Self-referential theories

Samuel A. Alexander*
Department of Mathematics, the Ohio State University

January 19, 2020

Abstract

We study the structure of families of theories in the language of arithmetic extended to allow these families to refer to one another and to themselves. If a theory contains schemata expressing its own truth and expressing a specific Turing index for itself, and contains some other mild axioms, then that theory is untrue. We exhibit some families of true self-referential theories that barely avoid this forbidden pattern.

1 Introduction

This is a paper about families of r.e. theories, each capable of referring to itself and the others. Many of this paper's results first appeared in the author's dissertation [1]. There, they were stated in terms of families of interacting mechanical knowing agents. Here, we will speak instead of families of self-referential r.e. theories. We hope this will more directly expose the underlying mathematics.

In epistemology, it is well-known that a (suitably idealized) truthful knowing machine capable of arithmetic, logic, and self-reflection, cannot know its own truth and its own code. This is due, in various guises, to authors such as Lucas [8], Benacerraf [3], Reinhardt [11], Penrose [9], and Putnam [10]. In terms of self-referential theories, a true theory satisfying certain assumptions cannot contain schemata stating its own truth and its own Gödel number (if such a theory did exist, we could program a machine knower that knows precisely its consequences). Reinhardt conjectured, and Carlson proved [5], a truthful machine knower can know (in a local sense, i.e., expressed by infinite schemata rather than a single axiom) that it is truthful and has some code, without knowing which. A true self-referential theory can (in a local sense) state its own truth and recursive enumerability. We showed [2] that, alternatively, a truthful machine can (in a local sense) exactly know its own code, if not required to know its own truth. A true theory can state (in a local sense) its own Gödel number.

Our goal is to generalize the above consistency results to multiple theories. The paper contains four main findings. In the following list of promises, except where otherwise stated, $\prec$ is an r.e. well-founded partial-order on ω, and *expresses* is meant in the local (infinite schema) sense.

1. There are true theories $(T_i)_{i\in\omega}$ such that T_i expresses a Gödel number of T_j (all i, j) and T_i expresses the truth of T_j (all $j \prec i$).
2. There are true theories $(T_i)_{i\in\omega}$ such that T_i expresses a Gödel number of T_j ($j \prec i$), the truth of T_j ($j \preceq i$), and the fact that T_j has some Gödel number (all i, j).
3. If $\prec$ is ill-founded, and if we extend the base language to include a predicate for computable ordinals and require the theories to include rudimentary facts about them, then 1 and 2 fail.
4. Finally, if we do not extend the base language as in 3, then there do exist ill-founded r.e. partial orders $\prec$ such that 1 and 2 hold.

Our proofs of 1 and 2 are constructive, but the proof of 4 is nonconstructive. In short, if 4 were false, either of 1 or 2 could be used to define the set WF of r.e. well-founded partial orders of ω using nothing but arithmetic and a truth predicate Tr for arithmetic. This is impossible since WF is Π^1_1-complete and Tr is Δ^1_1.

*Email: alexander@math.ohio-state.edu

2 Preliminaries

To us, *theory* and *schema* mean *set of sentences* (a *sentence* is a formula with no free variables).

Definition 1. (Standard Definitions)

1. When a first-order structure is clear from context, an *assignment* is a function s mapping first-order variables into the universe of that structure. If x is a variable and u is an element of the universe, $s(x|u)$ is the assignment that agrees with s except that it maps x to u.
2. We write $\mathscr{M} \models \phi[s]$ to indicate that the first-order structure $\mathscr{M}$ satisfies the formula ϕ relative to the assignment s. We write $\mathscr{M} \models \phi$ just in case $\mathscr{M} \models \phi[s]$ for every assignment s. If T is a theory, $\mathscr{M} \models T$ means that $\mathscr{M} \models \phi$ for every $\phi \in T$.
3. We write $\mathrm{FV}(\phi)$ for the set of free variables of ϕ.
4. We write $\phi(x|t)$ for the result of substituting term t for variable x in ϕ.
5. $\mathscr{L}_{\mathrm{PA}}$ is the language of Peano arithmetic, with constant symbol 0 and function symbols S, $+$, $\cdot$ with the usual arities. If $\mathscr{L}$ extends $\mathscr{L}_{\mathrm{PA}}$, an $\mathscr{L}$-structure *has standard first-order part* if it has universe $\mathbb{N}$ and interprets 0, S, $+$ and $\cdot$ as intended.
6. We define $\mathscr{L}_{\mathrm{PA}}$-terms $\overline{n}$ ($n \in \mathbb{N}$), called *numerals*, so that $\overline{0} \equiv 0$ and $\overline{n+1} \equiv S(\overline{n})$.
7. We fix a computable bijection $\langle \bullet, \bullet, \bullet \rangle : \mathbb{N}^3 \to \mathbb{N}$. Being computable, this is $\mathscr{L}_{\mathrm{PA}}$-definable, so we may freely act as if $\mathscr{L}_{\mathrm{PA}}$ contained a function symbol for this bijection. Similarly we may act as if $\mathscr{L}_{\mathrm{PA}}$ contained a binary predicate symbol $\bullet \in W_\bullet$ for membership in the nth r.e. set W_n.
8. Whenever a computable language is clear from context, $\phi \mapsto \ulcorner \phi \urcorner$ denotes Gödel numbering.
9. A *valid* formula is one that is true in every structure.
10. A *universal closure* of ϕ is a sentence $\forall x_1 \cdots \forall x_n \phi$ where $\mathrm{FV}(\phi) \subseteq \{x_1, \ldots, x_n\}$. We write $\mathrm{ucl}(\phi)$ to denote a generic universal closure of ϕ.

Note that if $\mathscr{M}$ is a structure and ψ is a universal closure of ϕ, in order to prove $\mathscr{M} \models \psi$ it suffices to let s be an arbitrary assignment and show $\mathscr{M} \models \phi[s]$.

To formalize self-referential theories, we employ an extension of first-order logic where languages may contain new unary connective symbols. This logic is borrowed from [5].

Definition 2. (The Base Logic) A language $\mathscr{L}$ of the *base logic* is a first-order language $\mathscr{L}_0$ together with a class of symbols called *operators*. Formulas of $\mathscr{L}$ are defined as usual, with the clause that $\mathbf{T}_i\vDash\phi$ is a formula whenever ϕ is a formula and $\mathbf{T}_i\vDash$ is an operator. Syntactic parts of Definition 1 extend to the base logic in obvious ways (we define $\mathrm{FV}(\mathbf{T}_i\vDash\phi) = \mathrm{FV}(\phi)$). An $\mathscr{L}$-structure $\mathscr{M}$ is a first-order $\mathscr{L}_0$-structure $\mathscr{M}_0$ together with a function that takes one operator $\mathbf{T}_i\vDash$, one $\mathscr{L}$-formula ϕ, and one assignment s, and outputs either True or False—in which case we write $\mathscr{M} \models \mathbf{T}_i\vDash\phi[s]$ or $\mathscr{M} \not\models \mathbf{T}_i\vDash\phi[s]$, respectively—satisfying the following three requirements.

1. Whether or not $\mathscr{M} \models \mathbf{T}_i\vDash\phi[s]$ does not depend on $s(x)$ if $x \notin \mathrm{FV}(\phi)$.
2. If ϕ and ψ are *alphabetic variants* (meaning that one is obtained from the other by renaming bound variables so as to respect the binding of the quantifiers), then $\mathscr{M} \models \mathbf{T}_i\vDash\phi[s]$ if and only if $\mathscr{M} \models \mathbf{T}_i\vDash\psi[s]$.
3. For variables x and y such that y is substitutable for x in $\mathbf{T}_i\vDash\phi$, $\mathscr{M} \models \mathbf{T}_i\vDash\phi(x|y)[s]$ if and only if $\mathscr{M} \models \mathbf{T}_i\vDash\phi[s(x|s(y))]$.

The definition of $\mathscr{M} \models \phi[s]$ for arbitrary $\mathscr{L}$-formulas is obtained from this by induction. Semantic parts of Definition 1 extend to the base logic in obvious ways.

Traditionally the operator $\mathbf{T}_i\vDash$ would be written K_i, and the formula $K_i\phi$ would be read like "agent i knows ϕ". For the present paper, the added intuition would not be worth the philosophical distraction.

Theorem 3. (Completeness and compactness) Suppose $\mathscr{L}$ is an r.e. language in the base logic.

1. The set of valid $\mathscr{L}$-formulas is r.e.
2. For any r.e. $\mathscr{L}$-theory Σ, $\{\phi : \Sigma \models \phi\}$ is r.e.
3. There is an effective procedure, given (a Gödel number of) an r.e. $\mathscr{L}$-theory Σ, to find (a Gödel number of) $\{\phi : \Sigma \models \phi\}$.
4. If Σ is an $\mathscr{L}$-theory and $\Sigma \models \phi$, there are $\sigma_1, \ldots, \sigma_n \in \Sigma$ such that[1] $\sigma_1 \to \cdots \to \sigma_n \to \phi$ is valid.

Proof. By interpreting the base logic within first-order logic (for details see [1]). □

Definition 4. If $\mathscr{L}$ is a first-order language and I is an index set, let $\mathscr{L}(I)$ be the language (in the base logic) consisting of $\mathscr{L}$ along with operators $\mathbf{T}_i\vDash$ for all $i \in I$.

In case I is a singleton, $\mathscr{L}_{\mathrm{PA}}(I)$ is a form of Shapiro's [12] language of Epistemic Arithmetic.

Definition 5.

- For any $\mathscr{L}_{\mathrm{PA}}(I)$-formula ϕ with $\mathrm{FV}(\phi) = \{x_1, \ldots, x_n\}$, and for assignment s (into $\mathbb{N}$), let ϕ^s be the sentence
$$\phi^s \equiv \phi(x_1|\overline{s(x_1)}) \cdots (x_n|\overline{s(x_n)})$$
obtained by replacing all free variables in ϕ by numerals for their s-values.
- For any language $\mathscr{L}$ extending $\mathscr{L}_{\mathrm{PA}}$, if $\mathscr{M}$ is an $\mathscr{L}$-structure, then $\mathscr{M}$ is said to *interpret formulas by substitution* if $\mathscr{M}$ has standard first-order part and the following property holds: for every $\mathscr{L}$-formula ϕ and assignment s, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^s$.

For example, if $s(x) = 0$ and $s(y) = 2$ then $(\forall z(x = y + z))^s \equiv \forall z(0 = S(S(0)) + z)$.

Definition 6. If $\mathbf{T} = (T_i)_{i\in I}$ is an I-indexed family of $\mathscr{L}_{\mathrm{PA}}(I)$-theories and $\mathscr{N}$ is an $\mathscr{L}_{\mathrm{PA}}(I)$-structure, we say $\mathscr{N} \models \mathbf{T}$ if $\mathscr{N} \models T_i$ for all $i \in I$.

Definition 7. Suppose $\mathbf{T} = (T_i)_{i\in I}$ is an I-indexed family of $\mathscr{L}_{\mathrm{PA}}(I)$-theories. The *intended structure* for $\mathbf{T}$ is the $\mathscr{L}_{\mathrm{PA}}(I)$-structure $\mathscr{M}_{\mathbf{T}}$ with standard first-order part, interpreting the operators $\mathbf{T}_i\vDash$ $(i \in I)$ as follows:

$$\mathscr{M}_{\mathbf{T}} \models \mathbf{T}_i\vDash\phi[s] \text{ if and only if } T_i \models \phi^s.$$

If $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$, we say $\mathbf{T}$ is *true*.

Lemma 8. For any family $\mathbf{T} = (T_i)_{i\in I}$ of $\mathscr{L}_{\mathrm{PA}}(I)$-theories, $\mathscr{M}_{\mathbf{T}}$ interprets formulas by substitution.

Proof. In other words, we must show that for every $\mathscr{L}_{\mathrm{PA}}(I)$-formula ϕ and assignment s, $\mathscr{M}_{\mathbf{T}} \models \phi[s]$ if and only if $\mathscr{M}_{\mathbf{T}} \models \phi^s$. The proof is a straightforward induction. □

Definition 9. By the *axioms of Peano arithmetic for* $\mathscr{L}_{\mathrm{PA}}(I)$ we mean the axioms of Peano arithmetic, with induction extended to $\mathscr{L}_{\mathrm{PA}}(I)$.

Lemma 10. For any $\mathscr{L}_{\mathrm{PA}}(I)$-structure $\mathscr{M}$, if $\mathscr{M}$ interprets formulas by substitution, then $\mathscr{M}$ satisfies the axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(I)$.

Proof. Let $\mathscr{M}$ be any $\mathscr{L}_{\mathrm{PA}}(I)$-structure which interprets formulas by substitution. This means $\mathscr{M}$ has standard-first order part and for every formula ϕ and assignment s, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^s$.

Let σ be an axiom of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(I)$. If σ is not an instance of induction, then $\mathscr{M} \models \sigma$ since $\mathscr{M}$ has standard first-order part. But suppose σ is $\mathrm{ucl}(\phi(x|0) \to \forall x(\phi \to \phi(x|S(x))) \to \forall x\phi)$. To see $\mathscr{M} \models \sigma$, let s be an arbitrary assignment and assume $\mathscr{M} \models \phi(x|0)[s]$ and $\mathscr{M} \models \forall x(\phi \to \phi(x|S(x)))[s]$. By assumption, $\mathscr{M} \models \phi^{s(x|0)}$ and $\forall m \in \mathbb{N}$, if $\mathscr{M} \models \phi^{s(x|m)}$ then $\mathscr{M} \models \phi(x|S(x))^{s(x|m)}$. Evidently $\phi(x|S(x))^{s(x|m)} \equiv \phi^{s(x|m+1)}$. By mathematical induction, $\forall m \in \mathbb{N}$, $\mathscr{M} \models \phi^{s(x|m)}$. By assumption, $\mathscr{M} \models \forall x\phi[s]$. □

[1] We write $A \to B \to C$ for $A \to (B \to C)$, and likewise for longer chains.

Definition 11. Suppose $\mathbf{T} = (T_i)_{i\in I}$ is a family $\mathscr{L}_{\mathrm{PA}}(I)$-theories. If $\mathbf{T}^+ = (T_i^+)_{i\in I}$ is another such family, we say $\mathbf{T} \subseteq \mathbf{T}^+$ if $T_i \subseteq T_i^+$ for every $i \in I$. If T is a single $\mathscr{L}_{\mathrm{PA}}(I)$-theory, we say $T \subseteq \mathbf{T}$ if $T \subseteq T_i$ for all $i \in I$. If $\mathbf{T}^1 = (T_i^1)_{i\in I}$ and $\mathbf{T}^2 = (T_i^2)_{i\in I}$ are families of $\mathscr{L}_{\mathrm{PA}}(I)$-theories, $\mathbf{T}^1 \cup \mathbf{T}^2$ is the family $\mathbf{T}' = (T_i')_{i\in I}$ where each $T_i' = T_i^1 \cup T_i^2$. Arbitrary unions $\bigcup_{n\in X} \mathbf{T}^n$ are defined similarly.

Definition 12. Suppose $\mathbf{T} = (T_i)_{i\in I}$ is a family of $\mathscr{L}_{\mathrm{PA}}(I)$-theories. For each $i \in I$, we say T_i is $\mathbf{T}_i\vDash$-*closed* if $\mathbf{T}_i\vDash\phi \in T_i$ whenever $\phi \in T_i$. We say $\mathbf{T}$ is *closed* if each T_i is $\mathbf{T}_i\vDash$-closed.

Definition 13. If I is an r.e. index set, a family $\mathbf{T} = (T_i)_{i\in I}$ is *r.e.* just in case $\{(\phi, i) : \phi \in T_i\}$ is r.e.

