

There are an uncountable number of possible Gödel Numberings[14], but there are some that are more useful to us that others. A *useful* system of numbering would be one that respects the inductive formation rules of terms and formulae, i.e.$\ulcorner \phi \to \psi \urcorner$ is a describable function of $\ulcorner \phi \urcorner$, $\ulcorner \to \urcorner$, and $\ulcorner \psi \urcorner$. Such numberings are useful as definable formulae can be given that determine whether $x$ is the Gödel Number of a formula, or a proof.

What we will see in this section is that a useful way of encoding everything of interest (only well-formed formulae and proofs) if we can give a formula of **PA** that encodes the notion of ordered pairs into **PA**.

[14] $|\{\text{Terms of } \mathbf{PA}\}| = |\mathbb{N}| = \aleph_0$, hence $|\{\text{Gödel Numberings}\}| > |\{\text{injective functions from } \mathbb{N} \text{ to } \mathbb{N}\}| = 2^{\aleph_0}$

To begin this encoding, we first will consider each formation rule of terms and formulae as the creation of ordered pairs:

Variables: – $x\prime$ is $\langle x, \prime \rangle$

Terms: – $(t + t')$ is $\langle +, \langle t, t' \rangle \rangle$

– $(t \times t')$ is $\langle \times, \langle t, t' \rangle \rangle$

– $\mathbf{s}t$ is $\langle \mathbf{s}, t \rangle$

Formulae: – $t = t'$ is $\langle =, \langle t, t' \rangle \rangle$

– $(\phi \to \psi)$ is $\langle \to, \langle \phi, \psi \rangle \rangle$

– $\forall x \phi$ is $\langle \forall, \langle x, \phi \rangle \rangle$

Recall the definition of a proof in **PA**: a proof of $\phi$ is a finite sequence whose last formula is $\phi$ and for each formula of the sequence $x$, $x$ is an axiom, or follows from previous terms by the rules MP or GEN. We can give similar definitions for variables, terms, atomic formulae using the above definitions in terms of ordered pairs. For example: $t$ is a term if there is a finite sequence $a = \langle a_1, \ldots, a_n \rangle$ such that $a_n = t$, and for each $1 \leq i \leq n$ $a_i$ is a variable, $\perp$ or 0, or there are $j, k < i$ such that $a_i = \langle +, \langle a_j, a_k \rangle \rangle$ or $\langle \times, \langle a_j, a_k \rangle \rangle$ or $\langle a, a_j \rangle$. We can also say what it is for one formula to follow from another by a rule of deduction, for example: $\phi$ follows from $\psi$ and $\chi$ by MP if $\phi$ is a formula and $\psi$ is a formula and $\chi = \langle \to, \langle \psi, \phi \rangle \rangle$.

Now to encode the terms and formulae of **PA**, we can give each symbol of our language a unique Gödel number, and then give a 2-place function that assigns Gödel Numbers to pairs. Ordered pairs are uniquely defined by their constituent elements (in that order) then any 2-place function $f$ such that $f(x, y) = f(x', y')$ implies $x = x'$ and $y = y'$ will work, and will respect the formation rules for terms, formula and proofs. For example, if $\overline{\ulcorner \forall \urcorner} = a$, $\overline{\ulcorner x \urcorner} = b$ and $\overline{\ulcorner \phi \urcorner} = c$, then this way of numbering formulae gives $\overline{\ulcorner \forall x \phi \urcorner} = f(a, f(b, c))$, and no other formula, term or variable will receive this Gödel Number. If we give each symbol of

our language an odd Gödel Number, then the function (which is certainly definable in **PA**) $\langle x, y\rangle = 2((x+y)(x+y)+x+1)$ satisfies the requirements of encoding pairs, and gives each pair an even Gödel Number, so this system of numbering is injective.

Useful consequences of the above encoding of ordered pairs are that $x, y < \langle x, y\rangle$, and pairs may be ordered as follows: if $x < x'$ then $\langle x, y\rangle < \langle x', y\rangle$, and if $y < y'$ then $\langle x, y\rangle < \langle x, y'\rangle$. We may also assume that $\Sigma_1$ pterms may be given describing the projection functions $\pi_1(\langle x, y\rangle) = x$ and $\pi_2(\langle x, y\rangle) = y$.

We will now see that the notion of finite sequences too can be encoded into **PA** by means of ordered pairs, and a function due to Gödel. This allows the definition of terms, formulae and derivations to be given complete translations into the language of **PA**.

**Definition 3.16.** *$Rm(x, d, r)$ is the $\Sigma_0$ pterm $((r < d \wedge \exists q\, x = q \times d + r) \vee (d = 0 \wedge r = x))$, expressing the function $r = rm(x, d) = x$ mod $d$.*

**Definition 3.17** (Gödel's $\beta$-function)**.** *$\beta(a, b, i)$ is the function expressed by the $\Sigma_0$ pterm $Rm(a, 1 + ((i + 1) \times b), r)$*

**Lemma 3.18** (Gödel's $\beta$-function Lemma)**.** *For any finite sequences $\langle s_1, \ldots, s_n\rangle$, there exist natural numbers $a$ and $b$ such that for all $1 \leq i \leq n$, $\beta(a, b, i) = s_i$. Further, for $k = max(n, s_1, \ldots, s_n) + 1$, $a$ and $b$ can be chosen such that $b < lcm[i + 1 : i < k] + 1$ and $a < lcm[1 + (i + 2)b : i < n]$ (where lcm stands for lowest common multiple, which is a pterm).*

*Proof.* See [1], p31-32. ■

Given this lemma, and given we have a function defining the notion on an ordered pair in **PA**, we will hence encode each finite sequence $s = \langle s_1, \ldots, s_n\rangle$ as the triple $\langle\langle a, b\rangle, n\rangle$, where $\langle a, b\rangle$ are the least pair of natural numbers that satisfy the conditions of Gödel's $\beta$-function Lemma. Consequentially we obtain the following definitions:

**Definition 3.19.** *$FinSeq(s)$ is the $\Sigma_0$ formula:*

$\exists a < s \exists b < s \exists k < s(s = \langle\langle a, b\rangle, k\rangle \wedge \forall c < s \forall d < s(\langle c, d\rangle < \langle a, b\rangle \rightarrow \exists i < k \beta(c, d, i) \neq \beta(a, b, i))$

**Definition 3.20.** *$lh(s) = \pi_2(s)$ is the length of the finite sequence $s = \langle\langle a, b\rangle, n\rangle$.*

**Definition 3.21.** *$s_i = \beta(\pi_1(\pi_1(s)), \pi_2(\pi_1(s)), i)$ is the value of the $i$th term of the finite sequence $s = \langle\langle a, b\rangle, n\rangle$.*

Now that we have said what it is for a number to code either an ordered pair or a finite sequence, it is now possible to give formulae of **PA** determine whether $n$ is the Gödel Number of a term, a formula, or proof of **PA**.

**Definition 3.22.** *1. $Variable(x)$ is defined as:*

$\exists s(FinSeq(s) \wedge \neg lh(s) = 0 \wedge s_1 = \overline{\ulcorner v \urcorner} \wedge s_{lh(s)} = x \wedge (1 < lh(s) \rightarrow \forall i \leq lh(s)\, s_i = \langle s_{i-1}, \overline{\ulcorner \prime \urcorner}\rangle))$

*2. $Term(x)$ is defined as:*

$\exists s(FinSeq(s) \wedge \neg lh(s) = 0 \wedge s_{lh(s)} = x \wedge \forall i \leq lh(s)(Variable(s_i) \vee s_i = \overline{\ulcorner 0 \urcorner} \vee$
$\exists j < i \exists k < i(s_i = \langle \overline{\ulcorner + \urcorner}, \langle s_j, s_k\rangle\rangle \vee s_i = \langle \overline{\ulcorner \times \urcorner}, \langle s_j, s_k\rangle\rangle \vee s_i = \langle \overline{\ulcorner \boldsymbol{s} \urcorner}, s_j\rangle)))$

*3. $AtomicFormula(x)$ is defined as:*

$x = \overline{\ulcorner \bot \urcorner} \vee \exists y < x \exists z < x\, (Term(y) \wedge Term(z) \wedge x = \langle \overline{\ulcorner = \urcorner}, \langle y, z\rangle\rangle)$

*4. $Formula(x)$ is defined as:*

$\exists s(FinSeq(s) \wedge \neg lh(s) = 0 \wedge s_{lh(s)} = x \wedge \forall i \leq lh(s)(AtomicFormula(s_i) \vee$
$\exists j < i \exists k < i(s_i = \langle \overline{\ulcorner \rightarrow \urcorner}, \langle s_j, s_k\rangle\rangle \vee \exists w < s_i(Variable(w) \wedge s_i = \langle \overline{\ulcorner \forall \urcorner}, \langle w, s_j\rangle\rangle)))$

N.B: In the definitions above there are unbounded quantifiers, which would suggest that the above formulae are all $\Sigma_1$. However, the $\beta$ -function allows that bounds can be put on these variables, and hence they are in fact $\Sigma_0$. We omit the proof of this here, but a similar proof is given in [1], p42.

We can give formulae for each of the axioms of **PA** that says what it is to be the Gödel Number of a formula of that form, for example, for axiom A1, $(\phi \to (\psi \to \phi)$:

$x$ is the Gödel Number of a copy of the axiom A1 if:
$\exists y < x \exists z < x(Formula(y) \wedge Formula(z) \wedge x = \langle \overline{\ulcorner \to \urcorner}, \langle y, \langle \overline{\ulcorner \to \urcorner}, \langle z, y \rangle \rangle \rangle)$

We will not give explicit formulae for all the other axioms of **PA**, as it is a long list. We can now define what it is to be the Gödel Number of a proof of **PA**, and hence define the formula $Pr(x)$.

**Definition 3.23.** *1. $Axiom(x)$ is the conjunction of the formulae that describe each of the axioms of **PA**.*

*2. $ByMP(x, y, z)$ which says that $x$ follows from $y$, $z$ by MP is:*
$Formula(y) \wedge Form(z) \wedge z = \langle \overline{\ulcorner \to \urcorner}, \langle y, x \rangle \rangle$

*3. $ByGEN(x, y)$ which says that $x$ follows from $y$ by GEN is:*
$Formula(y) \wedge \exists z < x Variable(z) \wedge x = \langle \overline{\ulcorner \forall \urcorner}, \langle w, y \rangle \rangle$

*4. $Proof(y, x)$ which says that $y$ is a proof of $x$ is:*
$FinSeq(y) \wedge y_{lh(y)} = x \wedge \forall i \leq lh(y)(Axiom(y_i) \vee \exists j < i \exists k < i(ByMP(y_i, y_j, y_k) \vee ByGEN(y_i, y_j))$

**Definition 3.24.** *$Pr(x)$ is defined to be $\exists y Proof(y, x)$.*

Note that $Proof(y, x)$ is $\Sigma_0$, and hence $Pr(x)$ is a $\Sigma_1$ formula.

Further to the Hilbert-Bernays-Löb Derivability Conditions and Löb's Theorem:

1. If $\vdash \phi$, then $\vdash Pr(\overline{\ulcorner \phi \urcorner})$

2. $\vdash Pr(\overline{\ulcorner \phi \to \psi \urcorner}) \to (Pr(\overline{\ulcorner \phi \urcorner}) \to Pr(\overline{\ulcorner \psi \urcorner}))$

3. $\vdash Pr(\overline{\ulcorner \phi \urcorner}) \to Pr(\overline{\ulcorner Pr(\overline{\ulcorner \phi \urcorner}) \urcorner})$

**LT**. If $\mathbf{PA} \vdash Pr(\overline{\ulcorner \phi \urcorner}) \rightarrow \phi$, then $\mathbf{PA} \vdash \phi$

The following theorem can also be proven:

**Theorem 3.25.** *If $\phi$ is a $\Sigma_1$ formula, then $\boldsymbol{PA} \vdash \phi \rightarrow Pr(\overline{\ulcorner \phi \urcorner})$*

*Proof.* See [1], pp46-49. ■

## 3.5 The Generalised Diagonal Lemma

**Definition 3.26.** *Let $SUB(y, x_0, \ldots, x_n, z)$ be a pterm for the $(n+2)$-place function $sub(y, x_0, \ldots, x_n)$, whose value at $a, b_0, \ldots, b_n$ is $\overline{\ulcorner G_a(\bar{b_0}, \ldots, \bar{b_n}) \urcorner}$ where $G_a(x_0, \ldots, x_n)$ is the formula with Gödel Number $a$.*

**Lemma 3.27.** *$SUB(y, x_0, \ldots, x_n, z)$ is a $\Sigma_1$ formula.*

*Proof.* We sketch a proof, making some assumptions about certain formula that can be given and that are $\Sigma_1$. For $SUB$ to be a $\Sigma_1$ pterm, we must make some (reasonable) assumptions about the system of Gödel Numbering we have in place.

Let $NUM(x, y)$ be the $\Sigma_1$ formula:

$$\exists s(FinSeq(s) \wedge lh(s) = x + 1 \wedge s_0 = \overline{\ulcorner 0 \urcorner} \wedge \forall i < x s_{i+1} = \langle \overline{\ulcorner s \urcorner}, s_i \rangle \wedge s_x = y)$$

We can see that $NUM(x, y)$ is a $\Sigma_1$ pterm describing the function $y = num(x) = \overline{\ulcorner x \urcorner}$.

By the definition of $Variable(x)$ and of the function $\langle x, y \rangle$, there exists a pterm $VAR(x, y)$ describing the function $var(x) = \overline{\ulcorner v_x \urcorner}$ (where $v_x$ is the concatenation of the variable symbol $v$ with $x$ occurrences of $\prime$). We may assume that $VAR(x, y)$ is $\Sigma_1$.

We assume that we are given a $\Sigma_1$ pterm $subst(t, i, x)$ such that it returns the Gödel Number of the formula obtained by substituting the term with the value of $t$ for all free occurrences of the variable with Gödel Number $i$ in the formula with Gödel Number $x$.

Let $SUB(y, v_{k_0}, \ldots, v_{k_n}, z)$ be:

$z = subst(num(v_{k_n}), var(k_n), \ldots, subst((num(v_{k_1}), var(k_1), subst((num(v_{k_0}), var(k_0), y) \ldots)$

By the definitions of *num* and *subst*, at $a, b_0, \ldots, b_n$ this gives the Gödel Number of the formula obtained by substituting for each $1 \leq i \leq n$, the term with the value of $num(\bar{b_i}) = \overline{\ulcorner \bar{b_i} \urcorner}$ for all free occurrences of the variable with Gödel Number $var(k_i) = \overline{\ulcorner v_{k_i} \urcorner}$, in the formula with Gödel Number $a$. This is just the definition of $SUB$.

Further, as *subst*, *var*, and *num* are all $\Sigma_1$ pterms, it follows that $SUB(y, v_{k_0}, \ldots, v_{k_n}, z)$ is also $\Sigma_1$. ■

**Theorem 3.28** (The Generalised Diagonal Lemma)**.** *Suppose that $y_0, \ldots, y_n, z_1, \ldots, z_m$ are distinct variables, and that $\phi_0(y_1, \ldots, y_n, \boldsymbol{z}), \ldots, \phi_n(y_1, \ldots, y_n, \boldsymbol{z})$ are formulae of the language of **PA** in which all free variables are among $y_0, \ldots, y_n, z_1, \ldots, z_m$. Then, there exist formulae $\chi_0(\boldsymbol{z}), \ldots, \chi_n(\boldsymbol{z})$ of the language of **PA** in which all free variables are among $z_1, \ldots, z_m$, such that:*

*For each $0 \leq i \leq n$, $\boldsymbol{PA} \vdash \chi_i(\boldsymbol{z}) \leftrightarrow \phi_i(\overline{\ulcorner \chi_0(\boldsymbol{z}) \urcorner}, \ldots, \overline{\ulcorner \chi_n(\boldsymbol{z}) \urcorner}, \boldsymbol{z})$*

*Proof.* For each $0 \leq i \leq n$, let $p_i$ be the Gödel Number of:

$$\phi_i(sub(x_0, x_0, \ldots, x_n), \ldots, sub(x_n, x_0, \ldots, x_n), \mathbf{z})$$

and let $\chi_i(\mathbf{z})$ be the formula:

$$\phi_i(sub(\bar{p_0}, \bar{p_0}, \ldots, \bar{p_n}), \ldots, sub(\bar{p_n}, \bar{p_0}, \ldots, \bar{p_n}), \mathbf{z})$$

From this and the definition of *sub*, we can deduce that:

$$sub(\bar{p_i}, \bar{p_0}, \ldots, \bar{p_n}) = \overline{\ulcorner G_{p_i}(\bar{p_0}, \ldots, \bar{p_n}, \mathbf{z}) \urcorner} = \overline{\ulcorner \chi_i(\mathbf{z}) \urcorner}$$

That is to say, $SUB(\bar{p_i}, \bar{p_0}, \ldots, \bar{p_n}, \overline{\ulcorner \chi_i(\mathbf{z}) \urcorner})$ is true. Equivalently we have that:

$$\chi_i(\mathbf{z}) = \phi_i(sub(\bar{p_0}, \bar{p_0}, \ldots, \bar{p_n}), \ldots, sub(\bar{p_n}, \bar{p_0}, \ldots, \bar{p_n}), \mathbf{z}) = \phi_i(\overline{\ulcorner \chi_0(\mathbf{z}) \urcorner}, \ldots, \overline{\ulcorner \chi_n(\mathbf{z}) \urcorner}, \mathbf{z})$$

In other words, $\chi_i(\mathbf{z}) \leftrightarrow \phi_i(\overline{\ulcorner \chi_0(\mathbf{z}) \urcorner}, \ldots, \overline{\ulcorner \chi_n(\mathbf{z}) \urcorner}, \mathbf{z})$ is true.