3 Generic Axioms

If $\mathbf{T}$ is a family of theories whose truth was in doubt, and if we state a theorem removing that doubt, we often state more: that $\mathbf{T}\cup\mathbf{S}$ is true, where $\mathbf{S}$ is some background theory of provability, including non-controversial things like Peano arithmetic or the schema $\mathrm{ucl}(\mathbf{T}_i\vDash(\phi \to \psi) \to \mathbf{T}_i\vDash\phi \to \mathbf{T}_i\vDash\psi)$. The choice of $\mathbf{S}$ is somewhat arbitrary, or at best based on tradition. We will avoid this arbitrary choice by stating results in the form: "$\mathbf{T}$ is true together with any background theory of provability such that…"

Definition 14. A family $\mathbf{T}$ of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories is *closed-r.e.-generic* if $\mathbf{T}$ is r.e. and $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}$ for every closed r.e. family $\mathbf{U} \supseteq \mathbf{T}$ of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories.

Lemma 15. If $\mathbf{T}$ is a union of closed-r.e.-generic families and $\mathbf{T}$ is r.e., then $\mathbf{T}$ is closed-r.e.-generic.

Proof. Straightforward. □

Definition 16. For $i \in I$ and for T an $\mathscr{L}_{\mathrm{PA}}(I)$-theory, we write $[T]_i$ for the family $\mathbf{T} = (T_k)_{k\in I}$ where $T_i = T$ and $T_k = \emptyset$ for all $k \neq i$.

3.1 Closed-r.e.-generic Building Blocks

In this subsection, we will exhibit some examples of closed-r.e.-generic families. They can be combined in diverse ways, via Lemma 15, to form background theories of provability. This will allow us to state Theorem 24 below in a generalized way, essentially saying that a certain doubted theory is consistent with *any* background theory of provability made up of closed-r.e.-generic building blocks. The alternative would be for us to arbitrarily choose one such background theory and build it directly into Theorem 24, which would cause the core details in the proof of Theorem 24 to get jumbled up with unimportant distractions.

It is common for a theory to state its own closure under modus ponens. When there are multiple theories, it is less clear whether each individual theory should only state its own closure thereunder, or the closure of all the other theories, or of some subset thereof. With the following lemma, we avoid arbitrarily imposing a decision along these lines.

Lemma 17. For any $i, j \in \omega$, the following family is closed-r.e.-generic:

- $[S]_i$ where S is: (j-Deduction) the schema $\mathrm{ucl}(\mathbf{T}_j\vDash(\phi \to \psi) \to \mathbf{T}_j\vDash\phi \to \mathbf{T}_j\vDash\psi)$.

Proof. Let $\mathbf{U} = (U_k)_{k\in\omega}$ be any closed r.e. family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories such that $\mathbf{U} \supseteq [S]_i$ where S is j-Deduction. We must show $\mathscr{M}_{\mathbf{U}} \models [S]_i$. In other words we must show $\mathscr{M}_{\mathbf{U}} \models \mathrm{ucl}(\mathbf{T}_j\vDash(\phi \to \psi) \to \mathbf{T}_j\vDash\phi \to \mathbf{T}_j\vDash\psi)$ for any ϕ, ψ. Let s be an assignment and assume $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_j\vDash(\phi \to \psi)[s]$ and $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_j\vDash\phi[s]$, we must show $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_j\vDash\psi[s]$. By Definition of $\mathscr{M}_{\mathbf{U}}$, $U_j \models (\phi \to \psi)^s$ and $U_j \models \phi^s$. Clearly $(\phi \to \psi)^s \equiv \phi^s \to \psi^s$ so by modus ponens $U_j \models \psi^s$, that is, $\mathscr{M}_{\mathbf{U}} \models \psi[s]$. □

It might not be controversial to require that a theory express its own ability to prove valid sentences, but in a multi-theory context, should we require each theory to express that much about all its fellow theories? The following lemma allows us to avoid arbitrarily declaring the right answer to that question. Part 2 of this lemma illustrates an interesting combinatorial property of closed-r.e.-generic building blocks. Some schemas would not be suitable building blocks by themselves, but when paired with other schemas, the combination can become a suitable building block.

Lemma 18. For any $i, j \in \omega$, the following families are closed-r.e.-generic:

1. $[S]_i$ where S is: (Assigned Validity) the schema ϕ^s (ϕ valid, s an assignment).
2. [Assigned Validity]$_i \cup [S]_j$ where S is: (i-Validity) ucl($\mathbf{T}_i\vDash\phi$) for ϕ valid.

Proof. Both (1) and (2) are r.e. by Theorem 3.

(1) Let $\mathbf{U} = (U_k)_{k\in\omega}$ be a closed r.e. superset of $[S]_i$ where S is Assigned Validity. We must show $\mathscr{M}_{\mathbf{U}} \models [S]_i$. If $\phi \in [S]_i$ then ϕ is ϕ_0^s for some valid ϕ_0 and some assignment s. Since ϕ_0 is valid, $\mathscr{M}_{\mathbf{U}} \models \phi_0[s]$. By Lemma 8, $\mathscr{M}_{\mathbf{U}} \models \phi_0^s$.

(2) Let $\mathbf{U} = (U_k)_{k\in\omega}$ be any closed r.e. family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories such that U_i contains Assigned Validity and U_j contains i-Validity. By (1), $\mathscr{M}_{\mathbf{U}}$ satisfies Assigned Validity. It remains to show $\mathscr{M}_{\mathbf{U}}$ satisfies i-Validity. Let ϕ be valid and s an assignment. Since U_i contains Assigned Validity, $U_i \models \phi^s$, so by definition of $\mathscr{M}_{\mathbf{U}}$, $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_i\vDash\phi[s]$. □

In modal logic, some papers treat the so-called *positive introspection axiom* (also known as the *KK axiom*) as one of the fundamental axioms of knowledge, and some do not. Rather than join either side, we prefer instead to study the combinatorial structure of the axiom, asking: are there other schemas we can add to it to make the combination closed-r.e.-generic?

Lemma 19. For any $i, j \in \omega$, the following family is closed-r.e.-generic:

- [Assigned Validity]$_i \cup [i$-Validity$]_i \cup [i$-Deduction$]_i \cup [S]_j$ where S is:

  (i-Introspection) the schema ucl($\mathbf{T}_i\vDash\phi \to \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi$).

Proof. Recursive enumerability is by Theorem 3. Let $\mathbf{U} = (U_k)_{k\in\omega}$ be any closed r.e. family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories such that U_i contains Assigned Validity, i-Validity and i-Deduction, and U_j contains i-Introspection. Then $\mathscr{M}_{\mathbf{U}}$ satisfies Assigned Validity and i-Validity by Lemma 18. By Lemma 17, $\mathscr{M}_{\mathbf{U}}$ satisfies i-Deduction. For i-Introspection, let s be an assignment and assume $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_i\vDash\phi[s]$, we will show $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi[s]$. Since $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_i\vDash\phi[s]$, $U_i \models \phi^s$. By Theorem 3, there are $\sigma_1, \ldots, \sigma_n \in U_i$ such that $\sigma_1 \to \cdots \to \sigma_n \to \phi^s$ is valid. Since U_i contains i-Validity, $U_i \models \mathbf{T}_i\vDash(\sigma_1 \to \cdots \to \sigma_n \to \phi^s)$. By repeated applications of i-Deduction contained in U_i, $U_i \models \mathbf{T}_i\vDash\sigma_1 \to \cdots \to \mathbf{T}_i\vDash\sigma_n \to \mathbf{T}_i\vDash\phi^s$. Since $\mathbf{U}$ is closed, U_i is $\mathbf{T}_i\vDash$-closed and so contains $\mathbf{T}_i\vDash\sigma_1, \ldots, \mathbf{T}_i\vDash\sigma_n$. So $U_i \models (\mathbf{T}_i\vDash\phi)^s$ and $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi[s]$. □

The following lemma shows that arithmetic is generic, which will enable us to state a later result (Theorem 24) in such a way that it is clear that the result is neither contingent on the presence, nor the absense, of arithmetic in the theories in question.

Lemma 20. For any $i \in \omega$, $[S]_i$ is closed-r.e.-generic, where S is the set of axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(\omega)$.

Proof. By Lemmas 8 and 10. □

Carlson proved [5] that it is consistent for an idealized knowing machine to know "I am a machine" (without knowing which specific machine it is). The following lemma sheds additional light: not only is it consistent for a knowing machine to know "I am a machine", in fact that knowledge is generic: it does not depend heavily on specific arbitrary decisions about the background theory of provability.

Lemma 21. For any $i, j \in \omega$, the following family is closed-r.e.-generic.

- $[S]_i$ where S is: (j-SMT) (See [5] and [11]) ucl($\exists e\forall x(\mathbf{T}_j\vDash\phi \leftrightarrow x \in W_e)$), $e \notin \mathrm{FV}(\phi)$.

Proof. Suppose $\mathbf{U} = (U_i)_{i\in\omega}$ is a closed r.e. family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories and $\mathbf{U} \supseteq [S]_i$ where S is j-SMT. We must show $\mathscr{M}_{\mathbf{U}} \models [S]_i$. That is, given ϕ with $e \notin \mathrm{FV}(\phi)$, we must show $\mathscr{M}_{\mathbf{U}} \models \mathrm{ucl}(\exists e\forall x(\mathbf{T}_j\vDash\phi \leftrightarrow x \in W_e))$. Let s be an assignment and let $x_1, \ldots, x_k = \mathrm{FV}(\phi)\backslash\{x\}$. Since U_j is r.e., by the S-m-n theorem there is some n such that $W_n = \{m : U_j \models \phi(x|\overline{m})(x_1|\overline{s(x_1)}) \cdots (x_k|\overline{s(x_k)})\}$. Since $e \notin \mathrm{FV}(\phi)$, and $\mathscr{M}_{\mathbf{U}}$ has standard first-order part, it follows that $\mathscr{M}_{\mathbf{U}} \models \forall x(\mathbf{T}_j\vDash\phi \leftrightarrow x \in W_e)[s(e|n)]$. □

Finally, the following lemma offers a way to obtain new building blocks from old. This can be combined with Lemma 21 to advance from "I am a machine" to "I know I am a machine".

Lemma 22. For any $i, j \in \omega$ and any closed-r.e.-generic family $\mathbf{T} = (T_k)_{k\in\omega}$, $\mathbf{T} \cup [S]_i$ is closed-r.e.-generic, where S is the schema: $\mathbf{T}_j{\vDash}\phi$ $(\phi \in T_j)$.

Proof. Suppose $\mathbf{U} = (U_i)_{i\in\omega} \supseteq \mathbf{T} \cup [S]_i$ is closed and r.e. Right away $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}$ because $\mathbf{T}$ is closed-r.e.-generic. It remains to show that $\mathscr{M}_{\mathbf{U}} \models [S]_i$, i.e., that $\mathscr{M}_{\mathbf{U}} \models S$. Fix $\phi \in T_j$ and let s be any assignment. Since ϕ is a sentence, $\phi \equiv \phi^s$ and thus $T_j \models \phi^s$. Since $U_j \supseteq T_j$, $U_j \models \phi^s$. By definition of $\mathscr{M}_{\mathbf{U}}$, $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}_j{\vDash}\phi[s]$. □

We gather Lemmas 17–22 together into the following summary.

Corollary 23. For any $i, j \in \omega$, each of the following families is closed-r.e.-generic.

1. $[j\text{-Deduction}]_i$.
2. $[\text{Assigned Validity}]_i$.
3. $[\text{Assigned Validity}]_i \cup [i\text{-Validity}]_j$.
4. $[\text{Assigned Validity}]_i \cup [i\text{-Validity}]_i \cup [i\text{-Deduction}]_i \cup [i\text{-Introspection}]_j$.
5. $[S]_i$ where S is the set of axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(\omega)$.
6. $[j\text{-SMT}]_i$.
7. $\mathbf{T} \cup [S]_i$, for any closed-r.e.-generic $\mathbf{T}$, where S is the schema: $\mathbf{T}_j{\vDash}\phi$ $(\phi \in T_j)$.

The above building blocks are not exhaustive. In choosing building blocks, a primary concern was to facilitate creation of background provability theories strong enough to make our consistency result (Theorem 24) generalize Carlson's consistency result [5]. If that were our lone motivation, we could restrict Corollary 23 to only those families where $i = j$, but a secondary motivation was to provide inter-theory versions of those restricted building blocks. It would be interesting to investigate questions about whether the above building-blocks are minimal. For example, in Lemma 19, is it really necessary to bundle j-Introspection with all three other schemas? For now, we will leave those questions open.

4 First Consistency Result: Prioritizing Exact Codes

The following theorem fulfils the first promise from the introduction.

Theorem 24. Suppose $\prec$ is an r.e. well-founded partial order on ω and $\mathbf{T}^0 = (T^0_i)_{i\in\omega}$ is closed-r.e.-generic. For each $n \in \mathbb{N}$, let $\mathbf{T}(n) = (T_i(n))_{i\in\omega}$ where each $T_i(n)$ is the smallest $\mathbf{T}_i{\vDash}$-closed theory containing the following:

1. The axioms in T^0_i.
2. $\forall x(\mathbf{T}_j{\vDash}\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ whenever $j \in \omega$, $\mathrm{FV}(\phi) \subseteq \{x\}$.
3. $\mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \to \phi)$ whenever $j \prec i$.

There is some $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true.

Proof. By the *S-m-n* Theorem, there is a total computable $f : \mathbb{N} \to \mathbb{N}$ such that $\forall n \in \mathbb{N}$,

$$W_{f(n)} = \{\langle\ulcorner\phi\urcorner, i, m\rangle : \mathrm{FV}(\phi) \subseteq \{x\} \text{ and } T_i(n) \models \phi(x|\overline{m})\}.$$

Using the Recursion Theorem, fix $n \in \mathbb{N}$ such that $W_{f(n)} = W_n$. For brevity write $\mathbf{T}$ for $\mathbf{T}(n)$ and T_i for $T_i(n)$. We will show $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$. This is a self-referential statement: to show T_i is true includes showing $\mathscr{M}_{\mathbf{T}} \models \mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \to \phi)$, which is essentially the statement that T_j is true. Hence the restriction $j \prec i$, which allows induction since $\prec$ is well founded. We will show, by $\prec$-induction on i, that $\mathscr{M}_{\mathbf{T}} \models T_i$ for every $i \in \omega$. Fix $i \in \omega$ and assume $\mathscr{M}_{\mathbf{T}} \models T_j$ for all $j \prec i$. Suppose $\sigma \in T_i$, we will show $\mathscr{M}_{\mathbf{T}} \models \sigma$.

Case 1: $\sigma \in T_i^0$. Then $\mathscr{M}_{\mathbf{T}} \models \sigma$ because $\mathbf{T}^0$ is closed-r.e.-generic and $\mathbf{T} \supseteq \mathbf{T}^0$ is closed r.e.

Case 2: σ is $\forall x(\mathbf{T}_j{\vDash}\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ for some $j \in \omega$, $\mathrm{FV}(\phi) \subseteq \{x\}$. Let s be an assignment, $m \in \mathbb{N}$. The following are equivalent.

$$
\begin{aligned}
\mathscr{M}_{\mathbf{T}} &\models \mathbf{T}_j{\vDash}\phi[s(x|m)] && \\
T_j &\models \phi^{s(x|m)} && \text{(Definition of } \mathscr{M}_{\mathbf{T}}\text{)} \\
T_j &\models \phi(x|\overline{m}) && \text{(Since } \mathrm{FV}(\phi) \subseteq \{x\}\text{)} \\
\langle\ulcorner\phi\urcorner, j, m\rangle &\in W_n && \text{(By definition of } n\text{)} \\
\mathscr{M}_{\mathbf{T}} &\models \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, \overline{m}\rangle \in W_{\overline{n}} && (\mathscr{M}_{\mathbf{T}} \text{ has standard first-order part}) \\
\mathscr{M}_{\mathbf{T}} &\models \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}}[s(x|m)]. && \text{(Lemma 8)}
\end{aligned}
$$

Case 3: σ is $\mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \to \phi)$ for some $j \prec i$. Let s be an assignment and assume $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}_j{\vDash}\phi[s]$. This means $T_j \models \phi^s$. By our $\prec$-induction hypothesis, $\mathscr{M}_{\mathbf{T}} \models T_j$, so $\mathscr{M}_{\mathbf{T}} \models \phi^s$. By Lemma 8, $\mathscr{M}_{\mathbf{T}} \models \phi[s]$.