By substitution of equivalent terms, $\mathbf{PA} \vdash SUB(\bar{p_i}, \bar{p_0}, \ldots, \bar{p_n}, \overline{\ulcorner \chi_i(\mathbf{z}) \urcorner}) \rightarrow (\chi_i(\mathbf{z}) \leftrightarrow \phi_i(\overline{\ulcorner \chi_0(\mathbf{z}) \urcorner}, \ldots, \overline{\ulcorner \chi_n(\mathbf{z}) \urcorner}, \mathbf{z}))$.

By Lemma 3.27, $SUB(\bar{p_i}, \bar{p_0}, \ldots, \bar{p_n}, \overline{\ulcorner \chi_i(\mathbf{z}) \urcorner})$ is a $\Sigma_1$ sentence and true, so by Theorem 3.14, it is indeed the case that $\mathbf{PA} \vdash SUB(\bar{p_i}, \bar{p_0}, \ldots, \bar{p_n}, \overline{\ulcorner \chi_i(\mathbf{z}) \urcorner})$, as required. ■

# 4 Arithmetical Soundness of GL

In this section, we will show that every theorem of **GL** is provable under any appropriate translation in **PA**; the Arithmetical Soundness of **GL**.[15]

**Definition 4.1.** *A **realisation** is a function $*$ from the set of sentence letters of our propositional language to the sentences of **PA**.*

A realisation can be naturally extended to a *translation* from all modal sentences to the language of **PA**.

**Definition 4.2.** *The **translation** $S^*$ of a modal sentence $S$ induced by a realisation $*$ is defined inductively as follows:*

1. $\bot^*=\bot$
2. *If $S$ is a propositional variable, then $S^* = *(S)$*
3. *If $S = T \to U$, then $S^* = T^* \to U^*$*
4. *If $S = \Box T$, then $S^* = Pr(\overline{\ulcorner T^* \urcorner})$*

**Theorem 4.3** (Arithmetical Soundness of GL)**.** *If **GL** $\vdash S$, then for every realisation $*$, **PA** $\vdash S^*$.*

*Proof.* Let us first remind ourselves of the derivability conditions and Löb's Theorem:

1. If $\vdash S$, then $\vdash Pr(\overline{\ulcorner S \urcorner})$
2. $\vdash Pr(\overline{\ulcorner S \to T \urcorner}) \to (Pr(\overline{\ulcorner S \urcorner}) \to Pr(\overline{\ulcorner T \urcorner}))$
3. $\vdash Pr(\overline{\ulcorner S \urcorner}) \to Pr(\overline{\ulcorner Pr(\overline{\ulcorner S \urcorner}) \urcorner})$

**LT**. If **PA** $\vdash Pr(\overline{\ulcorner S \urcorner}) \to S$, then **PA** $\vdash S$

[15] This section is adapted from [1], chapter 3.

We proceed by induction on the lengths of proofs. Let $S_1, \ldots, S_n = S$ be a proof in **GL** of $S$.

Base cases:

Let $n = 1$, so $S$ is an axiom of **GL**. The axioms A1-A3 are the same for **GL** and for **PA** (they are just the propositional axioms), so if $S$ is one such axiom, clearly for any realisation $*$, $\mathbf{PA} \vdash S^*$.

If $S = \Box(T \to U) \to (\Box T \to \Box U)$, then

$$\begin{aligned} S^* &= (\Box(T \to U) \to (\Box T \to \Box U))^* \\ &= \Box(T^* \to U^*) \to (\Box T^* \to \Box U^*) \\ &= Pr(\overline{\ulcorner T^* \to U^* \urcorner}) \to (Pr(\overline{\ulcorner T^* \urcorner}) \to Pr(\overline{\ulcorner U^* \urcorner})). \end{aligned}$$

This is just a case of derivability condition 2 above, and hence $\mathbf{PA} \vdash S^*$.

If $S = \Box(\Box T \to T) \to \Box T$, then $S^* = Pr(\overline{\ulcorner Pr(\overline{\ulcorner T^* \urcorner}) \to T^* \urcorner}) \to Pr(\overline{\ulcorner T^* \urcorner})$. In the following we will write $PT$ for $Pr(\overline{\ulcorner T^* \urcorner})$, and $U$ for $P(PT \to T)$ for readability and space:

$$\begin{aligned} \text{By condition 2, } &\vdash P(U \to PT) \to (PU \to PPT) && (1) \\ \text{and } &\vdash U \to (PPT \to PT) && (2) \\ \text{By condition 3, } &\vdash U \to PU && (3) \\ \text{By (2) and prop. logic, } &\vdash PPT \to (U \to PT) && (4) \\ \text{By (1) and prop. logic, } &\vdash PU \to (P(U \to PT) \to PPT) && (5) \\ \text{From (3) and (5), } &\vdash U \to (P(U \to PT) \to PPT) && (6) \\ \text{From (2) and (6), } &\vdash U \to ((P(U \to PT) \to PPT) \wedge (PPT \to PT)) && (7) \\ \text{So by prop. logic, } &\vdash U \to ((P(U \to PT) \to PT) && (8) \\ \text{and hence } &\vdash P(U \to PT) \to (U \to PT) && (9) \\ \text{Hence by Löb's Theorem } &\vdash U \to PT && (10) \end{aligned}$$

$U \to PT$ is just $S^*$.

Hence we have now shown that all axioms of **GL** are theorems of **PA** under any realisation.

Inductive cases:

Assume for all $1 \leq i < n$ for each realisation $*$, $\mathbf{PA} \vdash S_i^*$, and that $S$ follows from previous sentences by one of the deduction rules. If $S$ follows by MP from $S_i$ and $S_j = S_i \to S$, then $\mathbf{PA} \vdash S_i^*$ and $\mathbf{PA} \vdash S_j^* = S_i^* \to S^*$. Whence by MP in **PA**, $\mathbf{PA} \vdash S^*$.

If $S$ follows from $S_i$ by the Box Rule, then $S = \Box S_i$. If $\mathbf{PA} \vdash S_i^*$, then by derivability condition 1, $\mathbf{PA} \vdash Pr(\overline{\ulcorner S_i^* \urcorner}) = (\Box S)^*$.

Thus we have shown by induction that for any theorem $S$ of **GL** under any realisation $*$, $\mathbf{PA} \vdash A^*$. ■

The converse of this theorem is Solovay's Arithmetic Completeness Theorem, which we prove in the following section.

# 5 The Arithmetical Completeness Theorem

This section will be dedicated to proving that for any modal sentence $S$, if for every translation $*$, $\mathbf{PA} \vdash S^*$, then $\mathbf{GL} \vdash S$. This result was first proved by Robert Solovay[16], and is known as Solovay's Arithmetical Completeness Theorem. It was not until this theorem was proved that we knew that the addition of the arithmetisation of Löb's theorem to the modal system **K** gives a complete axiomatisation of the logic of provability in **PA**.[17]

## 5.1 Outlining the Proof

Let us now begin looking at how to prove Solovay's Theorem:

[16] [6].

[17] We follow [1], chapter 9, in our exposition.

We will prove this theorem by contraposition, so let us take $A$ to be a modal sentence such that $\mathbf{GL} \nvdash A$. We need to exhibit a realisation function $*$ such that $\mathbf{PA} \nvdash A^*$.

As $\mathbf{GL} \nvdash A$, we know by Theorem 2.29 that there exists a finite transitive, converse wellfounded model $\mathcal{M} = \langle W, R, I \rangle$ such that for some $w \in W$ $V_{\mathcal{M}}(A, w) = 0$, and for all $v \neq w$, $Rwv$. We may assume that $W = \{1, \ldots, n\}$ for some $n \in \mathbb{N}$, and that $w = 1$.

We extend $\mathcal{M}$ to a new model $\mathcal{M}' = \langle W', R', I' \rangle$ such that:

- $W' = W \cup \{0\}$
- $R' = R \cup \{\langle 0, i \rangle : 1 \leq i \leq n\}$
- $I'(p_j, i) = \begin{cases} I(p_j, i), \text{ if } 1 \leq i \leq n \\ I(p_j, 1), \text{ if } i = 0 \end{cases}$

$R'$ inherits the transitivity and converse wellfoundedness of $R$, and hence we know it is also irreflexive.

We will embed the model $\mathcal{M}'$ into **PA**. For each $i \in \{0, \ldots, n\}$, we will give a sentence of **PA**, $S_i$, which we will call the Solovay sentences, that are constructed such that they bear a relation to one another that is isomorphic to $R'$. Further, we will give a realisation function $*$ such that the following holds for all subsentences $B$ of $A$:

1. If $V_{\mathcal{M}}(B, i) = 1$, then $\mathbf{PA} \vdash S_i \to B^*$
2. If $V_{\mathcal{M}}(B, i) = 0$, then $\mathbf{PA} \vdash S_i \to \neg B^*$

So truth at a world in $\mathcal{M}$ has a corresponding notion in **PA** at each of the $S_i$. If such an embedding does truly reflect the mechanics of our modal semantics within **PA**, then as $V_{\mathcal{M}'}(\neg\Box A, 0) = 0$, it should follow that $\mathbf{PA} \vdash S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$, whence if $S_0$ is true and $S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$ is true, the result follows.

## 5.2 Constructing the Solovay Sentences

Our goal will be to give a function $f : \mathbb{N} \to \{0, \ldots, n\}$ and construct sentences $S_0, \ldots, S_n$ of **PA** such that $S_i$ states that the limit of $f$ is $i$.

**Definition 5.1.** *Suppose $\phi$ is some function whose domain is $\mathbb{N}$. The **limit** of $\phi$ exists and is equal to $a$ if for some $m \in \mathbb{N}$, $\phi(m) = a$ and for all $n > m$, $\phi(n) = a$. We denote the limit of $\phi$ as $lim(\phi)$.*

**Lemma 5.2.** *Suppose $Q$ is some transitive and irreflexive binary relation on $\{1, \ldots, n\}$, and $\phi : \mathbb{N} \to \{0, \ldots, n\}$ is such that if $\phi(m) = i$, then either $\phi(m+1) = i$ or $\phi(m+1) = j$ for some $j$ such that $Qij$. Then:*

1. *$lim(\phi)$ exists.*
2. *If for some $m \in \mathbb{N}$ $\phi(m) = i$, then either $lim(\phi) = i$ or $lim(\phi) = j$ for some $j$ such that $Qij$.*

*Proof.* 1. If $lim(\phi)$ does not exist, then for all $q \in \mathbb{N}$ there exists $m > q$ such that $\phi(m) \neq \phi(q)$. Let $a_0 = 0$ and let $a_{i+1}$ be the minimal $m > a_i$ such that $\phi(m) \neq \phi(a_i)$. By definition of $\phi$, for each $i$, $Q\phi(a_i)\phi(a_{i+1})$. As $\{1, \ldots, n\}$ is finite, there are $j$, $k$ such that $\phi(a_j) = \phi(a_k)$ and hence by transitivity of $Q$, $Q\phi(a_j)\phi(a_k)$. But this contradicts the irreflexivity of $Q$. Hence $lim(\phi)$ exists.
2. This follows from the transitivity of $Q$. ■

Suppose we are given a function $f : \mathbb{N} \to \{0, \ldots, n\}$. Let $F(a, b)$ be a formula of **PA** defining the binary relation $\{\langle a, b\rangle : f(a) = b\}$. Then we can define the Solovay sentences as:

$$S_i = \exists x \forall y (x \leq y \to \exists z (z = \bar{i} \wedge F(y, z)))$$

We define $f$ as follows:

f1. $f(0) = 0$

f2. If $f(m) = i$, then $f(m+1) = i$ unless $m$ is the Gödel number of the proof in **PA** of $\neg S_j$ for some $j$ such that $R'ij$, in which case $f(m+1) = j$.

(f2) implies that $f$ is such that if $f(m) = i$, then either $f(m+1) = i$ or $f(m+1) = j$ for some $j$ such that $Rij$. Thus by Lemma 5.2, $f$ has the additional conditions:

f3. $lim(f)$ exists.

f4. If for some $m \in \mathbb{N}$ $f(m) = i$, then either $lim(f) = i$ or $lim(f) = j$ for some $j$ such that $R'ij$.

**Definition 5.3.** *Let $notlim(x, y)$ be the $\Sigma_1$ pterm of a function whose value for each pair $\langle m, j\rangle$ is the Gödel number of the formula $\neg\exists x\forall y(x \leq y \rightarrow \exists z(z = \bar{j} \wedge G_m))$.*

Hence for some formula $H(y, z)$ such that $\ulcorner H(y, z)\urcorner = m$ defining a function $h : \mathbb{N} \rightarrow \mathbb{N}$, $notlim(\bar{m}, \bar{j})$ says that the $lim(h) \neq j$.

There is an apparent problem with the definitions of $f$ and the $S_i$'s that we have given above: condition (f2) imposed on $f$ tells us that $f$ is defined in terms of the Solovay Sentences, which in turn are defined in terms of the function $f$. We will see that this apparent circularity is *only* apparent, by appealing to the Generalised Diagonal Lemma.

(f1) and (f2) together tell us that if $f$ exists, the following holds: $f(a) = b$ if and only if there is a finite sequence, $s$, of length $a + 1$ such that $s_0 = 0$, $s_a = b$ and such that for all $x < a$, if $s_x = i$ and $\neg Proof(x, \overline{\ulcorner\neg S_j\urcorner})$ for some $j$ such that $R'ij$, then $s_{x+1} = j$ and $s_{x+1} = s_x$ otherwise.

We use this partial formalisation in the following lemma to prove the existence of $f$ and the $S_i$.

**Lemma 5.4.** *There exists some formula $F(a, b)$ of **PA** such that:*
$\boldsymbol{PA} \vdash F(a, b) \leftrightarrow \exists s(FinSeq(s) \wedge lh(s) = a + 1 \wedge s_0 = 0 \wedge s_a = b \wedge \forall x < a \bigwedge_{0\leq i\leq n}\{s_x = \bar{i} \rightarrow (\bigwedge_{j:R'ij}\{Proof(x, \overline{\ulcorner\neg S_j\urcorner}) \rightarrow s_{x+1} = \bar{j}\} \wedge (\bigwedge_{j:R'ij}\{\neg Proof(x, \overline{\ulcorner\neg S_j\urcorner})\} \rightarrow s_{x+1} = s_x))\})$

*Proof.* Let $\theta(y,a,b) := \exists s(FinSeq(s) \wedge lh(s) = a+1 \wedge s_0 = 0 \wedge s_a = b \wedge \forall x < a \bigwedge_{0 \leq i \leq n} \{s_x = \bar{i} \rightarrow (\bigwedge_{j:R'ij} \{Proof(x, notlim(y, \bar{j})) \rightarrow s_{x+1} = \bar{j}\} \wedge (\bigwedge_{j:R'ij} \{\neg Proof(x, notlim(y, \bar{j}))\} \rightarrow s_{x+1} = s_x))\})$.

By the Generalised Diagonal Lemma (Theorem 3.28), there exists a formula $\psi(a,b)$ in which all free variables are among $a$ and $b$ such that:

$(*)$ $\mathbf{PA} \vdash \psi(a,b) \leftrightarrow \theta(\overline{\ulcorner \psi(a,b) \urcorner}, a, b)$

Notice further that if we allow our $F(a,b)$ to be a formula satisfying $(*)$, then we have that $notlim(\overline{\ulcorner F(a,b) \urcorner}, \bar{j}) = \overline{\ulcorner \neg S_j \urcorner}$, which gives us the lemma. ■

Note that $F(a,b)$ is definitely $\Sigma_1$, as it is provably equivalent to a formula of the form $\exists s\, \psi$, where $\psi$ is $\Sigma_0$. We show in section 5.3.2 that $F(a,b)$ is $\Sigma_0$ in **PA**.

By the construction of the Solovay sentences we see there is a relation between them that exactly corresponds to $R'$: $R'ij$ if and only if the definition of $f$ allows that $f(m) = i$ then $f(m+1) = j$.

**Lemma 5.5.** *If* $0 \leq i < j \leq n$, *then* $\boldsymbol{PA} \vdash \neg(S_i \wedge S_j)$

*Proof.* We will prove by induction on the variable $a$ that $\mathbf{PA} \vdash \exists! b F(a,b)$.

Base case:

By Lemma 5.4, $\mathbf{PA} \vdash F(0,b) \leftrightarrow \exists s(FinSeq(s) \wedge lh(s) = \bar{1} \wedge s_0 = 0 \wedge s_0 = b)$. **PA** proves such a sequence exists and is unique, so $\mathbf{PA} \vdash \exists! b(\exists s(FinSeq(s) \wedge lh(s) = \bar{1} \wedge s_0 = 0 \wedge s_0 = b))$, and hence $\mathbf{PA} \vdash \exists! b F(0,b)$.

Inductive case:

The following argument can be formalised in **PA**. Assume for some $m \in \mathbb{N}$ that $\mathbf{PA} \vdash \exists! b F(\bar{m}, b)$, and let $k$ be the unique number satisfying $F(\bar{m}, \bar{k})$, and let $t$ be a finite sequence satisfying the right-hand side of the biconditional in Lemma 5.4, for the case where $a = \bar{m}$ and $b = \bar{k}$. For all such $t$, $t_m = k$. Then let $t'$ be the finite sequence of length $m+2$ extending $t$, where $t'_{m+1} = j$ if $m$ is the Gödel number of $S_j$ for some $j$ such that $R'kj$

and $t'_{m+1} = k = t_m$ otherwise. $t'_{m+1}$ is determined uniquely by the value of $t_m$, which in turn exists uniquely. Hence $\mathbf{PA} \vdash \exists! b F(\overline{m+1}, b)$, as **PA** proves the existence of the finite sequence $t'$.