Case 4: σ is only present in T_i because of the clause that T_i is $\mathbf{T}_i{\vDash}$-closed. Then σ is $\mathbf{T}_i{\vDash}\sigma_0$ for some $\sigma_0 \in T_i$. Being in T_i, σ_0 is a sentence, so for any assignment s, $\sigma_0 \equiv \sigma_0^s$, $T_i \models \sigma_0^s$, and finally $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}_i{\vDash}\sigma_0[s]$.

By $\prec$-induction, $\mathscr{M}_{\mathbf{T}} \models T_i$ for all $i \in \omega$. This shows $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$, that is, $\mathbf{T}$ is true. □

The first promise from the introduction is met: for any r.e. well-founded partial order $\prec$ on ω, there are theories $(T_n)_{n\in\omega}$ such that $\forall i, j, k \in \omega$ with $j \prec i$, T_i expresses the truth of T_j, and T_i expresses a Gödel number of T_k. In order to fulfill the second promise we will extend Carlson's notion of *stratification* to the case of multiple operators, and introduce *stratifiers*, a tool used to deal with subtleties that arise when multiple self-referential theories refer to one another.

In [2] the technique behind Theorem 24 was used to exhibit a machine that knows its own code.

5 Stratification

For the second promise from the introduction, we need to prove a result like Theorem 24 where T_i includes $\mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \to \phi)$ for all $j \preceq i$, not just $j \prec i$. This rules out the direct $\prec$-induction of the type used above. Induction on formula complexity will not work either: we would need to show all of T_i consistent just to show $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}_i{\vDash}(1 = 0) \to (1 = 0)$. Instead, we will use ordinal induction. But there are no ordinals anywhere in sight. To obtain ordinals to induct on, we will modify the theories we care about, in a process called *stratification*. We will start with some informal motivational remarks. Readers who would like to advance directly to the formal definitions can safely skip Subsection 5.1.

5.1 Motivation for Stratification

As explained above, we would like to invoke ordinal induction, but there are no ordinals in sight. In order to make ordinal induction relevant, we will do the following. We will extend the background language to contain not only the operators $\mathbf{T}_i{\vDash}$ $(i \in \omega)$, but also operators $\mathbf{T}_i^{\alpha}{\vDash}$ $(i \in \omega,\ \alpha \in \epsilon_0 \cdot \omega)$. And instead of focusing directly on T_i, we will focus on a theory U_i such that the result U_i^- of erasing superscripts from U_i is $U_i^- = T_i$. The intended interpretation of $\mathbf{T}_i^{\alpha}{\vDash}\phi[s]$ will be $U_i \cap \alpha \models \phi^s$, where $U_i \cap \alpha$ is the set of axioms of U_i whose superscripts are $< \alpha$. Thus, we may think of U_i as a version of T_i with extra information about the structure of T_i. We will show (Theorem 50), for certain formulas ϕ whose superscripts are positive multiples of ϵ_0, that ϕ holds (in the intended interpretation) if and only if ϕ^- holds. We will use this, after proving that U_i holds, to conclude that T_i also holds.

Suppose we would like T_i to contain the axiom $\mathbf{T}_i{\vDash}(1 + 1 = 2)$. Then, as we carry out the procedure in the above paragraph, we would ensure that U_i contain all sentences of the form $\mathbf{T}_i^{\alpha}{\vDash}(1 + 1 = 2)$. This would have the side effect that for any $\beta > \alpha$, $U_i \cap \beta \models \mathbf{T}_i^{\alpha}{\vDash}(1 + 1 = 2)$, so that $\mathbf{T}_i^{\beta}{\vDash}\mathbf{T}_i^{\alpha}{\vDash}(1 + 1 = 2)$ would hold in structures with the intended interpretation.

Next, suppose that for every arithmetical sentence ϕ, we would like T_i to include

$$\mathbf{T}_i{\vDash}\phi \to \mathbf{T}_i{\vDash}\mathbf{T}_i{\vDash}\phi.$$

Then we would arrange that U_i contain

$$\mathbf{T}_i^{\alpha}\vDash\phi \to \mathbf{T}_i^{\beta}\vDash\mathbf{T}_i^{\alpha}\vDash\phi$$

(whenever $\beta > \alpha$). The reason for the β is as follows. The intended interpretation of $\mathbf{T}_i^{\alpha}\vDash\phi$ shall be $U_i \cap \alpha \models \phi$. Thus, it would make no sense to put the axiom $\mathbf{T}_i^{\alpha}\vDash\phi \to \mathbf{T}_i^{\alpha}\vDash\mathbf{T}_i^{\alpha}\vDash\phi$ into U_i: the fact that $U_i \cap \alpha \models \phi$ does not generally imply that $U_i \cap \alpha \models \mathbf{T}_i^{\alpha}\vDash\phi$, since $U_i \cap \alpha$ is limited to formulas in which all superscripts are $< \alpha$. At least $\mathbf{T}_i^{\alpha}\vDash\phi \to \mathbf{T}_i^{\beta}\vDash\mathbf{T}_i^{\alpha}\vDash\phi$ is plausible.

Again, suppose that for some $j \prec i$, we would like for T_i to include

$$\mathbf{T}_i\vDash(\mathbf{T}_j\vDash(1=0) \to (1=0)).$$

We would arrange that U_i contain (for all α):

$$\mathbf{T}_i^{\alpha}\vDash(\mathbf{T}_j\vDash(1=0) \to (1=0)).$$

Note the lack of superscript on $\mathbf{T}_j\vDash$. The intuition is that U_i is a version of T_i with extra information about the structure of T_i (namely, that said structure arises from an increasing family of theories), but without any additional information about the structure of T_j.

Similarly, suppose we would like T_i to include

$$\mathbf{T}_j\vDash(\mathbf{T}_i\vDash(1=0)) \to \mathbf{T}_i\vDash(1=0).$$

We would arrange that U_i contain (for each α):

$$\mathbf{T}_j\vDash(\mathbf{T}_i\vDash(1=0)) \to \mathbf{T}_i^{\alpha}\vDash(1=0).$$

Note the lack of superscript on the $\mathbf{T}_i\vDash$ within the scope of $\mathbf{T}_j\vDash$. As above, the intuition is that U_i is a version of T_i with extra information about the structure of T_i. It does not have any extra information about the structure of T_j—not even about what T_j says about T_i. This is important because, when $j \prec i$, we would like T_i to contain axioms declaring, essentially, the Gödel number of T_j. This Gödel number would be hardcoded into such axioms, and thus there would be no hope of such axioms remaining true if T_j were changed.

5.2 Stratification Formal Details

To get a foothold for induction, instead of considering a particular theory T_i, we will be considering copies of T_i with ordinal-number superscripts added. To recover information about the original T_i from these modified theories, we will need to use sophisticated results from [4] about the structure of the ordinals.

Definition 25. We define a binary relation $\leq_1$ on Ord by transfinite recursion so that for all $\alpha, \beta \in \text{Ord}$, $\alpha \leq_1 \beta$ if and only if $\alpha \leq \beta$ and $(\alpha, \leq, \leq_1)$ is a Σ_1-elementary substructure of $(\beta, \leq, \leq_1)$.

The following theorem is based on calculations from [4]. It was used by Carlson to prove Reinhardt's conjecture [5]. We state it here without proof.

Theorem 26.

1. The binary relation $\leq_1$ is a recursive partial ordering on $\epsilon_0 \cdot \omega$.
2. For all positive integers $m \leq n$, $\epsilon_0 \cdot m \leq_1 \epsilon_0 \cdot n$.
3. For any $\alpha \leq \beta \in \text{Ord}$, $\alpha \leq_1 \beta$ if and only if the following statement is true. For every finite set $X \subseteq \alpha$ and every finite set $Y \subseteq [\alpha, \beta)$, there is a set $X < \widetilde{Y} < \alpha$ such that $X \cup \widetilde{Y} \cong_{(\leq, \leq_1)} X \cup Y$.

The usefulness of Theorem 26 will appear in Theorem 38, but first we need some machinery.

Definition 27. Let $\mathcal{I} = ((\epsilon_0 \cdot \omega) \times \omega) \sqcup \omega$. Thus $\mathscr{L}_{\text{PA}}(\mathcal{I})$ contains operators $\mathbf{T}_{(\alpha,i)}\vDash$ for all $\alpha \in \epsilon_0 \cdot \omega$, $i \in \omega$, along with operators $\mathbf{T}_i\vDash$ for all $i \in \omega$. As abbreviation, we write $\mathbf{T}_i^{\alpha}\vDash$ for $\mathbf{T}_{(\alpha,i)}\vDash$, and refer to α as its *superscript*.

Definition 28. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, $\mathrm{On}(\phi) \subseteq \epsilon_0 \cdot \omega$ denotes the set of superscripts appearing in ϕ.

Definition 29. Suppose $i \in \omega$. The *i-stratified* formulas of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$ are defined as follows (where ϕ ranges over $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formulas).

1. If ϕ is $\mathbf{T}_j{\vDash}\phi_0$ for some $j \neq i$, then ϕ is i-stratified if and only if ϕ is an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula.
2. If ϕ is $\mathbf{T}_j^\alpha{\vDash}\phi_0$ for some $j \neq i$, then ϕ is not i-stratified.
3. If ϕ is $\mathbf{T}_i{\vDash}\phi_0$, then ϕ is not i-stratified.
4. If ϕ is $\mathbf{T}_i^\alpha{\vDash}\phi_0$, then ϕ is i-stratified if and only if ϕ_0 is i-stratified and $\alpha > \mathrm{On}(\phi_0)$.
5. If ϕ is $\neg\phi_0$, $\phi_1 \to \phi_2$, or $\forall x \phi_0$, then ϕ is i-stratified if and only if its immediate subformula(s) are.
6. If ϕ is atomic, then ϕ is i-stratified.

An $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory T is i-stratified if ϕ is i-stratified whenever $\phi \in T$. An $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ is *very i-stratified* if ϕ is i-stratified and $\mathrm{On}(\phi) \subseteq \{\epsilon_0 \cdot 1, \epsilon_0 \cdot 2, \ldots\}$.

For example:

- $\mathbf{T}_7^\omega{\vDash}\mathbf{T}_7^5{\vDash}(1=0) \to \mathbf{T}_8{\vDash}(1=0)$ is 7-stratified but not 6- or 8-stratified.
- $\mathbf{T}_7^5{\vDash}\mathbf{T}_7^\omega{\vDash}(1=0)$ is not 7-stratified, nor is $\mathbf{T}_7^5{\vDash}\mathbf{T}_7{\vDash}(1=0)$.
- $\mathbf{T}_7^5{\vDash}\mathbf{T}_8{\vDash}\mathbf{T}_7{\vDash}(1=0)$ is 7-stratified but $\mathbf{T}_7^5{\vDash}\mathbf{T}_8{\vDash}\mathbf{T}_7^4{\vDash}(1=0)$ is not.

We will not make use of the following lemma, but we state it to further illuminate Definition 29.

Lemma 30. Suppose ϕ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula, $i \in \omega$. Then ϕ is i-stratified if and only if all of the following conditions hold.

1. For all $j \in \omega$ and $\alpha \in \epsilon_0 \cdot \omega$, if $\mathbf{T}_j^\alpha{\vDash}$ occurs in ϕ, then $j = i$.
2. Every occurrence of $\mathbf{T}_i{\vDash}$ in ϕ is inside the scope of $\mathbf{T}_j{\vDash}$ for some $j \neq i$.
3. $\mathbf{T}_i^\alpha{\vDash}$ never occurs in ϕ inside the scope of $\mathbf{T}_j{\vDash}$, for any $\alpha \in \epsilon_0 \cdot \omega$ or any $j \in \omega$.
4. For all $\alpha, \beta \in \epsilon_0 \cdot \omega$, if $\mathbf{T}_i^\alpha{\vDash}$ occurs in ϕ inside the scope of $\mathbf{T}_i^\beta{\vDash}$, then $\beta > \alpha$.

Proof. Straightforward. □

Definition 31. Suppose $X \subseteq \epsilon_0 \cdot \omega$ and $h : X \to \epsilon_0 \cdot \omega$ is order preserving. For each $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, define an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula $h(\phi)$ inductively as follows:

1. If ϕ is $\neg\phi_0$, $\phi_1 \to \phi_2$, or $\forall x \phi_0$, then $h(\phi)$ is $\neg h(\phi_0)$, $h(\phi_1) \to h(\phi_2)$, or $\forall x h(\phi_0)$, respectively.
2. If ϕ is atomic or $\mathbf{T}_i{\vDash}\phi_0$, then $h(\phi) \equiv \phi$.
3. If ϕ is $\mathbf{T}_i^\alpha{\vDash}\phi_0$ where $\alpha \in X$, then $h(\phi) \equiv \mathbf{T}_i^{h(\alpha)}{\vDash}h(\phi_0)$.
4. If ϕ is $\mathbf{T}_i^\alpha{\vDash}\phi_0$ where $\alpha \notin X$, then $h(\phi) \equiv \mathbf{T}_i^\alpha{\vDash}h(\phi_0)$.

In practice, we will mainly be interested in ϕ when ϕ is i-stratified for some i, in which case $\mathbf{T}_j^\alpha{\vDash}$ cannot occur within the scope of $\mathbf{T}_k{\vDash}$ in ϕ for any k, j. For such ϕ, $h(\phi)$ is simply the result of applying h to every superscript in ϕ that is in X.

For example if $X = \{1, \omega\}$, $h(1) = 0$, and $h(\omega) = \omega \cdot 2 + 1$, then

$$h\left(\mathbf{T}_i^0{\vDash}(1=0) \to \mathbf{T}_i^1{\vDash}(1=0) \to \mathbf{T}_i^\omega{\vDash}(1=0)\right) \quad \equiv \quad \mathbf{T}_i^0{\vDash}(1=0) \to \mathbf{T}_i^0{\vDash}(1=0) \to \mathbf{T}_i^{\omega\cdot 2+1}{\vDash}(1=0).$$

In practice, we will primarily be interested in applying Definition 31 in the case where $\mathrm{On}(\phi) \subseteq X$.

Definition 32. Suppose $X \subseteq \epsilon_0 \cdot \omega$ and $h : X \to \epsilon_0 \cdot \omega$ is order preserving. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}$, we define an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $h(\mathscr{N})$ that has the same universe as $\mathscr{N}$, agrees with $\mathscr{N}$ on $\mathscr{L}_{\mathrm{PA}}(\omega)$, and interprets $\mathscr{L}_{\mathrm{PA}}(\mathcal{I}) \backslash \mathscr{L}_{\mathrm{PA}}(\omega)$ so that

$$h(\mathscr{N}) \models \mathbf{T}_i^{\alpha} \vDash \phi[s] \text{ if and only if } \mathscr{N} \models h(\mathbf{T}_i^{\alpha} \vDash \phi)[s].$$

Lemma 33. Suppose $X \subseteq \epsilon_0 \cdot \omega$, $h : X \to \epsilon_0 \cdot \omega$ is order preserving, and $\mathscr{N}$ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $h(\mathscr{N}) \models \phi[s]$ if and only if $\mathscr{N} \models h(\phi)[s]$.

Proof. By induction. □

Corollary 34. Suppose $X \subseteq \epsilon_0 \cdot \omega$ and $h : X \to \epsilon_0 \cdot \omega$ is order preserving. For any valid $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, $h(\phi)$ is valid.

Proof. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}$ and assignment s, $h(\mathscr{N}) \models \phi[s]$ by validity, so $\mathscr{N} \models h(\phi)[s]$ by Lemma 33. □

Definition 35. If $X \subseteq \mathrm{Ord}$ and $h : X \to \mathrm{Ord}$, we call h a *covering* if h is order preserving and whenever $x, y \in X$ and $x \leq_1 y$, $h(x) \leq_1 h(y)$.

Definition 36. Suppose $i \in \omega$. An $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory T is *i-unistratified* if the following conditions hold:

1. T is i-stratified.
2. (Uniformity) Whenever $\phi \in T$, $X \subseteq \epsilon_0 \cdot \omega$, $\mathrm{On}(\phi) \subseteq X$, and $h : X \to \epsilon_0 \cdot \omega$ is a covering, then $h(\phi) \in T$.

Definition 37. If T is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory and $\alpha \in \epsilon_0 \cdot \omega$, let $T \cap \alpha$ be the set $\{\phi \in T : \mathrm{On}(\phi) \subseteq \alpha\}$ of sentences in T that do not contain any superscripts $\geq \alpha$.

Theorem 38. (The Collapse Theorem) Suppose T is an i-unistratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory.

1. If n is a positive integer and $\mathrm{On}(\phi) \subseteq \epsilon_0 \cdot n$, then $T \models \phi$ if and only if $T \cap (\epsilon_0 \cdot n) \models \phi$.
2. If $\alpha \leq_1 \beta$ and $\mathrm{On}(\phi) \subseteq \alpha$, then $T \cap \alpha \models \phi$ if and only if $T \cap \beta \models \phi$.

Proof. Note that since T is i-unistratified, in particular T is i-stratified. We will prove (1), the proof of (2) is similar.

($\Leftarrow$) Immediate since $T \cap (\epsilon_0 \cdot n) \subseteq T$.

($\Rightarrow$) Assume $T \models \phi$. By Theorem 3 there are $\sigma_1, \ldots, \sigma_k \in T$ such that

$$\Phi \equiv \sigma_1 \to \cdots \to \sigma_k \to \phi$$

is valid. Let $X = \mathrm{On}(\Phi) \cap (\epsilon_0 \cdot n)$, $Y = \mathrm{On}(\Phi) \cap [\epsilon_0 \cdot n, \infty)$, note $|X|, |Y| < \infty$.

Since Y is finite, there is some integer $n' > n$ such that $Y \subseteq \epsilon_0 \cdot n'$. By Theorem 26 part 2, $\epsilon_0 \cdot n \leq_1 \epsilon_0 \cdot n'$. By Theorem 26 part 3, there is some $X < \widetilde{Y} < \epsilon_0 \cdot n$ such that $X \cup \widetilde{Y} \cong_{(\leq, \leq_1)} X \cup Y$.

Let $h : X \cup Y \to X \cup \widetilde{Y}$ be a $(\leq, \leq_1)$-isomorphism. Since $\mathrm{On}(\phi) \subseteq \epsilon_0 \cdot n$, $h(\phi) = \phi$. By Corollary 34,

$$h(\Phi) \equiv h(\sigma_1) \to \cdots \to h(\sigma_k) \to \phi$$

is valid. Since T is i-unistratified, $h(\sigma_1), \ldots, h(\sigma_k) \in T$. Finally since $\mathrm{range}(h) < \epsilon_0 \cdot n$, $h(\sigma_1), \ldots, h(\sigma_k) \in T \cap (\epsilon_0 \cdot n)$, showing $T \cap (\epsilon_0 \cdot n) \models \phi$. □

Loosely speaking, what we have done in Theorem 38 is we have taken a proof of ϕ and we have *collapsed* the proof, shrinking its ordinals by using Theorem 26 part 3.

Definition 39. For every $i \in \omega$ we define the following $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-schema:

- (i-Collapse) $\mathrm{ucl}(\mathbf{T}_i^{\alpha} \vDash \phi \leftrightarrow \mathbf{T}_i^{\beta} \vDash \phi)$ whenever $\mathbf{T}_i^{\alpha} \vDash \phi$ is i-stratified and $\alpha \leq_1 \beta$.

Definition 40. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, ϕ^- is the result of erasing all superscripts from ϕ. If T is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory, $T^- = \{\sigma^- : \sigma \in T\}$.

For example, if ϕ is $\mathbf{T}_5^\omega\vDash(1=0)\to\mathbf{T}_5^{\omega+1}\vDash\mathbf{T}_5^\omega\vDash(1=0)$, then ϕ^- is $\mathbf{T}_5\vDash(1=0)\to\mathbf{T}_5\vDash\mathbf{T}_5\vDash(1=0)$.

Lemma 41. If T is i-unistratified then for every $\phi\in T$ there is some $\psi\in T$ such that ψ is very i-stratified and $\psi^-\equiv\phi^-$.

Proof. Let $X=\mathrm{On}(\phi)=\{\alpha_1<\cdots<\alpha_n\}$, $Y=\{\epsilon_0\cdot 1,\ldots,\epsilon_0\cdot n\}$, and define $h:X\to Y$ by $h(\alpha_j)=\epsilon_0\cdot j$. Clearly h is order preserving; by Theorem 26 part 2, h is a covering. Since T is i-unistratified, T contains $\psi\equiv h(\phi)$. Clearly ψ is very i-stratified and $\psi^-\equiv\phi^-$. □

Definition 42. For any $\mathscr{L}_{\mathrm{PA}}(\omega)$-structure $\mathscr{N}$, we define an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}^-$ that has the same universe as $\mathscr{N}$, agrees with $\mathscr{N}$ on $\mathscr{L}_{\mathrm{PA}}(\omega)$, and interprets $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})\backslash\mathscr{L}_{\mathrm{PA}}(\omega)$ as follows. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, $\alpha\in\epsilon_0\cdot\omega$, $i\in\mathbb{N}$, and assignment s,

$$\mathscr{N}^-\models\mathbf{T}_i^\alpha\vDash\phi[s]\text{ if and only if }\mathscr{N}\models(\mathbf{T}_i^\alpha\vDash\phi)^-[s].$$

Lemma 43. Suppose $\mathscr{N}$ is an $\mathscr{L}_{\mathrm{PA}}(\omega)$-structure. For every $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{N}^-\models\phi[s]$ if and only if $\mathscr{N}\models\phi^-[s]$.

Proof. By induction. □

Corollary 44. If ϕ is a valid $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula, then ϕ^- is a valid $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula.

Proof. Similar to the proof of Corollary 34. □

A converse-like statement holds for Corollary 44 as well.

Lemma 45. For any valid $\mathscr{L}_{\mathrm{PA}}(\omega)$-sentence ϕ and $i\in\omega$, there is a valid very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentence ψ such that $\psi^-\equiv\phi$.

Proof. Let $\psi\mapsto\psi^+$ be the function taking $\mathscr{L}_{\mathrm{PA}}(\omega)$-formulas to $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formulas defined as follows.

1. If ψ is atomic, or of the form $\mathbf{T}_j\vDash\psi_0$ with $j\neq i$, then $\psi^+\equiv\psi$.
2. If ψ is $\mathbf{T}_i\vDash\psi_0$, then $\psi^+\equiv\mathbf{T}_i^{\epsilon_0\cdot n}\vDash\psi_0^+$, where $n=\min\{m\in\mathbb{N}:\epsilon_0\cdot m>\mathrm{On}(\psi_0^+)\}$.
3. If ψ is $\neg\psi_0$, $\psi_0\to\psi_1$, or $\forall x\psi_0$, then ψ^+ is $\neg\psi_0^+$, $\psi_1^+\to\psi_2^+$, or $\forall x\psi_0^+$, respectively.

It is straightforward to show ϕ^+ is very i-stratified. We claim ϕ^+ is valid. Let $\mathscr{M}$ be any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure, we will show $\mathscr{M}\models\phi^+$. Let $\mathscr{M}^+$ be the $\mathscr{L}_{\mathrm{PA}}(\omega)$-structure with the same universe as $\mathscr{M}$, which agrees with $\mathscr{M}$ on the interpretation of arithmetic and of $\mathbf{T}_j\vDash$ for $j\neq i$, and which interprets $\mathbf{T}_i\vDash$ as follows:

$$\mathscr{M}^+\models\mathbf{T}_i\vDash\psi[s]\text{ if and only if }\mathscr{M}\models(\mathbf{T}_i\vDash\psi)^+[s].$$

Since ϕ is valid, $\mathscr{M}^+\models\phi$. It follows that $\mathscr{M}\models\phi^+$. □

Definition 46. Let $i\in\omega$. We define the following $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-schemas.

- (i-Strativalidity) $\mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi)$ whenever ϕ is a valid $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula and $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified.
- (i-Stratideduction) $\mathrm{ucl}(\mathbf{T}_i^\alpha\vDash(\phi\to\psi)\to\mathbf{T}_i^\alpha\vDash\phi\to\mathbf{T}_i^\alpha\vDash\psi)$ whenever this formula is i-stratified.

Definition 47. An $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory T is *i-straticlosed* if the following conditions hold:

1. T is i-unistratified.
2. T includes i-Strativalidity, i-Stratideduction and i-Collapse.
3. For every $\phi\in T$, if $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified then $\mathbf{T}_i^\alpha\vDash\phi\in T$.

A family $\mathbf{T}=(T_i)_{i\in\omega}$ is *straticlosed* if each T_i is i-straticlosed.

The following theorem serves as an omnibus of results from Section 5 of [5].

Theorem 48. (Proof Stratification) Suppose T is an i-straticlosed $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory. Then:

1. Whenever $T \cap \alpha \models \phi$, $\mathbf{T}_i^\alpha \vDash \phi$ is an i-stratified sentence, and $\beta > \alpha$, then $T \cap \beta \models \mathbf{T}_i^\alpha \vDash \phi$.
2. For any very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentences ρ and σ, if $\rho^- \equiv \sigma^-$ then $T \models \rho \leftrightarrow \sigma$.
3. For any very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentence ϕ, $T \models \phi$ if and only if $T^- \models \phi^-$.

Proof. Note that since T is i-straticlosed, in particular T is i-unistratified and hence, i-stratified.

Claim 0: Any time $T \models \mathbf{T}_i^\alpha \vDash (\rho \leftrightarrow \sigma)$ and this is i-stratified, $T \models \mathbf{T}_i^\alpha \vDash \rho \leftrightarrow \mathbf{T}_i^\alpha \vDash \sigma$.

Assume the hypotheses. By i-Strativalidity, $T \models \mathbf{T}_i^\alpha \vDash ((\rho \leftrightarrow \sigma) \to (\rho \to \sigma))$. By i-Stratideduction,

$$T \models \mathbf{T}_i^\alpha \vDash ((\rho \leftrightarrow \sigma) \to (\rho \to \sigma)) \to \mathbf{T}_i^\alpha \vDash (\rho \leftrightarrow \sigma) \to \mathbf{T}_i^\alpha \vDash (\rho \to \sigma)$$
$$\text{and } T \models \mathbf{T}_i^\alpha \vDash (\rho \to \sigma) \to \mathbf{T}_i^\alpha \vDash \rho \to \mathbf{T}_i^\alpha \vDash \sigma.$$

It follows that $T \models \mathbf{T}_i^\alpha \vDash \rho \to \mathbf{T}_i^\alpha \vDash \sigma$. The reverse implication is similar.

Claim 1: If $T \cap \alpha \models \phi$, $\mathbf{T}_i^\alpha \vDash \phi$ is an i-stratified sentence, and $\beta > \alpha$, then $T \cap \beta \models \mathbf{T}_i^\alpha \vDash \phi$.

Given $T \cap \alpha \models \phi$, there are $\sigma_1, \ldots, \sigma_n \in T \cap \alpha$ such that $\sigma_1 \to \cdots \to \sigma_n \to \phi$ is valid. By instances of i-Strativalidity and i-Stratideduction contained in $T \cap \beta$, $T \cap \beta \models \mathbf{T}_i^\alpha \vDash \phi$.

Claim 2: If ρ and σ are very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentences and $\rho^- \equiv \sigma^-$, then $T \models \rho \leftrightarrow \sigma$.

By induction on ρ. Note that ρ is not of the form $\mathbf{T}_j^\alpha \vDash \rho_0$ (with $j \neq i$), as that is not i-stratified. If ρ is $\mathbf{T}_j \vDash \rho_0$ then $\rho \equiv \rho^- \equiv \sigma^- \equiv \sigma$ and the claim is immediate.

The only nontrivial remaining case is when ρ is $\mathbf{T}_i^\alpha \vDash \rho_0$. Since ρ is very i-stratified, this implies $\alpha = \epsilon_0 \cdot n$ (some positive integer n) and ρ_0 is very i-stratified. Since $\sigma^- \equiv \rho^-$ and σ is very stratified, this implies $\sigma \equiv \mathbf{T}_i^{\epsilon_0 \cdot m} \vDash \sigma_0$ for some positive integer m and very i-stratified σ_0 with $\sigma_0^- \equiv \rho_0^-$. Assume $m \leq n$, the other case is similar.

By induction, $T \models \rho_0 \leftrightarrow \sigma_0$. By compactness, there is a natural $\ell \geq n$ such that $T \cap (\epsilon_0 \cdot \ell) \models \rho_0 \leftrightarrow \sigma_0$. By Claim 1, $T \models \mathbf{T}_i^{\epsilon_0 \cdot \ell} \vDash (\rho_0 \leftrightarrow \sigma_0)$; Claim 0 then gives $T \models \mathbf{T}_i^{\epsilon_0 \cdot \ell} \vDash \rho_0 \leftrightarrow \mathbf{T}_i^{\epsilon_0 \cdot \ell} \vDash \sigma_0$. The claim now follows since T contains i-Collapse and $\epsilon_0 \cdot m \leq_1 \epsilon_0 \cdot n \leq_1 \epsilon_0 \cdot \ell$ (Theorem 26 part 2).

Claim 3: If ϕ is an i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentence and $T \models \phi$, then $T^- \models \phi^-$.

By compactness, find $\sigma_1, \ldots, \sigma_n \in T$ such that $\sigma_1 \to \cdots \to \sigma_n \to \phi$ is valid. By Corollary 44, so is $\sigma_1^- \to \cdots \to \sigma_n^- \to \phi^-$, witnessing $T^- \models \phi^-$.

Claim 4: If ϕ is a very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentence and $T^- \models \phi^-$, then $T \models \phi$.

By compactness, there is a valid sentence

$$\Phi \equiv \sigma_1^- \to \cdots \to \sigma_n^- \to \phi^-$$

where each $\sigma_j \in T$. By Lemma 45, there is a valid very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentence Ψ such that $\Psi^- \equiv \Phi$. And because $\Psi^- \equiv \Phi$, this implies

$$\Psi \equiv \sigma_1^* \to \cdots \to \sigma_n^* \to \phi^*$$

where each $(\sigma_j^*)^- \equiv \sigma_j^-$, $(\phi^*)^- \equiv \phi^-$, and $\sigma_1^*, \ldots, \sigma_n^*, \phi^*$ are very i-stratified.

By Lemma 41, there are very i-stratified $\sigma_1^{**}, \ldots, \sigma_n^{**} \in T$ with each $(\sigma_j^{**})^- \equiv \sigma_j^- \equiv (\sigma_j^*)^-$. By Claim 2, $T \models \phi^* \leftrightarrow \phi$, and for $j = 1, \ldots, n$, $T \models \sigma_j^{**} \leftrightarrow \sigma_j^*$. Thus

$$T \models (\sigma_1^{**} \to \cdots \to \sigma_n^{**} \to \phi) \leftrightarrow \Psi,$$

and since Ψ is valid and the $\sigma_j^{**} \in T$, this shows $T \models \phi$. □

Definition 49. If $\mathbf{T} = (T_i)_{i \in \omega}$ is a straticlosed family of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories, its *stratification*, written $\mathrm{Str}(\mathbf{T})$, is the family $\mathrm{Str}(\mathbf{T}) = (S_i)_{i \in \mathcal{I}}$, where for every $i \in \omega$, $S_i = T_i^-$ and $\forall \alpha \in \epsilon_0 \cdot \omega$, $S_{(\alpha,i)} = T_i \cap \alpha$.