So by induction on the variable $a$, $\mathbf{PA} \vdash \forall a \exists! b F(a, b)$ and furthermore for $i \neq j$, $\mathbf{PA} \vdash \forall a \exists! b F(a, b) \rightarrow \neg(S_i \wedge S_j)$. The result follows by MP. ■

**Definition 5.6.** *Given a finite transitive irreflexive frame $\mathcal{N} = \langle X, Q \rangle$ and $w \in X$, the **rank** $\rho_{\mathcal{N}}(w)$ as the greatest $n \in \mathbb{N}$ such that there exist $w_n = w, \ldots, w_0 \in X$ such that for all $0 \leq i < n$, $Qw_{n-i}w_{n-i-1}$ holds.*

This is well-defined for finite transitive irreflexive frames, as for all $i < j$, by transitivity $Qw_j w_i$ and by irreflexivity $w_j \neq w_i$, so there is always a greatest such $n \in \mathbb{N}$.

**Lemma 5.7.** *Given a finite transitive irreflexive frame $\mathcal{N} = \langle X, Q \rangle$, if $Qwv$ then $\rho_{\mathcal{N}}(w) > \rho_{\mathcal{N}}(v)$.*

*Proof.* If $\rho_{\mathcal{N}}(v) = n$, then there exist $v_n = v, \ldots, v_0 \in X$ such that for all $0 \leq i < n$ $Qv_{n-i}v_{n-i-1}$ holds, then for all $0 \leq i < n+1$ $Qv_{n+1-i}v_{n-i}$ where $v_{n+1} = w$, and hence $\rho_{\mathcal{N}}(w) > \rho_{\mathcal{N}}(v)$. ■

**Lemma 5.8.** $\mathbf{PA} \vdash F(a, \bar{i}) \rightarrow (S_i \vee \bigvee_{j:R'ij} \{S_j\})$

*Proof.* $\langle W', R' \rangle$ is a finite transitive irreflexive frame, so the rank $\rho(i)$ is well-defined. We proceed by induction on the rank of each $i \in \{0, \ldots, n\}$.

Base case:

If $\rho(i) = 0$, then $\{j : R'ij\}$ is empty. It follows from Lemma 5.4 and by induction in **PA** that $\mathbf{PA} \vdash \forall x(a \leq x \rightarrow F(x, \bar{i})$, and hence we can show that $\mathbf{PA} \vdash F(a, \bar{i}) \rightarrow S_i$, which as $\{j : R'ij\}$ is empty is the required result.

Inductive case:

Assume that for all $l < k$ that the lemma holds for $j$ such that $\rho(j) = l$, and let $i$ be such that $\rho(i) = k$. Then for all $j$ such that $R'ij$, by Lemma 5.7 $\rho(j) < \rho(i)$, and by assumption

(i) $\mathbf{PA} \vdash F(a, \bar{j}) \to (S_j \vee \bigvee_{m:R'jm}\{S_m\})$

By Lemma 5.4,

(ii) $\mathbf{PA} \vdash F(a, \bar{i}) \to \forall x(a \leq x \to (F(x, \bar{i}) \vee \bigvee_{j:R'ij}\{F(x, \bar{j})\}))$.

(i) and (ii) together imply that

$$\mathbf{PA} \vdash F(a, \bar{i}) \to (\forall x(a \leq x \to (F(x, \bar{i}) \vee \bigvee_{j:R'ij}\{S_j \vee \bigvee_{m:R'jm}\{S_m\}\}))$$

which can equivalently be written as

$$\mathbf{PA} \vdash F(a, \bar{i}) \to (\forall x(a \leq x \to F(x, \bar{i})) \vee \bigvee_{j:R'ij}\{S_j \vee \bigvee_{m:R'jm}\{S_m\}\})$$

In other words,

$$\mathbf{PA} \vdash F(a, \bar{i}) \to (S_i \vee \bigvee_{j:R'ij}\{S_j \vee \bigvee_{m:R'jm}\{S_m\}\})$$

Which by the transitivity of $R'$ is $F(a, \bar{i}) \to (S_i \vee \bigvee_{j:R'ij}\{S_j\})$, as required. ■

**Lemma 5.9.** $\boldsymbol{PA} \vdash (S_0 \vee \cdots \vee S_n)$

*Proof.* It follows from Lemma 5.8 and that $R'0i$ for all $0 < i \leq n$ that $\mathbf{PA} \vdash F(a, 0) \to (S_0 \vee \cdots \vee S_n)$. We also know that $\mathbf{PA} \vdash F(0, 0)$, so by modus ponens, $\mathbf{PA} \vdash (S_0 \vee \cdots \vee S_n)$. ■

**Lemma 5.10.** *If* $R'ij$, *then* $\boldsymbol{PA} \vdash S_i \to \neg Pr(\overline{\ulcorner \neg S_j \urcorner})$

*Proof.* First note that **PA** proves that any theorem has infinitely many proofs, as one can repeat the last formula of any such proof, and obtain a new one (with a different Gödel number) i.e $\mathbf{PA} \vdash Pr(\overline{\ulcorner \phi \urcorner}) \to \forall x \exists y(y > x \wedge Proof(y, \overline{\ulcorner \phi \urcorner}))$.

Suppose that $S_i$ holds, and for some $j$, $R'ij$. Then for some $m \in \mathbb{N}$, $\forall x(x > \bar{m} \to F(x, \bar{i}))$ holds. Assume for contradiction that $\mathbf{PA} \vdash \neg S_j$. Then we know $\mathbf{PA} \vdash Pr(\overline{\ulcorner \neg S_j \urcorner})$ and hence $\mathbf{PA} \vdash \forall x \exists y(y > x \wedge Proof(y, \overline{\ulcorner \neg S_j \urcorner}))$. In particular $\mathbf{PA} \vdash \exists y(y > \bar{m} \wedge Proof(y, \overline{\ulcorner \neg S_j \urcorner}))$. This implies for some $p > m$, $\mathbf{PA} \vdash F(\bar{p}, \bar{j})$, which contradicts[18] $\forall x(x > \bar{m} \to F(x, \bar{i}))$. So $\mathbf{PA} \nvdash \neg S_j$. We can formalise this in **PA** which gives $\mathbf{PA} \vdash S_i \to \neg Pr(\overline{\ulcorner \neg S_j \urcorner})$, as required. ■

[18] A consistent $\Sigma_0$-complete system is also $\Sigma_0$-sound, [5], p72, so this indeed follows.

**Lemma 5.11.** *If* $1 \leq i$, *then* $\boldsymbol{PA} \vdash S_i \to Pr(\overline{\ulcorner \neg S_i \urcorner})$

*Proof.* It follows from Lemma 5.4 that $\mathbf{PA} \vdash F(a, \bar{i}) \to \exists x Proof(x, \overline{\ulcorner \neg S_i \urcorner})$. Further, $\mathbf{PA} \vdash S_i \to \exists a F(a, \bar{i})$ and hence it follows that $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner \neg S_i \urcorner})$ ■

**Lemma 5.12.** *If* $1 \leq i$, *then* $\boldsymbol{PA} \vdash S_i \to Pr(\overline{\ulcorner \bigvee_{j:R'ij} \{S_j\} \urcorner})$

*Proof.*

$F(a, b)$ is $\Sigma_1$, so by Thm 3.25, $\mathbf{PA} \vdash \exists a F(a, \bar{i}) \to Pr(\overline{\ulcorner \exists a F(a, \bar{i}) \urcorner})$ (1)

$$\mathbf{PA} \vdash S_i \to \exists a F(a, \bar{i}) \tag{2}$$

By (1) and (2), $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner \exists a F(a, \bar{i}) \urcorner})$ (3)

By Lemma 5.11, as $1 \leq i$, $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner \neg S_i \urcorner})$ (4)

By Lemma 5.8, $\mathbf{PA} \vdash \exists a F(a, \bar{i}) \to (S_i \vee \bigvee_{j:R'ij} \{S_j\})$ (5)

By the derivability cond., $\mathbf{PA} \vdash Pr(\overline{\ulcorner \exists a F(a, \bar{i}) \urcorner}) \to Pr(\overline{\ulcorner (S_i \vee \bigvee_{j:R'ij} \{S_j\}) \urcorner})$ (6)

So, $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner (S_i \vee \bigvee_{j:R'ij} \{S_j\}) \urcorner})$ (7)

Further, $\mathbf{PA} \vdash (Pr(\overline{\ulcorner \neg S_i \urcorner}) \wedge Pr(\overline{\ulcorner (S_i \vee \bigvee_{j:R'ij} \{S_j\}) \urcorner})) \to Pr(\overline{\ulcorner \bigvee_{j:R'ij} \{S_j\} \urcorner})$ (8)

Whence, $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner \bigvee_{j:R'ij} \{S_j\} \urcorner})$ (9)

■

Define the realisation $*$ as follows:

- For each propositional variable, $p_k^* = \begin{cases} \bigvee \{S_i : I'(p_k, i) = 1\} & \text{if } \{S_i : I'(p_k, i) = 1\} \neq \emptyset \\ \bot & \text{otherwise.} \end{cases}$

This induces a translation over all modal sentences in the standard way as defined in section 4.

**Lemma 5.13.** *For all $i$ such that $1 \leq i \leq n$ and all subsentences $B$ of $A$, if $V_{\mathcal{M}}(B, i) = 1$ then $\boldsymbol{PA} \vdash S_i \to B^*$, and if $V_{\mathcal{M}}(B, i) = 0$, then $\boldsymbol{PA} \vdash S_i \to \neg B^*$*

*Proof.* We prove this by induction on the length of $B$. Let $1 \leq i \leq n$.

Base case:

If $B = \bot$, then $V_{\mathcal{M}}(\bot, i) = 0$ and $\mathbf{PA} \vdash S_i \to \neg \bot$.

Supppose $B = p_k$ for some $k \in \mathbb{N}$, and let $X = \{S_j : I'(p_k, j) = 1\}$.

If $V_{\mathcal{M}}(p_k, i) = 1$, then by definition of $\mathcal{M}'$, $I'(p_k, i) = 1$ and hence $S_i \in X$. So $\mathbf{PA} \vdash S_i \to \bigvee X$, i.e. $\mathbf{PA} \vdash S_i \to p_k^*$.

If $V_{\mathcal{M}}(p_k, i) = 0$, then by definition of $\mathcal{M}'$, $I'(p_k, i) = 0$ and hence $S_i \notin X$. If $X = \emptyset$ then $p_k^* = \bot$, and $\mathbf{PA} \vdash S_i \to \neg \bot$. If $X \neq \emptyset$, then for all $S_j \in X$, $\mathbf{PA} \vdash S_i \to \neg S_j$. So $\mathbf{PA} \vdash S_i \to \bigwedge\{\neg S_j : I'(p_k, j) = 1\}$; equivalently, $\mathbf{PA} \vdash S_i \to \neg \bigvee X$ (and $\bigvee X = p_k^*$).

Inductive case:

Assume that the lemma holds for all subsentences of some subsentence $B$ of $A$. Then we have three cases:

i. $B = \neg C$

ii. $B = C \to D$

iii. $B = \Box C$

For (i), if $V_{\mathcal{M}}(B, i) = 1$ then $V_{\mathcal{M}}(C, i) = 0$ and by the inductive hypothesis $\mathbf{PA} \vdash S_i \to \neg C^*$, so $\mathbf{PA} \vdash S_i \to B^*$. Similarly, if $V_{\mathcal{M}}(B, i) = 0$ then $V_{\mathcal{M}}(C, i) = 1$ and by the inductive hypothesis $\mathbf{PA} \vdash S_i \to C^*$, so $\mathbf{PA} \vdash S_i \to \neg B^*$.

For (ii), if $V_{\mathcal{M}}(C \to D, i) = 1$ then $V_{\mathcal{M}}(C, i) = 0$ or $V_{\mathcal{M}}(D, i) = 1$. Hence by the inductive hypothesis, either $\mathbf{PA} \vdash S_i \to \neg C^*$ or $\mathbf{PA} \vdash S_i \to D^*$. Hence $\mathbf{PA} \vdash (S_i \to \neg C^*) \vee (S_i \to D^*)$

and hence by definition of translation and by logic, $\mathbf{PA} \vdash S_i \to (C \to D)^*$. Similarly if $V_{\mathcal{M}}(C \to D, i) = 0$ we deduce by the inductive hypothesis that $\mathbf{PA} \vdash S_i \to C^*$ and $\mathbf{PA} \vdash S_i \to \neg D^*$, and it follows that $\mathbf{PA} \vdash S_i \to \neg(C \to D)^*$.

For (iii), if $V_{\mathcal{M}}(\Box C, i) = 1$ then for all $j$ such that $Rij$, $V_{\mathcal{M}}(C, j) = 1$, and by the inductive hypothesis for all such $j$, $\mathbf{PA} \vdash S_i \to C^*$. As $1 \leq i$, by definition of $\mathcal{M}'$ we have that $Rij$ if and only if $R'ij$, and hence $\mathbf{PA} \vdash \bigvee_{j:R'ij}\{S_j\} \to C^*$. By the derivability conditions and MP, we deduce that $\mathbf{PA} \vdash Pr(\overline{\ulcorner \bigvee_{j:R'ij}\{S_j\} \urcorner}) \to Pr(\overline{\ulcorner C^* \urcorner})$. By definition, $Pr(\overline{\ulcorner C^* \urcorner}) = B^*$, By Lemma 5.12, $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner \bigvee_{j:R'ij}\{S_j\} \urcorner})$, so as $\Pr C^* = B^*$ it follows that $\mathbf{PA} \vdash S_i \to B^*$.

Now if $V_{\mathcal{M}}(\Box C, i) = 0$, then there exists $j$ such that $Rij$ and $V_{\mathcal{M}}(C, j) = 0$. By the inductive hypothesis, $\mathbf{PA} \vdash S_j \to \neg C^*$, and hence by logic and the derivability conditions $\mathbf{PA} \vdash \neg Pr(\overline{\ulcorner \neg S_j \urcorner}) \to \neg Pr(\overline{\ulcorner C^* \urcorner})$. As $i > 0$ and $Rij$, also $R'ij$, so by Lemma 5.10, $\mathbf{PA} \vdash S_i \to \neg Pr(\overline{\ulcorner \neg S_j \urcorner})$. It follows that $\mathbf{PA} \vdash S_i \to \neg Pr(\overline{\ulcorner C^* \urcorner})$, as required.

Hence we have proved the lemma by induction. ■

**Lemma 5.14.** $\boldsymbol{PA} \vdash S_1 \to \neg A^*$

*Proof.* By assumption, $V_{\mathcal{M}}(A, 1) = 0$ and hence by Lemma 5.13, as $A$ is a subsentence of itself, $\mathbf{PA} \vdash S_1 \to \neg A^*$. ■

**Lemma 5.15.** $\boldsymbol{PA} \vdash S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$

*Proof.* $\mathbf{PA} \vdash S_1 \to \neg A^*$, hence by contraposition, $\mathbf{PA} \vdash A^* \to \neg S_1$. By the derivability conditions and modus ponens, $\mathbf{PA} \vdash Pr(\overline{\ulcorner A^* \urcorner}) \to Pr(\overline{\ulcorner \neg S_1 \urcorner})$. Again by contraposition, $\mathbf{PA} \vdash \neg Pr(\overline{\ulcorner \neg S_1 \urcorner}) \to \neg Pr(\overline{\ulcorner A^* \urcorner})$, and further by Lemma 5.10, $\mathbf{PA} \vdash S_0 \to \neg Pr(\overline{\ulcorner \neg S_1 \urcorner})$, and therefore by logic $\mathbf{PA} \vdash S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$. ■

## 5.3 The Main Result

Finally, with all these Lemmas, we can prove Solovay's Theorem. This last step cannot be formalised in **PA**. Recall that for the result we require that $\mathbf{PA} \nvdash A^*$. If we could formalise

this within **PA**, then $\mathbf{PA} \vdash \neg Pr(\overline{\ulcorner A^* \urcorner})$. This implies that Peano Arithmetic can prove its own consistency. However, we know by Gödel's Second Incompleteness Theorem[19] that this cannot be the case, if we assume that **PA** is a consistent theory (which we do).

A further assumption we must make about **PA** is for the final step of the proof is that for all the Solovay sentences or their negations, if they are provable, then they are true, and that for each $i > 0$, $S_i \to Pr(\overline{\ulcorner \neg S_i \urcorner})$, and $S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$ are true. We stated above that the formula $F(x, y)$ can be given such that it is $\Sigma_0$, and as each $S_i$ is provably equivalent to $\exists x \forall y(x \leq y \to F(y, \bar{i}))$, each $S_i$ is $\Sigma_2$ and $\neg S_i$ is $\Pi_2$. Also each $S_i \to Pr(\overline{\ulcorner \neg S_i \urcorner})$ and $S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$ are $\Sigma_2$. So a condition on **PA** whose assumption entails this is $\Sigma_2$- and $\Pi_2$-soundness. $\Sigma_2$-soundness entails $\Pi_2$-soundness, so we need only assume $\Sigma_2$-soundness.

**Theorem 5.16** (Arithmetical Completeness Theorem for GL)**.** *For a modal sentence $S$, if for every translation $*$, $\boldsymbol{PA} \vdash S^*$, then $\boldsymbol{GL} \vdash S$.*

*Proof.* Recall we require for our fixed modal sentence $A$ that $\mathbf{PA} \nvdash A^*$. For $1 \leq i$, by Lemma 5.11, $\mathbf{PA} \vdash S_i \to Pr(\overline{\ulcorner \neg S_i \urcorner})$. Assume for contradiction that $S_i$ is true. Then $Pr(\overline{\ulcorner \neg S_i \urcorner})$ is true and hence $\mathbf{PA} \vdash \neg S_i$. So $\neg S_i$ must be true; a contradiction. Thus $S_i$ is false.