Theorem 50. (The Stratification Theorem) Suppose $\mathbf{T} = (T_i)_{i \in \omega}$ is a straticlosed family of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories. For any $i \in \omega$, any very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, and any assignment s, $\mathscr{M}_{\mathrm{Str}(\mathbf{T})} \models \phi[s]$ if and only if $\mathscr{M}_{\mathrm{Str}(\mathbf{T})} \models \phi^-[s]$.

Proof. By induction on ϕ. The only nontrivial case is when ϕ is $\mathbf{T}_i^\alpha \vDash \psi$. Since ϕ is very i-stratified, ψ is very i-stratified and we may write $\alpha = \epsilon_0 \cdot n$ for some positive integer n, $\mathrm{On}(\psi) \subseteq \epsilon_0 \cdot n$. The following are equivalent.

$$\begin{aligned}
\mathscr{M}_{\mathrm{Str}(\mathbf{T})} &\models \mathbf{T}_i^{\epsilon_0 \cdot n} \vDash \psi[s] \\
T_i \cap (\epsilon_0 \cdot n) &\models \psi^s && \text{(Definition of } \mathscr{M}_{\mathrm{Str}(\mathbf{T})}\text{)} \\
T_i &\models \psi^s && \text{(Theorem 38)} \\
T_i^- &\models (\psi^s)^- && \text{(Theorem 48)} \\
T_i^- &\models (\psi^-)^s && \text{(Clearly } (\psi^s)^- \equiv (\psi^-)^s\text{)} \\
\mathscr{M}_{\mathrm{Str}(\mathbf{T})} &\models \mathbf{T}_i \vDash \psi^-[s]. && \text{(Definition of } \mathscr{M}_{\mathrm{Str}(\mathbf{T})}\text{)}
\end{aligned}$$

□

6 Stratifiers

In order to apply theorems from the previous section, it is necessary to work with families $\mathbf{T} = (T_i)_{i \in \omega}$ where each T_i is i-stratified. If we want T_i^- to (locally) express the truthfulness of T_j^-, we cannot simply add a schema like $\mathrm{ucl}(\mathbf{T}_j \vDash \phi \to \phi)$ to T_i, because this is not necessarily i-stratified: for example, the particular instance $\mathbf{T}_j \vDash \mathbf{T}_i \vDash (1=0) \to \mathbf{T}_i \vDash (1=0)$ is not i-stratified. But neither is, say, $\mathbf{T}_j \vDash \mathbf{T}_i^\alpha \vDash (1=0) \to \mathbf{T}_i^\alpha \vDash (1=0)$, where $\mathbf{T}_i^\alpha \vDash$ occurs within the scope of $\mathbf{T}_j \vDash$. We will use a schema $\mathrm{ucl}(\mathbf{T}_j \vDash \phi \to \phi^+)$, where $\bullet^+$ varies over what we call i-stratifiers.

Definition 51. Suppose $X \subseteq \epsilon_0 \cdot \omega$, $|X| = \infty$, and $i \in \omega$. The *i-stratifier given by X* is the function $\phi \mapsto \phi^+$ taking $\mathscr{L}_{\mathrm{PA}}(\omega)$-formulas to $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formulas as follows.

1. If ϕ is atomic or of the form $\mathbf{T}_j \vDash \phi_0$ with $j \neq i$, then $\phi^+ \equiv \phi$.
2. If ϕ is $\mathbf{T}_i \vDash \phi_0$ then $\phi^+ \equiv \mathbf{T}_i^\alpha \vDash \phi_0^+$ where $\alpha = \min\{x \in X \,:\, x > \mathrm{On}(\phi_0^+)\}$.
3. If ϕ is $\neg\psi$, $\psi \to \rho$, or $\forall x \psi$, then ϕ^+ is $\neg\psi^+$, $\psi^+ \to \rho^+$ or $\forall x \psi^+$, respectively.

By an *i-stratifier* we mean an i-stratifier given by some X. By the *i-veristratifier* we mean the i-stratifier given by $X = \{\epsilon_0 \cdot 1, \epsilon_0 \cdot 2, \ldots\}$.

For example, if $\bullet^+$ is the i-veristratifier and $j \neq i$ then

$$(\mathbf{T}_j \vDash \mathbf{T}_i \vDash (1=0) \to \mathbf{T}_i \vDash \mathbf{T}_i \vDash (1=0))^+ \quad \equiv \quad \mathbf{T}_j \vDash \mathbf{T}_i \vDash (1=0) \to \mathbf{T}_i^{\epsilon_0 \cdot 2} \vDash \mathbf{T}_i^{\epsilon_0} \vDash (1=0).$$

Lemma 52. Suppose $Z \subseteq \epsilon_0 \cdot \omega$, $h : Z \to \epsilon_0 \cdot \omega$ is order preserving, $i \in \omega$, and $\bullet^+$ is an i-stratifier. For any $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula θ with $\mathrm{On}(\theta^+) \subseteq Z$, there is a computable i-stratifier $\bullet^*$ with $\theta^* \equiv h(\theta^+)$.

Proof. Let $X_0 = \{h(\alpha) \,:\, \alpha \in \mathrm{On}(\theta^+)\}$, let $X = X_0 \cup \{\alpha \in \epsilon_0 \cdot \omega \,:\, \alpha > X_0\}$, and let $\bullet^*$ be the i-stratifier given by X. By induction, for every subformula θ_0 of θ, $\theta_0^* \equiv h(\theta_0^+)$. □

Definition 53. By a *stratifier-set*, we mean a finite set

$$I = \{\bullet^{+_1}, \ldots, \bullet^{+_k}\}$$

where each $\bullet^{+_p}$ is an i_p-stratifier for some $i_p \in \omega$, and $i_1, \ldots, i_k$ are distinct. With I as above, we write $\mathrm{Indices}(I)$ for $\{i_1, \ldots, i_k\}$. We say I is *computable* if each $\bullet^{+_p}$ is computable.

For example, if $\bullet^{+_1}$ is a 1-stratifier, $\bullet^{+_2}$ is a 5-stratifier, and $\bullet^{+_3}$ is a 2-stratifier, then $I = \{\bullet^{+_1}, \bullet^{+_2}, \bullet^{+_3}\}$ is a stratifier-set and $\mathrm{Indices}(I) = \{1, 5, 2\}$. For a non-example, if $\bullet^{*_1}$ and $\bullet^{*_2}$ are distinct 1-stratifiers, then $\{\bullet^{*_1}, \bullet^{*_2}\}$ is not a stratifier-set, because it fails the distinctness condition.

Definition 54. 1. Suppose $\mathscr{N}$ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure and I is a stratifier-set. We define an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}^I$ as follows. The universe and interpretation of arithmetic of $\mathscr{N}^I$ agree with those of $\mathscr{N}$, as do the interpretations of $\mathbf{T}_i\vDash$ ($i \notin \mathrm{Indices}(I)$) and $\mathbf{T}_i^\alpha\vDash$ (any α, i). For each $i \in \mathrm{Indices}(I)$, let $\bullet^+ \in I$ be the corresponding i-stratifier, and let $\mathscr{N}^I$ interpret $\mathbf{T}_i\vDash$ as follows. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, we consider two cases.

   (a) If ϕ is an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula, then $\mathscr{N}^I \models \mathbf{T}_i\vDash\phi[s]$ if and only if $\mathscr{N} \models (\mathbf{T}_i\vDash\phi)^+[s]$.

   (b) If ϕ is not an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula, then $\mathscr{N}^I \models \mathbf{T}_i\vDash\phi[s]$ if and only if $\mathscr{N} \models \mathbf{T}_i\vDash\phi[s]$.

2. For any $i \in \omega$, any i-stratifier $\bullet^+$, and any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}$, let $\mathscr{N}^+ = \mathscr{N}^I$ where $I = \{\bullet^+\}$ is the stratifier-set containing only $\bullet^+$.

Case 1b in Definition 54 is somewhat arbitrary. We will only ever really care about whether $\mathscr{N}^I \models \mathbf{T}_i\vDash\phi[s]$ when $\mathbf{T}_i\vDash\phi$ is j-stratified for some j. If ϕ is not an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula then $\mathbf{T}_i\vDash\phi$ is not j-stratified for any j.

Lemma 55. (Compare Lemma 43) Suppose $\mathscr{N}$ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure, $i \in \omega$, and $\bullet^+$ is an i-stratifier. For every $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ and assignment s, $\mathscr{N}^+ \models \phi[s]$ if and only if $\mathscr{N} \models \phi^+[s]$.

Proof. By induction. □

Lemma 56. For any $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, any $i \in \omega$, and any i-stratifier $\bullet^+$, ϕ is valid if and only if ϕ^+ is valid.

Proof.

($\Rightarrow$) Assume ϕ is valid. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}$ and assignment s, $\mathscr{N}^+ \models \phi[s]$ by validity, so $\mathscr{N} \models \phi^+[s]$ by Lemma 55.

($\Leftarrow$) By Corollary 44. □

Lemma 57. Suppose $\mathscr{M}$ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure, I_0 is a stratifier-set, $i \in \omega$, $i \notin \mathrm{Indices}(I_0)$, and $\bullet^+$ is an i-stratifier. Let $I = I_0 \cup \{\bullet^+\}$. Then $\mathscr{M}^I = (\mathscr{M}^{I_0})^+$. Furthermore, $\mathscr{M}^+$ and $\mathscr{M}^I$ agree on the interpretation of $\mathbf{T}_i\vDash$.

Proof. Straightforward. □

Lemma 58. Suppose $i \in \omega$ and suppose $\mathscr{M}$ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure with the property that for every very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^-[s]$. Suppose I is a stratifier-set such that $i \notin \mathrm{Indices}(I)$. Then for every very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{M}^I \models \phi[s]$ if and only if $\mathscr{M}^I \models \phi^-[s]$.

Proof. By induction on ϕ. Let s be an assignment. The only interesting cases are the following.

Case 1: ϕ is $\mathbf{T}_j\vDash\psi$ for some j. Then $\phi^- \equiv \phi$ and the claim is trivial.

Case 2: ϕ has the form $\mathbf{T}_j^\alpha\vDash\psi$ for some $j \neq i$. Impossible, this is not i-stratified.

Case 3: ϕ has the form $\mathbf{T}_i^\alpha\vDash\psi$. The following are equivalent:

$$\begin{aligned}
&\mathscr{M}^I \models \mathbf{T}_i^\alpha\vDash\psi[s] && \\
&\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\psi[s] && (\mathscr{M} \text{ and } \mathscr{M}^I \text{ agree on } \mathbf{T}_i^\alpha\vDash) \\
&\mathscr{M} \models (\mathbf{T}_i^\alpha\vDash\psi)^-[s] && (\text{By hypothesis}) \\
&\mathscr{M}^I \models (\mathbf{T}_i^\alpha\vDash\psi)^-[s]. && (\text{Since } i \notin \mathrm{Indices}(I), \mathscr{M} \text{ and } \mathscr{M}^I \text{ agree on } \mathbf{T}_i\vDash)
\end{aligned}$$

□

Lemma 59. Suppose $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{M}$ is an instance of Definition 7, and suppose I is a stratifier-set. Then $\mathscr{M}^I$ interprets formulas by substitution.

Proof. By induction on $|I|$. If $|I| = 0$, we are done by Lemma 8. Otherwise, we may decompose I as $I = I_0 \cup \{\bullet^+\}$ where $\bullet^+$ is an i-stratifier. By induction, $\mathscr{M}^{I_0}$ interprets formulas by substitution $(*)$. By Lemma 57, $\mathscr{M}^I = (\mathscr{M}^{I_0})^+$.

By definition of interpreting formulas by substitution, for every $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{M}^{I_0} \models \phi[s]$ if and only if $\mathscr{M}^{I_0} \models \phi^s$. We must show that for every such ϕ and s, $(\mathscr{M}^{I_0})^+ \models \phi[s]$ if and only if $(\mathscr{M}^{I_0})^+ \models \phi^s$.

We induct on ϕ. By Definition 54, $(\mathscr{M}^{I_0})$ and $(\mathscr{M}^{I_0})^+$ agree on all symbols except $\mathbf{T}_i\vDash$, and they agree on $\mathbf{T}_i\vDash\phi_0$ if ϕ_0 is not an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula. Thus the only nontrivial case is when ϕ is of the form $\mathbf{T}_i\vDash\phi_0$ for some $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ_0. Any such ϕ is itself an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula and thus susceptible to Lemma 55. The following are equivalent.

$$
\begin{aligned}
&(\mathscr{M}^{I_0})^+ \models \phi[s] && \\
&\quad \mathscr{M}^{I_0} \models \phi^+[s] && \text{(Lemma 55)} \\
&\quad \mathscr{M}^{I_0} \models (\phi^+)^s && \text{(By } (*)\text{)} \\
&\quad \mathscr{M}^{I_0} \models (\phi^s)^+ && \text{(Clearly } (\phi^+)^s \equiv (\phi^s)^+\text{)} \\
&(\mathscr{M}^{I_0})^+ \models \phi^s. && \text{(Lemma 55)}
\end{aligned}
$$

□

7 Generic Stratified Axioms

We now have enough technical machinery to fulfill the second promise from the Introduction. We will fulfill it in a general way, essentially saying: "The theories in question, whose truth were in doubt, are true together with any background theory of provability such that..." Just like in Section 3, we do this by introducing a notion of genericness. Throughout this section, $\prec$ is an r.e. well-founded partial-order of ω.

Definition 60. If $i \in \omega$, we say that a stratifier-set I is *above* i if $\forall j \in \mathrm{Indices}(I)$, $i \prec j$. We adopt the following convention: if I is above i then we will write I as $I(i)$ in order to remind ourselves that I is above i.

Definition 61. (Compare Definition 14) Suppose $\mathbf{T} = (T_i)_{i\in\omega}$ is an r.e. family of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories and each T_i is i-unistratified. We say $\mathbf{T}$ is *$\prec$-straticlosed-r.e.-generic* (or *straticlosed-r.e.-generic*, if $\prec$ is clear from context) if for every straticlosed r.e. family $\mathbf{U} \supseteq \mathbf{T}$, every $i \in \omega$, and every computable stratifier-set $I(i)$ above i, $\mathscr{M}^{I(i)}_{\mathrm{Str}(\mathbf{U})} \models T_i$.

Lemma 62. If the family $\mathbf{T} = (T_i)_{i\in\omega}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sets is r.e. and is a union of straticlosed-r.e.-generic families, then $\mathbf{T}$ is straticlosed-r.e.-generic.

Proof. Straightforward. □

7.1 Straticlosed-r.e.-generic Building Blocks

As in Section 3.1, we exhibit some examples of straticlosed-r.e.-generic families, which can be combined (via Lemma 62) to form background theories of provability. This will allow us to state Theorem 72 below in a generalized way, essentially saying that certain doubted theories are consistent with *any* background theory of provability built up from such blocks. This saves us from having to arbitrarily impose any particular background theory of provability.

In the following lemma, for part 3, the intuition is that the things T_i says about T_j need to be generic enough that they even remain true when a j-stratifier is applied to them. $\mathbf{T}_j\vDash(\phi \to \psi) \to \mathbf{T}_j\vDash\phi \to \mathbf{T}_j\vDash\psi$ is not so generic, because it could be that $(\mathbf{T}_j\vDash\phi)^+ \equiv \mathbf{T}_j^{\alpha}\vDash\phi^+$, $(\mathbf{T}_j\vDash\psi)^+ \equiv \mathbf{T}_j^{\beta}\vDash\psi^+$, where $\beta < \alpha$. For parts 1–2, the reason we cannot merge these parts into $[j\text{-Deduction}]_i$ ($j \not\prec i$) is because $[i\text{-Deduction}]_i$ is not i-stratified.

Lemma 63. (Compare Lemma 17) For any $i, j \in \omega$, each of the following families is straticlosed-r.e.-generic.

1. $[i$-Stratideduction$]_i$.
2. $[j$-Deduction$]_i$ (if $j \not\succeq i$).
3. $[S]_i$ (if $i \prec j$) where S is the following schema (ϕ, ψ range over $\mathscr{L}_{\mathrm{PA}}(\omega)$-formulas):

$$\text{(Weak } j\text{-Deduction) } \mathrm{ucl}(\mathbf{T}_j{\vDash}(\phi \to \psi) \to \mathbf{T}_j{\vDash}\phi \to \mathbf{T}_j{\vDash}(\psi \wedge (\phi \vee \neg\phi))).$$

Proof. Clearly these families are unistratified. Recursive enumerability follows from the fact that $\prec$ is r.e. In each case below, let $\mathbf{U} = (U_k)_{k\in\omega}$ be a straticlosed r.e. family extending the family in question. For brevity, let $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$.

(1) Let $I(i)$ be any computable stratifier-set above i, we must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}_i^\alpha{\vDash}(\phi \to \psi) \to \mathbf{T}_i^\alpha{\vDash}\phi \to \mathbf{T}_i^\alpha{\vDash}\psi)$ assuming this formula is i-stratified. Let s be an assignment and assume $\mathscr{M}^{I(i)} \models \mathbf{T}_i^\alpha{\vDash}(\phi \to \psi)[s]$ and $\mathscr{M}^{I(i)} \models \mathbf{T}_i^\alpha{\vDash}\phi[s]$. By Definition 54, $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_i^\alpha{\vDash}$, so $\mathscr{M} \models \mathbf{T}_i^\alpha{\vDash}(\phi \to \psi)[s]$. By definition of $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$, this means $U_i \cap \alpha \models (\phi \to \psi)^s$. Clearly $(\phi \to \psi)^s \equiv \phi^s \to \psi^s$, so $U_i \cap \alpha \models \phi^s \to \psi^s$. By similar reasoning, $U_i \cap \alpha \models \phi^s$. By modus ponens, $U_i \cap \alpha \models \psi^s$, which means $\mathscr{M} \models \mathbf{T}_i^\alpha{\vDash}\psi[s]$. Since $\mathscr{M}$ and $\mathscr{M}^{I(i)}$ agree on $\mathbf{T}_i^\alpha{\vDash}$, $\mathscr{M}^{I(i)} \models \mathbf{T}_i^\alpha{\vDash}\psi[s]$, as desired.

(2) Let $I(i)$ be any computable stratifier-set above i, we must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}_j{\vDash}(\phi \to \psi) \to \mathbf{T}_j{\vDash}\phi \to \mathbf{T}_j{\vDash}\psi)$.

Let s be an assignment and assume $\mathscr{M}^{I(i)} \models \mathbf{T}_j{\vDash}(\phi \to \psi)[s]$ and $\mathscr{M}^{I(i)} \models \mathbf{T}_j{\vDash}\phi[s]$. Since $I(i)$ is above i and $j \not\succeq i$, $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_j{\vDash}$, so $\mathscr{M} \models \mathbf{T}_j{\vDash}(\phi \to \psi)[s]$ and $\mathscr{M} \models \mathbf{T}_j{\vDash}\phi[s]$. By definition of $\mathscr{M}$, $U_j^- \models \phi^s \to \psi^s$ and $U_j^- \models \phi^s$, thus $U_j^- \models \psi^s$, so $\mathscr{M} \models \mathbf{T}_j{\vDash}\psi[s]$ and thus so does $\mathscr{M}^{I(i)}$.

(3) Let $I(i)$ be any computable stratifier-set above i, we must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}_j{\vDash}(\phi \to \psi) \to \mathbf{T}_j{\vDash}\phi \to \mathbf{T}_j{\vDash}(\psi \wedge (\phi \vee \neg\phi)))$. Let s be an assignment and assume $\mathscr{M}^{I(i)} \models \mathbf{T}_j{\vDash}(\phi \to \psi)[s]$ and $\mathscr{M}^{I(i)} \models \mathbf{T}_j{\vDash}\phi[s]$. If $j \notin \mathrm{Indices}(I(i))$, then $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_j{\vDash}$, so reason as in (2) above. If not, we can write $I(i) = I_0 \cup \{\bullet^+\}$ where $\bullet^+$ is a computable j-stratifier, and Lemma 57 ensures that $\mathscr{M}^{I(i)}$ and $\mathscr{M}^+$ agree on $\mathbf{T}_j{\vDash}$. By definition of $\mathscr{M}^+$, $\mathscr{M} \models (\mathbf{T}_j{\vDash}(\phi \to \psi))^+[s]$ and $\mathscr{M} \models (\mathbf{T}_j{\vDash}\phi)^+[s]$. Let $\alpha, \beta \in \epsilon_0 \cdot \omega$ be such that $(\mathbf{T}_j{\vDash}(\phi \to \psi))^+ \equiv \mathbf{T}_j^\alpha{\vDash}(\phi^+ \to \psi^+)$ and $(\mathbf{T}_j{\vDash}\phi)^+ \equiv \mathbf{T}_j^\beta{\vDash}\phi^+$. Then $\mathscr{M} \models \mathbf{T}_j^\alpha{\vDash}(\phi^+ \to \psi^+)[s]$ and $\mathscr{M} \models \mathbf{T}_j^\beta{\vDash}\phi^+[s]$. This means $U_j \cap \alpha \models (\phi^+ \to \psi^+)^s$ and $U_j \cap \beta \models (\phi^+)^s$. Since ϕ is a subformula of $\phi \to \psi$, it follows $\beta \leq \alpha$, thus $U_j \cap \alpha \models (\psi^+)^s$, and by tautology, $U_j \cap \alpha \models (\psi^+ \wedge (\phi^+ \vee \neg\phi^+))^s$. So $\mathscr{M} \models \mathbf{T}_j^\alpha{\vDash}(\psi^+ \wedge (\phi^+ \vee \neg\phi^+))[s]$. By Definition 51,

$$\mathbf{T}_j^\alpha{\vDash}(\psi^+ \wedge (\phi^+ \vee \neg\phi^+)) \equiv (\mathbf{T}_j{\vDash}(\psi \wedge (\phi \vee \neg\phi)))^+$$

(this is the reason for the $\phi \vee \neg\phi$ clause) and finally $\mathscr{M}^+ \models \mathbf{T}_j{\vDash}(\psi \wedge (\phi \vee \neg\phi))[s]$. □

In Lemma 18, we introduced Assigned Validity as a single schema for inclusion in T_i for any i. In the following lemma, we need to break the stratified version of Assigned Validity into different ω-indexed families because the stratified version of Assigned Validity intended for inclusion in T_i (for any particular i) needs to be i-stratified.

Lemma 64. (Compare Lemma 18) For any $i, j \in \omega$, each of the following families is straticlosed-r.e.-generic.

1. $[S]_i$ where S is: (i-Assigned Strativalidity) the schema ϕ^s (ϕ valid and i-stratified, s an assignment).
2. $[i$-Assigned Strativalidity$]_i \cup [i$-Strativalidity$]_i$.
3. $[i$-Assigned Strativalidity$]_i \cup [i$-Validity$]_j$ (if $j \neq i$).

Proof. For unistratifiedness, use Corollary 34. Recursive enumerability follows from the fact that $\prec$ is r.e. In each case below, let $\mathbf{U} = (U_k)_{k\in\omega}$ be a straticlosed r.e. family extending the family in question. For brevity, let $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$.

(1) Let $I(i)$ be any computable stratifier-set above i, let ϕ be any valid i-stratified formula, and let s be any assignment. Since ϕ is valid, $\mathscr{M}^{I(i)} \models \phi[s]$. By Lemma 59, $\mathscr{M}^{I(i)} \models \phi^s$, as desired.

(2) Let $I(i)$ be any computable stratifier-set above i. By (1), $\mathscr{M}^{I(i)} \models i$-Assigned Strativalidity. We must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi)$, where ϕ is any valid $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula and $\alpha < \epsilon_0 \cdot \omega$ is any ordinal such that $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified. Let s be any assignment. Since U_i contains i-Assigned Strativalidity, in particular U_i contains ϕ^s. Since $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified, α exceeds all the superscripts in ϕ (hence in ϕ^s), so $U_i \cap \alpha \models \phi^s$. By definition of $\mathscr{M}$, this means $\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\phi[s]$. By Definition 54, $\mathscr{M}$ and $\mathscr{M}^{I(i)}$ agree on $\mathbf{T}_i^\alpha\vDash$, so $\mathscr{M}^{I(i)} \models \mathbf{T}_i^\alpha\vDash\phi[s]$, as desired.

(3) By (1), $\mathscr{M}^{I(i)} \models i$-Assigned Strativalidity for every computable stratifier-set $I(i)$ above i. Let $J(j)$ be a computable stratifier-set above j, we must show $\mathscr{M}^{J(j)} \models i$-Validity. Let ϕ be a valid $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula, s an assignment.

Case 1: $i \notin \mathrm{Indices}(J(j))$. Then $\mathscr{M}^{J(j)}$ and $\mathscr{M}$ agree on $\mathbf{T}_i\vDash$. Let $\bullet^+$ be an i-stratifier. Since ϕ is valid, so is ϕ^+ (by Lemma 56), so $(\phi^+)^s \in U_i$ (since $[i\text{-Assigned Strativalidity}]_i$ is part of line 3). Clearly $((\phi^+)^s)^- \equiv \phi^s$, so $\phi^s \in U_i^-$, thus $\mathscr{M} \models \mathbf{T}_i\vDash\phi[s]$, and so does $\mathscr{M}^{J(j)}$.

Case 2: $i \in \mathrm{Indices}(J(j))$. Thus $j \prec i$ and we can write $J(j) = J_0 \cup \{\bullet^+\}$ for some computable i-stratifier $\bullet^+$. By Lemma 57, $\mathscr{M}^{J(j)}$ and $\mathscr{M}^+$ agree on $\mathbf{T}_i\vDash$. Let $\alpha \in \epsilon_0 \cdot \omega$ be such that $(\mathbf{T}_i\vDash\phi)^+ \equiv \mathbf{T}_i^\alpha\vDash\phi^+$. As in Case 1, $(\phi^+)^s$ is an instance of i-Assigned Strativalidity, so $(\phi^+)^s \in U_i$ (since $[i\text{-Assigned Strativalidity}]_i$ is part of line 3). In fact by choice of α, $(\phi^+)^s \in U_i \cap \alpha$, so $\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\phi^+[s]$, that is, $\mathscr{M} \models (\mathbf{T}_i\vDash\phi)^+[s]$. By Lemma 55, $\mathscr{M}^+ \models \mathbf{T}_i\vDash\phi[s]$. Since $\mathscr{M}^{J(j)}$ and $\mathscr{M}^+$ agree on $\mathbf{T}_i\vDash$, $\mathscr{M}^{J(j)} \models \mathbf{T}_i\vDash\phi[s]$. □

In Lemma 63 above, we had to weaken what T_i says about j-Deduction for $j \succ i$. No such weakening is needed in the following lemma. This is interesting because in modal logic, positive introspection is generally considered much more controversial and demanding than basic deduction.

Lemma 65. (Compare Lemma 19) For any $i, j \in \omega$, each of the following families is straticlosed-r.e.-generic.

1. $[i\text{-Assigned Strativalidity}]_i \cup [i\text{-Strativalidity}]_i \cup [i\text{-Stratideduction}]_i \cup [i\text{-Introspection}]_j$ $(j \neq i)$.
2. $[i\text{-Assigned Strativalidity}]_i \cup [i\text{-Strativalidity}]_i \cup [i\text{-Stratideduction}]_i \cup [S]_i$ where S is:

$$(i\text{-Stratrospection})\ \mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi \to \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi) \text{ whenever this is } i\text{-stratified.}$$

Proof. For unistratifiedness, use Corollary 34. Recursive enumerability follows from the fact that $\prec$ is r.e. In each case below, let $\mathbf{U} = (U_k)_{k\in\omega}$ be a straticlosed r.e. family extending the family in question. For brevity, let $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$.

(1) By Lemma 63 (part 1) and Lemma 64 (part 2), $\mathscr{M}^{I(i)} \models (i\text{-Assigned Strativalidity}) \cup (i\text{-Strativalidity}) \cup (i\text{-Stratideduction})$ for every computable stratifier-set $I(i)$ above i. Let $J(j)$ be a computable stratifier-set above j, we must show $\mathscr{M}^{J(j)} \models i$-Introspection. In other words, we must show $\mathscr{M}^{J(j)} \models \mathrm{ucl}(\mathbf{T}_i\vDash\phi \to \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi)$ for any $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ. Let s be any assignment and assume $\mathscr{M}^{J(j)} \models \mathbf{T}_i\vDash\phi[s]$.

Case 1: $i \notin \mathrm{Indices}(J(j))$. Then $\mathscr{M}^{J(j)}$ and $\mathscr{M}$ agree on $\mathbf{T}_i\vDash$. Thus $\mathscr{M} \models \mathbf{T}_i\vDash\phi[s]$. Let $\bullet^+$ be the veristratifier. By Theorem 50, $\mathscr{M} \models (\mathbf{T}_i\vDash\phi)^+[s]$. Let α be such that $(\mathbf{T}_i\vDash\phi)^+ \equiv \mathbf{T}_i^\alpha\vDash\phi^+$, so $\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\phi^+[s]$. By definition, this means $U_i \cap \alpha \models (\phi^+)^s$. Let β be such that $(\mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi)^+ \equiv \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi^+$, so $\beta > \alpha$. By Part 1 of Theorem 48, $U_i \cap \beta \models \mathbf{T}_i^\alpha\vDash(\phi^+)^s$. Thus $\mathscr{M} \models \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi^+[s]$. By Theorem 50, $\mathscr{M} \models (\mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi^+)^-[s]$, that is, $\mathscr{M} \models \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi[s]$. Since $\mathscr{M}$ and $\mathscr{M}^{J(j)}$ agree on $\mathbf{T}_i\vDash$, $\mathscr{M}^{J(j)} \models \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi[s]$, as desired.

Case 2: $i \in \mathrm{Indices}(J(j))$. Thus $j \prec i$ and we can write $J(j) = J_0 \cup \{\bullet^+\}$ for some computable i-stratifier $\bullet^+$. By Lemma 57, $\mathscr{M}^{J(j)}$ and $\mathscr{M}^+$ agree on $\mathbf{T}_i\vDash$. Thus $\mathscr{M}^+ \models \mathbf{T}_i\vDash\phi[s]$. By Lemma 55, $\mathscr{M} \models (\mathbf{T}_i\vDash\phi)^+[s]$. Let α be such that $(\mathbf{T}_i\vDash\phi)^+ \equiv \mathbf{T}_i^\alpha\vDash\phi^+$, so $\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\phi^+[s]$. By definition of $\mathscr{M}$, this means $U_i \cap \alpha \models (\phi^+)^s$. Let β be such that $(\mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi)^+ \equiv \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi^+$, so $\beta > \alpha$. By Part 1 of Theorem 48, $U_i \cap \beta \models \mathbf{T}_i^\alpha\vDash(\phi^+)^s$. Thus $\mathscr{M} \models \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi^+[s]$. In other words, $\mathscr{M} \models (\mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi)^+[s]$. By Lemma 55, $\mathscr{M}^+ \models \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi[s]$. Since $\mathscr{M}^+$ and $\mathscr{M}^{J(j)}$ agree on $\mathbf{T}_i\vDash$, $\mathscr{M}^{J(j)} \models \mathbf{T}_i\vDash\mathbf{T}_i\vDash\phi[s]$, as desired.