By Lemma 5.9, $\mathbf{PA} \vdash S_0 \vee \cdots \vee S_n$ and hence at least one of the $S_j$ is true. We can deduce that $S_0$ must be true. By Lemma 5.15, $\mathbf{PA} \vdash S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$, so $S_0 \to \neg Pr(\overline{\ulcorner A^* \urcorner})$ is true, and thus by MP, $\neg Pr(\overline{\ulcorner A^* \urcorner})$ is true. Which is to say **PA** doesn't prove $A^*$; in other words, $\mathbf{PA} \nvdash A^*$, as required. ■

An interesting point to be made about this final step of reasoning is that as $S_0$ is true, the limit of the function $f$ is 0, and in fact $f$ is a constant function: for all $n \in \mathbb{N}$, $f(n) = 0$. But if we were able to prove this fact in **PA**, then it follows that **PA** is an inconsistent system. It is remarkable that such a simple proposition as the constancy of a particular function would result in **PA** being inconsistent.

[19] Gödel's Second Incompleteness Theorem says that if a formal system includes a statement of its own consistency (relative to its own provability predicate), then that system is inconsistent

We can give an alternative final step of the proof[20] that reduces the requirement of $\Sigma_2$-soundness to $\Sigma_1$-soundness:

**Lemma 5.17.** $\boldsymbol{PA} \vdash \bigwedge_{1\leq i\leq n}\{Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i\} \rightarrow \neg Pr(\overline{\ulcorner A^* \urcorner})$

*Proof.*

$$\text{By Lemma 5.11, for each } i > 0 \vdash S_i \rightarrow (Pr(\overline{\ulcorner \neg S_i \urcorner}) \wedge S_i) \tag{1}$$

$$\text{Equivalently, } \vdash S_i \rightarrow \neg(Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i) \tag{2}$$

$$\text{And so for each } i > 0 \vdash (Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i) \rightarrow \neg S_i \tag{3}$$

$$\text{Hence, } \vdash \bigwedge_{1\leq i\leq n} \{Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i\} \rightarrow \bigwedge_{1\leq i\leq n} \{\neg S_i\} \tag{4}$$

$$\text{By Lemma 5.9, } \vdash S_0 \vee \cdots \vee S_n \tag{5}$$

$$\text{(4) and (5) together imply } \vdash \bigwedge_{1\leq i\leq n} \{Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i\} \rightarrow S_0 \tag{6}$$

$$\text{By Lemma 5.15, } \vdash S_0 \rightarrow \neg Pr(\overline{\ulcorner A^* \urcorner}) \tag{7}$$

$$\text{So by (6) and (7), } \vdash \bigwedge_{1\leq i\leq n} \{Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i\} \rightarrow \neg Pr(\overline{\ulcorner A^* \urcorner}) \tag{8}$$

■

If each $S_i$ is $\Sigma_2$, the formula in Lemma 5.17 can be shown to be $\Pi_2$. Now assume **PA** is $\Pi_2$-sound, and for contradiction that $\Phi = \bigwedge_{1\leq i\leq n}\{Pr(\overline{\ulcorner \neg S_i \urcorner}) \rightarrow \neg S_i\}$ is false. Then for each $i$, $Pr(\overline{\ulcorner \neg S_i \urcorner})$ is true, and $\neg S_i$ is false, so **PA** proves a false $\Pi_2$ formula; contradiction. Hence $\Phi$ is true. $\Phi \rightarrow \neg Pr(\overline{\ulcorner A^* \urcorner})$ is $\Pi_2$ and hence true, and so $Pr(\overline{\ulcorner A^* \urcorner})$ is true, as required. It can be easily shown that $\Sigma_1$-soundness entails $\Pi_2$ soundness. So we can require that **PA** be $\Sigma_1$-sound for Solovay's Theorem to hold[21].

[20] This is stated, but not proved, in [1], p130-131.

[21] In $\Sigma_0$-complete systems, $\Sigma_1$-soundness is equivalent to 1-consistency. This is the condition under which Gödel's Incompleteness Theorems can be proven for **PA** too.

### 5.3.1 Some Constructive Examples

How would this proof work for a particular example? We know that for $\Sigma_1$ formulae, $\mathbf{PA} \vdash \phi \to Pr(\overline{\ulcorner \phi \urcorner})$, but in general this is not the case. If we consider the modal sentence $A = p_0 \to \Box p_0$ and follow the proof of Solovay's Theorem, we will be able to construct a formula of **PA** that does not satisfy this provability result.

Let $\mathcal{M} = \langle W, R, I \rangle$, where $W = \{1, 2\}$, $R = \{\langle 1, 2 \rangle\}$, and $I$ is given by $I(p_0, 1) = 1$, $I(p_0, 2) = 0$, and all other pairs are assigned 0. $W$ is finite and $R$ is transitive and irreflexive, so this is an appropriate model for **GL**. We can see that for the valuation $V$ induced by $I$, $V(p_0 \to \Box p_0, 1) = 0$, as $V(p_0, 1) = 1$ and for some $i \in W$ such that $R1i$ (i.e. just 2), $V(p_0, i) = 0$, so $V(\Box p_0, 1) = 0$.

We will extend this model to a new one $\mathcal{M}' = \langle W', R', I' \rangle$, where $W' = \{0, 1, 2\}$, $R' = \{\langle 0, 1 \rangle, \langle 0, 2 \rangle, \langle 1, 2 \rangle\}$ and $I'$ extends $I$ by letting $I'(p_i, 0) = I(p_i, 1)$ for each $i \in \mathbb{N}$. This too is a **GL** counter model for $A$, where under the valuation $V'$ induced by $I'$, $V'(A, 0) = 0$.

We know by Lemma 5.4 that there exists $f : \mathbb{N} \to \{0, 1, 2\}$ and Solovay sentences $S_0$, $S_1$ and $S_2$. In the proof of Solovay's Theorem, we gave the realisation $*$ by $p_k^* = \bigvee_{i:I'(p_k,i)=1} \{S_i\}$, so in particular $p_0^* = S_0 \vee S_1$. The subsentences of $A$ are $p_0$, $\Box p_0$ and $A = p_0 \to \Box p_0$. The respective translation of these are: $S_0 \vee S_1$, $Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$, and $(S_0 \vee S_1) \to Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$. We can see that Lemma 5.13 holds for each of these as we know limits are provably unique:

$p_0$ is the easy case:

$V'(p_0, 1) = 1$, and $\mathbf{PA} \vdash S_1 \to (S_0 \vee S_1)$.

$V'(p_0, 2) = 0$ and $\mathbf{PA} \vdash S_2 \to \neg(S_0 \vee S_1)$.

$\Box p_0$ is a little more complex:

As $V'(\Box p_0, 1) = 0$, we want to show that $\mathbf{PA} \vdash S_1 \to \neg Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$. We know that $\mathbf{PA} \vdash (S_0 \vee S_1) \to \neg S_2$, whence $\mathbf{PA} \vdash \neg Pr(\overline{\ulcorner \neg S_2 \urcorner}) \to \neg Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$. But we know by Lemma 5.10 that $\mathbf{PA} \vdash S_1 \to \neg Pr(\overline{\ulcorner \neg S_2 \urcorner})$, so we have by logic that $\mathbf{PA} \vdash S_1 \to \neg Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$. As $V'(\Box p_0, 2) = 1$ (vacuously) we want to show that $\mathbf{PA} \vdash S_2 \to Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$. We know

that $\mathbf{PA} \vdash \neg S_2 \rightarrow (S_0 \vee S_1)$ and hence that $\mathbf{PA} \vdash Pr(\overline{\ulcorner \neg S_2 \urcorner}) \rightarrow Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$, and by Lemma 5.11, $\mathbf{PA} \vdash S_2 \rightarrow Pr(\overline{\ulcorner \neg S_2 \urcorner})$, so the result follows.

Finally for $A = p_0 \rightarrow \Box p_0$:

$V'(p_0 \rightarrow \Box p_0, 2) = 1$ is easy: we know from above that $\mathbf{PA} \vdash S_2 \rightarrow Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$, whence by propositional logic, $\mathbf{PA} \vdash S_2 \rightarrow ((S_0 \vee S_1) \rightarrow Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner}))$.

As $V'(p_0 \rightarrow \Box p_0, 1) = 0$ we here want to show that $\mathbf{PA} \vdash S_1 \rightarrow \neg((S_0 \vee S_1) \rightarrow Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner}))$. Equivalently we want to show that $\mathbf{PA} \vdash S_1 \rightarrow ((S_0 \vee S_1) \wedge \neg Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner}))$. By above, $\mathbf{PA} \vdash S_1 \rightarrow \neg Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$ and $\mathbf{PA} \vdash S_1 \rightarrow (S_0 \vee S_1)$, so by logic $\mathbf{PA} \vdash S_1 \rightarrow ((S_0 \vee S_1) \wedge \neg Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner}))$.