(2) Let $I(i)$ be any computable stratifier-set above i, we must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi \to \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi)$ assuming this is i-stratified (so $\beta > \alpha$). Let s be any assignment and assume $\mathscr{M}^{I(i)} \models \mathbf{T}_i^\alpha\vDash\phi[s]$. By Definition 54, $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_i^\alpha\vDash$, so $\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\phi[s]$. By definition of $\mathscr{M}$, this means $U_i \cap \alpha \models \phi^s$. By Part 1 of Theorem 48, $U_i \cap \beta \models \mathbf{T}_i^\alpha\vDash\phi^s$. Thus, $\mathscr{M} \models \mathbf{T}_i^\beta\vDash\mathbf{T}_i^\alpha\vDash\phi[s]$, and thus so does $\mathscr{M}^{I(i)}$ since it agrees with $\mathscr{M}$ on $\mathbf{T}_i^\beta\vDash$. □

For the next lemma, note that the proof shows more than is necessary, namely that the structures in question satisfy all the axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$, not just the i-stratified ones. But of course, the full set of Peano axioms for $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$ is not i-stratified.

Lemma 66. (Compare Lemma 20) For any $i \in \omega$, $[S]_i$ is straticlosed-r.e.-generic, where S is the set of those axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$ that are i-stratified.

Proof. Unistratifiedness and recursive enumerability are clear. Let $\mathbf{U} = (U_k)_{k\in\omega}$ be a straticlosed r.e. family extending $[S]_i$. By Lemma 59, $\mathscr{M}^{I(i)}_{\mathrm{Str}(\mathbf{U})}$ interprets formulas by substitution. By Lemma 10, $\mathscr{M}^{I(i)}_{\mathrm{Str}(\mathbf{U})}$ satisfies the axioms of Peano Arithmetic for $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$, as desired. □

Lemma 67. (Compare Lemma 21) For any $i, j \in \omega$, each of the following families is straticlosed-r.e.-generic.

1. $[j\text{-SMT}]_i$ $(j \neq i)$.
2. $[S]_i$, where S is: (i-Strati-SMT) $\mathrm{ucl}(\exists e \forall x(\mathbf{T}^\alpha_i \vDash \phi \leftrightarrow x \in W_e))$ when this is i-stratified, $e \notin \mathrm{FV}(\phi)$.

Proof. Unistratifiedness and recursive enumerability are clear. In each case below, let $\mathbf{U} = (U_k)_{k\in\omega}$ be a straticlosed r.e. family extending the family in question. For brevity, let $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$.

(1) Let $I(i)$ be any computable stratifier-set above i. We must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\exists e \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e))$ for every $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ with $e \notin \mathrm{FV}(\phi)$. Let s be an assignment and let $\{x_1, \ldots, x_k\} = \mathrm{FV}(\phi)\backslash\{x\}$.

Case 1: $j \notin \mathrm{Indices}(I(i))$. Then $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_j\vDash$. Since U_j^- is r.e., by the S-m-n theorem there is some n such that $W_n = \{m : U_j^- \models \phi(x|\overline{m})(x_1|\overline{s(x_1)}) \cdots (x_k|\overline{s(x_k)})\}$. Since $e \notin \mathrm{FV}(\phi)$ and $\mathscr{M}$ has standard first-order part, it follows that $\mathscr{M} \models \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e)[s(e|n)]$. By first-order semantics, $\mathscr{M} \models \exists e \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e)[s]$. Since $\mathscr{M}$ and $\mathscr{M}^{I(i)}$ agree on $\mathbf{T}_j\vDash$, $\mathscr{M}^{I(i)} \models \exists e \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e)[s]$, as desired.

Case 2: $j \in \mathrm{Indices}(I(i))$. Thus $i \prec j$ and we can write $I(i) = I_0 \cup \{\bullet^+\}$ for some computable j-stratifier $\bullet^+$. By Lemma 57, $\mathscr{M}^{I(i)}$ and $\mathscr{M}^+$ agree on $\mathbf{T}_j\vDash$. Let α be such that $(\mathbf{T}_j \vDash \phi)^+ \equiv \mathbf{T}^\alpha_j \vDash \phi^+$. Since $U_j \cap \alpha$ is r.e., by the S-m-n theorem there is some n such that $W_n = \{m : U_j \cap \alpha \models \phi^+(x|\overline{m})(x_1|\overline{s(x_1)}) \cdots (x_k|\overline{s(x_k)})\}$. Since $e \notin \mathrm{FV}(\phi)$ (thus $e \notin \mathrm{FV}(\phi^+)$), and since $\mathscr{M}$ has standard first-order part, it follows that $\mathscr{M} \models \forall x(\mathbf{T}^\alpha_j \vDash \phi^+ \leftrightarrow x \in W_e)[s(e|n)]$. By first-order semantics, $\mathscr{M} \models \exists e \forall x(\mathbf{T}^\alpha_j \vDash \phi^+ \leftrightarrow x \in W_e)[s]$. In other words, $\mathscr{M} \models (\exists e \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e))^+[s]$. By Lemma 55, $\mathscr{M}^+ \models \exists e \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e)[s]$. Since $\mathscr{M}^+$ and $\mathscr{M}^{I(i)}$ agree on $\mathbf{T}_j\vDash$, $\mathscr{M}^{I(i)} \models \exists e \forall x(\mathbf{T}_j \vDash \phi \leftrightarrow x \in W_e)[s]$, as desired.

(2) Let $I(i)$ be any computable stratifier-set above i, we must show $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\exists e \forall x(\mathbf{T}^\alpha_i \vDash \phi \leftrightarrow x \in W_e))$ for every $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ such that this is i-stratified and $e \notin \mathrm{FV}(\phi)$. Let s be any assignment and let $\{x_1, \ldots, x_k\} = \mathrm{FV}(\phi)\backslash\{x\}$. Since $U_i \cap \alpha$ is r.e., by the S-m-n theorem there is some n such that $W_n = \{m : U_i \cap \alpha \models \phi(x|\overline{m})(x_1|\overline{s(x_1)}) \cdots (x_k|\overline{s(x_k)})\}$. Since $e \notin \mathrm{FV}(\phi)$, and since $\mathscr{M}$ has standard first-order part, it follows that $\mathscr{M} \models \exists e \forall x(\mathbf{T}^\alpha_i \vDash \phi \leftrightarrow x \in W_e)[s]$. By Definition 54, $\mathscr{M}$ and $\mathscr{M}^{I(i)}$ agree on $\mathbf{T}^\alpha_i\vDash$, so $\mathscr{M}^{I(i)} \models \exists e \forall x(\mathbf{T}^\alpha_i \vDash \phi \leftrightarrow x \in W_e)[s]$, as desired. □

If $\mathbf{T} = (T_k)_{k\in\omega}$ is straticlosed-r.e.-generic, we cannot simply take an axiom ϕ from T_j and insert $\mathbf{T}_j\vDash\phi$ into T_i without violating straticlosed-r.e.-genericness, because such a ϕ is not necessarily i-stratified. Thus, the following lemma has a somewhat more complicated structure than Lemma 22.

Lemma 68. (Compare Lemma 22) Let $i, j \in \omega$ and suppose $\mathbf{T} = (T_k)_{k\in\omega}$ is straticlosed-r.e.-generic. Then each of the following families is straticlosed-r.e.-generic.

1. $\mathbf{T} \cup [S]_i$ where S is the schema $\mathbf{T}^\alpha_i \vDash \phi$ ($\phi \in T_i$ such that this is i-stratified).
2. $\mathbf{T} \cup [S]_i$ where S is the schema $\mathbf{T}_j \vDash \phi^-$ ($\phi \in T_j$, $j \not\succeq i$).

Proof. Unistratifiedness and recursive enumerability are clear. In each case below, let $\mathbf{U} = (U_k)_{k\in\omega}$ be a straticlosed r.e. family extending the family in question. For brevity, let $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$.

(1) Since $\mathbf{T}$ is straticlosed-r.e.-generic and $\mathbf{U} \supseteq \mathbf{T}$ is straticlosed and r.e., immediately $\mathscr{M}^{J(j)} \models \mathbf{T}$ (by Definition 61) for all $j \in \omega$ and any computable stratifier-set $J(j)$ above j. Let $I(i)$ be any computable stratifier-set above i. Suppose $\phi \in T_i$ and $\alpha \in \epsilon_0 \cdot \omega$ are such that $\mathbf{T}_i^{\alpha}\vDash\phi$ is i-stratified, and let s be any assignment. Since $U_i \supseteq T_i$, $\phi \in U_i$, in fact since $\mathbf{T}_i^{\alpha}\vDash\phi$ is i-stratified, it follows that $\phi \in U_i \cap \alpha$. Since ϕ is a sentence, $\phi \equiv \phi^s$, and so $U_i \cap \alpha \models \phi^s$, and so $\mathscr{M} \models \mathbf{T}_i^{\alpha}\vDash\phi[s]$. By Definition 54, $\mathscr{M}^{I(i)}$ agrees with $\mathscr{M}$ on $\mathbf{T}_i^{\alpha}\vDash$, so $\mathscr{M}^{I(i)} \models \mathbf{T}_i^{\alpha}\vDash\phi[s]$, as desired.

(2) Since $\mathbf{T}$ is straticlosed-r.e.-generic and $\mathbf{U} \supseteq \mathbf{T}$ is straticlosed and r.e., immediately $\mathscr{M}^{K(k)} \models \mathbf{T}$ (by Definition 61) for all $k \in \omega$ and any computable stratifier-set $K(k)$ above k. Let $I(i)$ be any computable stratifier-set above i. Suppose $\phi \in T_j$ where $j \not\geq i$. Let s be any assignment. Since $U_j \supseteq T_j$, $\phi \in U_j$. By Lemma 41, there is some very j-stratified $\psi \in U_j$ such that $\psi^- \equiv \phi^-$. Clearly since ϕ is a sentence, so is ψ. By compactness, there is some positive integer multiple α of ϵ_0 such that $U_j \cap \alpha \models \psi$. Since ψ is a sentence, $\psi \equiv \psi^s$ and thus $U_j \cap \alpha \models \psi^s$. Thus, $\mathscr{M} \models \mathbf{T}_j^{\alpha}\vDash\psi[s]$. By Theorem 50, $\mathscr{M} \models \mathbf{T}_j\vDash\psi^-[s]$, so by choice of ψ, $\mathscr{M} \models \mathbf{T}_j\vDash\phi^-[s]$. Since $I(i)$ is above i and $j \not\geq i$, $\mathscr{M}$ and $\mathscr{M}^{I(i)}$ agree on $\mathbf{T}_j\vDash$, so $\mathscr{M}^{I(i)} \models \mathbf{T}_j\vDash\phi^-[s]$, as desired. □

7.2 Stratifiable-r.e.-generic Building Blocks

We have established some straticlosed-r.e.-generic building blocks, but the goal of this paper is to better understand the structure of non-stratified theories—stratification is only a means to an end. Therefore, we introduce a corresponding non-stratified building-block notion.

Definition 69. If $\mathbf{T}^0 = (T_i^0)_{i\in\omega}$ where each T_i^0 is an $\mathscr{L}_{\mathrm{PA}}(\omega)$-theory, we say $\mathbf{T}^0$ is *$\prec$-stratifiable-r.e.-generic* (or *stratifiable-r.e.-generic* if $\prec$ is clear from context) if there is some $\prec$-straticlosed-r.e.-generic family $\mathbf{T} = (T_i)_{i\in\omega}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories such that each $T_i^- = T_i^0$.

Lemma 70. If $\mathbf{T} = (T_i)_{i\in\omega}$ is any straticlosed-r.e.-generic family of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories, then $\mathbf{T}^- = (T_i^-)_{i\in\omega}$ is a stratifiable-r.e.-generic family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories.

Proof. Straightforward. □

Corollary 71. (Compare Corollary 23) For all $i, j \in \omega$, each of the following families of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories is stratifiable-r.e.-generic.

1. $[j\text{-Deduction}]_i$ (if $i \not\prec j$).
2. [Weak j-Deduction]$_i$ (if $i \prec j$).
3. [Assigned Validity]$_i$.
4. [Assigned Validity]$_i \cup [i\text{-Validity}]_j$.
5. [Assigned Validity]$_i \cup [i\text{-Validity}]_i \cup [i\text{-Deduction}]_i \cup [i\text{-Introspection}]_j$.
6. $[S]_i$ where S is the axioms of Peano Arithmetic for $\mathscr{L}_{\mathrm{PA}}(\omega)$.
7. $[j\text{-SMT}]_i$.
8. (If $i \not\prec j$) $\mathbf{T} \cup [S]_i$, for any stratifiable-r.e.-generic $\mathbf{T} = (T_k)_{k\in\omega}$, where S is the schema: $\mathbf{T}_j\vDash\phi$ ($\phi \in T_j$).

Proof. By combining Lemma 70 with Lemmas 63–68. For parts involving validity, Lemma 56 can be used to provide valid stratified counterparts of valid non-stratified formulas. □

Comparing the stratifiable-r.e.-generic families we exhibited (Corollary 71) with the closed-r.e.-generic families we exhibited (Corollary 23), we see that the stratifiable-r.e.-generic families are weaker in exactly two ways:

1. They do not allow T_i to state full j-Deduction for T_j when $i \prec j$, instead allowing what we called Weak j-Deduction.

2. Their closure property is more restricted: if $\mathbf{T}^1 = (T^1_k)_{k\in\omega}$ is closed-r.e.-generic and $\mathbf{T}^2 = (T^2_k)_{k\in\omega}$ is stratifible-r.e.-generic, and if S_1 is the schema $\mathbf{T}_j\vDash\phi$ ($\phi \in T^1_j$), and if S_2 is the schema $\mathbf{T}_j\vDash\phi$ ($\phi \in T^2_j$), then Corollary 23 says $\mathbf{T}^1 \cup [S_1]_i$ is closed-r.e.-generic with no restrictions on j, whereas Corollary 71 only says that $\mathbf{T}^2 \cup [S_2]_i$ is stratifiable-r.e.-generic if $i \not\prec j$.

We leave it an open question to what extent Corollary 71 could be further strengthened. Our primary motivation in choosing building blocks was to facilitate creation of background provability theories at least strong enough to make our own consistency result (Theorem 72 below) generalize Carlson's consistency result [5]. If that were our lone motivation, we could restrict Corollary 71 to only those families where $i = j$, but a secondary motivation was to provide inter-theory versions of those restricted building blocks.

8 Second Consistency Result: Prioritizing Self-Truth

In this section, we continue to fix an r.e. well-founded partial-order $\prec$ of ω. The following theorem will satisfy the second promise from the introduction: it will exhibit true theories $(T_i)_{i\in\omega}$ such that T_i expresses a Gödel number of T_j ($j \prec i$) and the truth of T_j ($j \preceq i$). These theories can further be taken so that T_i expresses the fact that T_j has some Gödel number (all i, j), by Lemma 67.

Theorem 72. Let $\mathbf{T}^0 = (T^0_i)_{i\in\omega}$ be any stratifiable-r.e.-generic family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories. For every $i \in \omega$ and $n \in \mathbb{N}$, let $T_i(n)$ be the smallest $\mathbf{T}_i\vDash$-closed $\mathscr{L}_{\mathrm{PA}}(\omega)$-theory containing the following axioms.

1. The axioms contained in T^0_i.
2. Assigned Validity, i-Validity and i-Deduction.
3. $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \to \phi)$ whenever $j \preceq i$.
4. $\forall x(\mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ whenever $j \prec i$, $\mathrm{FV}(\phi) \subseteq \{x\}$.

Let each $\mathbf{T}(n) = (T_i(n))_{i\in\omega}$. There is some $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true.

Proof. By the S-m-n Theorem, there is a total computable $f : \mathbb{N} \to \mathbb{N}$ such that $\forall n \in \mathbb{N}$,

$$W_{f(n)} = \{\langle\ulcorner\phi\urcorner, j, m\rangle \in \mathbb{N} : \phi \text{ is an } \mathscr{L}_{\mathrm{PA}}(\omega)\text{-formula, } \mathrm{FV}(\phi) \subseteq \{x\}, \text{ and } T_j(n) \models \phi(x|\overline{m})\}.$$

By the Recursion Theorem, there is an $n \in \mathbb{N}$ such that $W_n = W_{f(n)}$. We will show $\mathbf{T}(n)$ is true. For the rest of the proof, we write $\mathbf{T}$ for $\mathbf{T}(n)$, T_i for $T_i(n)$.