In particular, we do indeed have that $\mathbf{PA} \vdash S_1 \rightarrow \neg A^*$ and whence we can deduce (as in Lemma 5.15) that $\mathbf{PA} \vdash S_0 \rightarrow \neg Pr(\overline{\ulcorner A^* \urcorner})$. Following the final steps of the proof as laid out in Theorem 5.16, we get that $\mathbf{PA} \nvdash (S_0 \vee S_1) \rightarrow Pr(\overline{\ulcorner S_0 \vee S_1 \urcorner})$. However, $A^*$ is not true: $S_0$ is true, so the antecedent is true, but (under the assumption that **PA** is consistent) $\mathbf{PA} \nvdash S_0$, and so if $A^*$ is true we must have that $\mathbf{PA} \vdash S_1$, which would imply that **PA** proves something false, again contradicting consistency.

We know that $\mathbf{PA} \nvdash S_0$, and under our realisation function, there is no $p_k$ such that $p_k^* = S_0$, as there is no propositional variable that is uniquely true at $0 \in W^*$. If we altered our evaluation function $I'$ so that $I'(p_1, 0) = 1$ and hence $p_1^* = S_0$, then we know of course that $\mathbf{GL} \nvdash p_1$, so this is consistent with Theorem 4.3, the Arithmetical Soundness Theorem of **GL**.

Can this method construct a true but unprovable formula? $\Box p_0 \rightarrow p_0$ has countermodel $\mathcal{M} = \langle W = \{1\}, R = \emptyset, I \rangle$ where $I$ is constantly 0, so we have that $V(\Box p_0 \rightarrow p_0, 1) = 0$. We extend this in the usual way to countermodel $\mathcal{M}' = \langle W' = \{0, 1\}, R' = \{\langle 0, 1 \rangle\}, I' \rangle$ where $I'$ is constantly 0. Then the translation the proof of Solovay's Theorem specifies us gives $p_0^* = \bot$, and it follows that $\mathbf{PA} \nvdash Pr(\overline{\ulcorner \bot \urcorner}) \rightarrow \bot$ which can be written as $\mathbf{PA} \nvdash \neg Pr(\overline{\ulcorner \bot \urcorner})$. This is just to say that **PA** cannot prove its own consistency; i.e. Gödel's Second Incompleteness Theorem.

### 5.3.2 Extending Solovay's Theorem

In this subsection we will pinpoint those non-logical properties of **PA** that were used to proof Solovay's Theorem, thus giving an upper bound for the minimal set of properties that any formal system must have in order to prove the result.

**Definition 5.18.** *Robinson Arithmetic, **Q**, is equivalent to **PA** without the induction schema, and with the additional axiom* $x = 0 \vee \exists y\, y = \boldsymbol{s}x$.

In [3], we find the following result:

**Theorem 5.19.** *Solovay's Arithmetical Completeness Theorem can be proved in any extension of **Q**, **T**, with the induction schema for* $\Sigma_0$ *formulae, and that proves* $\Sigma_1$*-completeness for its respective provability predicate: for all* $\Sigma_1$ *formulae,* $\phi$, $\boldsymbol{T} \vdash \phi \rightarrow Pr_{\boldsymbol{T}}(\overline{\ulcorner \phi \urcorner})$.

The following is my own argument for this result. Let us take **T** to be any formal system extending **Q** that can prove Solovay's Theorem. There are three lemmas to the proof of Theorem 5.16 in which we exploited non-logical properties of **PA**; Lemmas 5.4, 5.8 and 5.12. The rest follow by logic. We hence will see what **T** needs to prove these Lemmas.

Firstly, in Lemma 5.4, the Generalised Diagonal Lemma is used to prove that the function $f$ and the Solovay sentences are well-defined. The $\beta$ -function can be represented in **Q**, so in **T**, $FinSeq(x)$ can be defined and further a corresponding formula $Proof_{\mathbf{T}}(x, y)$ that says $x$ is a (**T**) proof of $y$ can be given. As in **PA**, this formula can be given such that it is $\Sigma_0$. Further we need to be able to define $notlim(x, y)$. The Diagonal Lemma can be proven in any formal system in which all primitive recursive functions can be represented, which is even the case in **Q**.

Secondly, in Lemmas 5.5 and 5.8, we used induction in **PA** on the formulae $\exists! b F(a, b)$ and $a \leq x \rightarrow F(x, \bar{i})$, on the variables $a$ and $x$ respectively. We will show in what circumstances these can be replaced with $\Sigma_0$ formulae.

In **T**, $F(x, \bar{i})$ is provably equivalent to the following:

$$\exists s(FinSeq(s) \wedge lh(s) = x+1 \wedge s_0 = 0 \wedge s_x = \bar{i} \wedge \forall y < x \bigwedge_{0 \leq j \leq n} \{s_y = \bar{j} \rightarrow$$
$$(\bigwedge_{k:R'jk} \{Proof_{\mathbf{T}}(y, \overline{\ulcorner \neg S_k \urcorner}) \rightarrow s_{y+1} = \bar{k}\} \wedge (\bigwedge_{k:R'jk} \{\neg Proof_{\mathbf{T}}(y, \overline{\ulcorner \neg S_k \urcorner})\} \rightarrow s_{y+1} = s_y))\})$$

Which is of the form $\exists s\, \theta(s, x, \bar{i})$. $FinSeq(s)$ is $\Sigma_0$, $lh(s)$, $s_0, \ldots, s_x$ are all $\Sigma_1$ pterms, and $Proof_{\mathbf{T}}(x, \overline{\ulcorner \neg (S_j) \urcorner})$ and its negation are $\Sigma_0$. Hence $F(x, \bar{i})$ is at the very least $\Sigma_1$, unless we can put a bound on its principle existential quantifier. As in the $\beta$-function lemma (Lemma 3.18) take $t = max[x+1, s_0, \ldots, s_{x+1}]$. The $s_i$ are bounded above by $n$, so take $w = max[x+1, n]$. Then, if $s = \langle\langle a, b\rangle, x+1\rangle$, $a$ and $b$ can be chosen such that $b < lcm[m+1 : m < t] + 1 \leq lcm[m+1 : m < w] \leq max[(x+1)^{x+1}, n^n] = q$, and $a < lcm[1+(m+1)b : m < x+1] \leq [(x+1)b]^{x+1} = p$. So we know that $s < \langle\langle p, q\rangle, x+2\rangle$. In order for this to be well-defined, we require that $\mathbf{T}$ proves the exponential function $z = x^y$ is total; $T \vdash \forall x \forall y \exists z z = x^y$. If this is the case, then $F(x, \bar{i})$ is provably equivalent to $\exists s(s < \langle\langle \bar{p}, \bar{q}\rangle, x+2\rangle \wedge \theta(s, x, \bar{i}))$, and thus $\Sigma_0$.

The formula $\exists! b F(a, b)$ can be replaced with $\exists b \leq n(F(a, b) \wedge \forall z \leq n(F(a, z) \rightarrow z = b))$ (which is $\Sigma_0$ if $F(a, b)$ is) and the result of Lemma 5.5 still follows. Also, as $a \leq x$ is $\Sigma_0$, both instances of induction in the proof of Solovay's Theorem are $\Sigma_0$-induction if the system has exponentiation.

Lemma 5.12 uses provable $\Sigma_1$-completeness (Theorem 3.25). So $\mathbf{T}$ must be such that for all $\Sigma_1$ formulae $\phi$, $T \vdash \phi \rightarrow \mathrm{Pr}_{\mathrm{T}}(\overline{\ulcorner \phi \urcorner})$.

So any formal system having these properties is strong enough to prove the Arithmetical Completeness Theorem.

# 6 Concluding Remarks

In this dissertation, we have examined the elements that are required to proof Solovay's Arithmetical Completeness Theorem. Also, we have illuminated these proofs by means of constructive examples and justified the intuition behind each step. We also saw that Solovay's

Theorem holds in far weaker systems of arithmetic than **PA**, and also saw what aspects of **PA** were important in proving the result.

We have learned that **GL** is the provability logic of **PA**, and thus highlighted the importance of Löb's Theorem in the study of our formal system of arithmetic. This tells that that the study of a large and interesting portion of **PA** can be reduced to the study of a decidable modal deductive system, and is just decidable itself. Further, **GL**'s semantic theory is neat and easier to visualise, and often countermodels to sentences can just be "seen".

# References


[1] George Boolos, *The Logic of Provability*, 1993: Cambridge University Press.

[2] Leon Henkin, *"A problem concerning provability"*, Journal of Symbolic Logic, 17 (1952), 160.

[3] Dick De Jongh, Marc Jumelet, Franco Montagna, *"On the Proof of Solovay's Theorem"*, Studia Logica, Vol. 50, No. 1 (1995), 51-69.

[4] Martin H. Löb, *"Solution of a problem of Leon Henkin"*, Journal of Symbolic Logic, 20 (1955), 115-118.

[5] Raymond M. Smulyan, *Gödel's Incompleteness Theorems*, 1992: Oxford University Press.

[6] Robert Solovay, *"Provability Interpretations of Modal Logic"*, Israel Journal of Mathematics 25 (1976), 287-304.

[7] *The On-Line Encyclopedia of Integer Sequences*, published electronically at https://oeis.org/A000112