The structure of the proof is as follows.

- ("Definition of $\mathbf{U}$" below) First, we will define a certain carefully-chosen family $\mathbf{U} = (U_i)_{i\in\omega}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories (with each $U_i^- = T_i$) and the $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$.
- ("Preliminary Result" below) Next, we will show that $\forall i \in \omega$, $\mathscr{M} \models U_i \cup T_i$. In order to deal with the difficulty mentioned at the beginning of Section 6, we will prove more than necessary, to obtain a strong $\prec$-induction hypothesis. Namely, we will prove, by $\prec$-induction, that $\forall i \in \omega$, for every stratifier-set $I(i)$ above i, $\mathscr{M}^{I(i)} \models U_i \cup T_i$.
  - (Claim 1 below) In order to prove $\mathscr{M}^{I(i)} \models U_i$, we will use induction on α to show that $\mathscr{M}^{I(i)} \models U_i \cap \alpha$ for all $\alpha \in \epsilon_0 \cdot \omega$.
  - (Case 3 below) Part of proving $\mathscr{M}^{I(i)} \models U_i \cap \alpha$ will be proving $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}^{\alpha_0}_i\vDash\phi \to \phi)$ whenever this is i-stratified, $\alpha_0 < \alpha$. This is where we will use the α-induction hypothesis.
  - (Case 4 below) Part of proving $\mathscr{M}^{I(i)} \models U_i \cap \alpha$ will be proving $\mathscr{M}^{I(i)} \models \mathrm{ucl}(\mathbf{T}_j\vDash\phi \to \phi^+)$ whenever $j \prec i$, ϕ is an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula, and $\bullet^+$ is an i-stratifier. This is where we will take advantage of our strong $\prec$-induction hypothesis.
- (Claims 2–3 below) Once we've established $\mathscr{M}^{I(i)} \models U_i$, we will essentially be able to conclude $\mathscr{M}^{I(i)} \models T_i$ using the Stratification Theorem (Theorem 50).

- At the very end of the proof, having established that $\forall i \in \omega$, $\mathscr{M} \models U_i \cup T_i$, we will use that to prove that $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$, i.e., that $\mathbf{T}$ is true.

Definition of U. Since $\mathbf{T}^0$ is stratifiable-r.e.-generic, there is a straticlosed-r.e-generic family $\mathbf{V} = (V_i)_{i\in\omega}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories such that each $V_i^- = T_i^0$. For every $i \in \mathbb{N}$, let U_i be the smallest i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory such that the following hold.

1. U_i contains V_i.
2. U_i contains i-Assigned Strativalidity, i-Strativalidity, i-Stratideduction and i-Collapse.
3. U_i contains $\mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi \to \phi)$ whenever $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified.
4. U_i contains $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \to \phi^+)$ for every $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, $j \prec i$, and i-stratifier $\bullet^+$.
5. U_i contains $\forall x(\mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ whenever $j \prec i$, $\mathrm{FV}(\phi) \subseteq \{x\}$ and ϕ is an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula.
6. Whenever $\phi \in U_i$ and $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified, $\mathbf{T}_i^\alpha\vDash\phi \in U_i$.

Let $\mathbf{U} = (U_i)_{i\in\omega}$. Observe that $\mathbf{U}$ is straticlosed and r.e. (to see U_i is i-unistratified, use Lemma 52; to see $\mathbf{U}$ is r.e., use Theorem 26 part 1); $\mathbf{U} \supseteq \mathbf{V}$; and for each $i \in \omega$, $U_i^- = T_i$.

Let $\mathscr{M} = \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$. Recall that $\mathrm{Str}(\mathbf{U})$ is the $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-family $(S_i)_{i\in\mathcal{I}}$ where $\forall i \in \omega$ and $\alpha \in \epsilon_0 \cdot \omega$, $S_i = U_i^- = T_i$ and $S_{(\alpha,i)} = U_i \cap \alpha$. For the reader's convenience, here is how (by definition) $\mathscr{M}$ interprets $\mathbf{T}_i\vDash$ and $\mathbf{T}_i^\alpha\vDash$ for all $i \in \omega$, $\alpha \in \epsilon_0 \cdot \omega$:

$$\mathscr{M} \models \mathbf{T}_i\vDash\phi[s] \text{ iff } T_i \models \phi^s,$$
$$\mathscr{M} \models \mathbf{T}_i^\alpha\vDash\phi[s] \text{ iff } U_i \cap \alpha \models \phi^s.$$

Preliminary Result. We would like to prove the following preliminary result: $\forall i \in \omega$, $\mathscr{M} \models U_i \cup T_i$. For the sake of a stronger induction hypothesis, we will prove that $\forall i \in \omega$, for every computable stratifier-set $I(i)$ above i, $\mathscr{M}^{I(i)} \models U_i \cup T_i$. This is more than enough because $\mathscr{M}^{I(i)} = \mathscr{M}$ when $I(i) = \emptyset$.

Fix $i \in \omega$. By $\prec$-induction, we have the following:

($*$) For every $j \prec i$, for every computable stratifier-set $J(j)$ above j, $\mathscr{M}^{J(j)} \models U_j \cup T_j$.

Let $I(i)$ be any computable stratifier-set above i. We must show $\mathscr{M}^{I(i)} \models U_i \cup T_i$.

Claim 1: $\forall \alpha \in \epsilon_0 \cdot \omega$, $\mathscr{M}^{I(i)} \models U_i \cap \alpha$.

By induction on α. Let $\sigma \in U_i \cap \alpha$.

Case 1: $\sigma \in V_i$. Then $\mathscr{M}^{I(i)} \models \sigma$ because $\mathbf{V}$ is straticlosed-r.e.-generic and $\mathbf{U} \supseteq \mathbf{V}$ is straticlosed and r.e.

Case 2: σ is an instance of i-Assigned Strativalidity, i-Strativalidity, or i-Stratideduction. Then $\mathscr{M}^{I(i)} \models \sigma$ by Lemma 63 or Lemma 64.

Case 3: σ is $\mathrm{ucl}(\mathbf{T}_i^{\alpha_0}\vDash\phi \to \phi)$ for some i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ such that $\mathbf{T}_i^{\alpha_0}\vDash\phi$ is i-stratified. Since $\sigma \in U_i \cap \alpha$, this forces $\alpha_0 < \alpha$. Let s be an assignment and assume $\mathscr{M}^{I(i)} \models \mathbf{T}_i^{\alpha_0}\vDash\phi[s]$, then:

$$\begin{aligned}
\mathscr{M}^{I(i)} &\models \mathbf{T}_i^{\alpha_0}\vDash\phi[s] && \text{(Assumption)}\\
\mathscr{M} &\models \mathbf{T}_i^{\alpha_0}\vDash\phi[s] && (\mathscr{M} \text{ and } \mathscr{M}^{I(i)} \text{ agree on } \mathbf{T}_i^{\alpha_0}\vDash \text{ by Def. 54})\\
U_i \cap \alpha_0 &\models \phi^s && \text{(Definition of } \mathscr{M})\\
\mathscr{M}^{I(i)} &\models \phi^s && (\text{By } \alpha\text{-induction, } \mathscr{M}^{I(i)} \models U_i \cap \alpha_0)\\
\mathscr{M}^{I(i)} &\models \phi[s]. && \text{(Lemma 59)}
\end{aligned}$$

Case 4: σ is $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \to \phi^+)$ for some $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, $j \prec i$, and i-stratifier $\bullet^+$. By Lemma 52 we may assume $\bullet^+$ is computable. Let $J(j)$ be the computable stratifier-set $J(j) = I(i) \cup \{\bullet^+\}$, which is above j

since $I(i)$ is above i and $j \prec i$. Let s be an assignment and assume $\mathscr{M}^{I(i)} \models \mathbf{T}_j\vDash\phi[s]$, then:

$$\begin{aligned}
\mathscr{M}^{I(i)} &\models \mathbf{T}_j\vDash\phi[s] && \text{(Assumption)}\\
\mathscr{M} &\models \mathbf{T}_j\vDash\phi[s] && \text{(Since } j \prec i \text{ and } I(i) \text{ is above } i, \mathscr{M}^{I(i)} \text{ and } \mathscr{M} \text{ agree on } \mathbf{T}_j\vDash)\\
T_j &\models \phi^s && \text{(Definition of } \mathscr{M})\\
\mathscr{M}^{J(j)} &\models \phi^s && \text{(Since } \mathscr{M}^{J(j)} \models T_j \text{ by } (*))\\
(\mathscr{M}^{I(i)})^+ &\models \phi^s && \text{(Lemma 57)}\\
\mathscr{M}^{I(i)} &\models (\phi^s)^+ && \text{(Lemma 55)}\\
\mathscr{M}^{I(i)} &\models (\phi^+)^s && \text{(Clearly } (\phi^s)^+ \equiv (\phi^+)^s)\\
\mathscr{M}^{I(i)} &\models \phi^+[s]. && \text{(Lemma 59)}
\end{aligned}$$

Case 5: σ is $\forall x(\mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ for some $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ with $\mathrm{FV}(\phi) \subseteq \{x\}$ and $j \prec i$. Let s be any assignment, say $s(x) = m$. The following biconditionals are equivalent:

$$\begin{aligned}
\mathscr{M}^{I(i)} &\models \mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}}[s]\\
\mathscr{M} &\models \mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}}[s] && (\mathscr{M}^{I(i)} \text{ and } \mathscr{M} \text{ agree on the symbols in question})\\
\mathscr{M} &\models \mathbf{T}_j\vDash\phi[s] \text{ iff } \mathscr{M} \models \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, \overline{m}\rangle \in W_{\overline{n}} && \text{(Lemma 59)}\\
\mathscr{M} &\models \mathbf{T}_j\vDash\phi[s] \text{ iff } \langle\ulcorner\phi\urcorner, j, m\rangle \in W_n && (\mathscr{M} \text{ has standard first-order part})\\
T_j &\models \phi^s \text{ iff } \langle\ulcorner\phi\urcorner, j, m\rangle \in W_n && \text{(Definition of } \mathscr{M})\\
T_j &\models \phi(x|\overline{m}) \text{ iff } \langle\ulcorner\phi\urcorner, j, m\rangle \in W_n. && \text{(Since } \mathrm{FV}(\phi) \subseteq \{x\})
\end{aligned}$$

The latter is true by definition of n.

Case 6: σ is an instance $\mathbf{T}_i^\beta\vDash\phi \leftrightarrow \mathbf{T}_i^\gamma\vDash\phi$ of i-Collapse (so $\beta \leq_1 \gamma$ and $\mathbf{T}_i^\beta\vDash\phi \leftrightarrow \mathbf{T}_i^\gamma\vDash\phi$ is i-stratified). Let s be an assignment, since $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_i^\beta\vDash$ and $\mathbf{T}_i^\gamma\vDash$, we need only show $\mathscr{M} \models \mathbf{T}_i^\beta\vDash\phi \leftrightarrow \mathbf{T}_i^\gamma\vDash\phi[s]$. In other words we must show $U_i \cap \beta \models \phi^s$ if and only if $U_i \cap \gamma \models \phi^s$. This is by Theorem 38.

Case 7: σ is $\mathbf{T}_i^{\alpha_0}\vDash\phi$ for some $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ such that $\mathbf{T}_i^{\alpha_0}\vDash\phi$ is i-stratified and $\phi \in U_i$. Since $\mathbf{T}_i^{\alpha_0}\vDash\phi$ is i-stratified, $\mathrm{On}(\phi) \subseteq \alpha_0$, so $\phi \in U_i \cap \alpha_0$. Thus $\mathscr{M} \models \mathbf{T}_i^{\alpha_0}\vDash\phi$, so $\mathscr{M}^{I(i)} \models \mathbf{T}_i^{\alpha_0}\vDash\phi$ since $\mathscr{M}^{I(i)}$ and $\mathscr{M}$ agree on $\mathbf{T}_i^{\alpha_0}\vDash$.

Cases 1–7 establish $\mathscr{M}^{I(i)} \models U_i \cap \alpha$. By arbitrariness of α, Claim 1 is proved.

Claim 2: For any assignment s and any very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, $\mathscr{M}^{I(i)} \models \phi[s]$ if and only if $\mathscr{M}^{I(i)} \models \phi^-[s]$.

By Theorem 50, for all such s and ϕ, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^-[s]$. The claim now follows from Lemma 58 ($i \notin \mathrm{Indices}(I(i))$ because $I(i)$ is above i).

Claim 3: $\mathscr{M}^{I(i)} \models T_i$.

For any $\sigma \in T_i$, there is some $\tau \in U_i$ such that $\tau^- \equiv \sigma$; since U_i is i-unistratified, we may take τ to be very i-stratified (Lemma 41). By Claim 1, $\mathscr{M}^{I(i)} \models U_i$, so $\mathscr{M}^{I(i)} \models \tau$. By Claim 2, $\mathscr{M}^{I(i)} \models \sigma$.

For each $i \in \omega$, letting $I(i) = \emptyset$, Claims 1–3 show that $\mathscr{M} \models U_i \cup T_i$. It follows that $\mathscr{M} \models \mathbf{T}$. Now, for every $i \in \omega$, $\mathscr{M}_{\mathbf{T}}$ interprets $\mathbf{T}_i\vDash$ as follows:

$$\mathscr{M}_{\mathbf{T}} \models \mathbf{T}_i\vDash\phi[s] \text{ iff } T_i \models \phi^s.$$

This is exactly the same way that $\mathscr{M}$ interprets $\mathbf{T}_i\vDash$. It follows that $\mathscr{M}$ and $\mathscr{M}_{\mathbf{T}}$ agree on $\mathscr{L}_{\mathrm{PA}}(\omega)$-formulas. Thus, since $\mathscr{M} \models \mathbf{T}$, $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$, i.e., $\mathbf{T}$ is true. □

9 Well-Foundation and Ill-Foundation

The following is a variation on Kleene's $\mathcal{O}$.

Definition 73. Simultaneously define $\mathcal{O} \subseteq \mathbb{N}$ and $|\bullet| : \mathcal{O} \to \mathrm{Ord}$ so that $\mathcal{O} \subseteq \mathbb{N}$ is the smallest set such that:

1. $0 \in \mathcal{O}$ (it represents the ordinal $|0| = 0$).
2. $\forall n \in \mathcal{O}$, $2^n \in \mathcal{O}$ (it represents the ordinal $|2^n| = |n| + 1$).
3. If φ_e (the eth partial recursive function) is total and $\mathrm{range}(\varphi_e) \subseteq \mathcal{O}$, then $3 \cdot 5^e \in \mathcal{O}$ (it represents the ordinal $|3 \cdot 5^e| = \sup\{|\varphi_e(0)|, |\varphi_e(1)|, \ldots\}$).

To avoid technical complications, we have differed from the usual Kleene's $\mathcal{O}$ in the following way: in the usual definition, in order for $3 \cdot 5^e$ to lie in $\mathcal{O}$, it is also required that $|\varphi_e(0)| < |\varphi_e(1)| < \cdots$.

Definition 74. $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}$ is the language of Peano arithmetic extended by a unary predicate $\mathcal{O}$. The following notions are defined by analogy with Section 2:

1. For any assignment s and $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-formula ϕ with $\mathrm{FV}(\phi)=\{x_1, \ldots, x_n\}$, $\phi^s \equiv \phi(x_1|\overline{s(x_1)}) \cdots (x_n|\overline{s(x_n)})$.
2. If $\mathbf{T} = (T_i)_{i \in I}$ is an I-indexed family of $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-theories, the *intended structure* for $\mathbf{T}$ is the $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-structure $\mathscr{M}_{\mathbf{T}}$ with universe $\mathbb{N}$, interpreting symbols of PA as usual and interpreting $\mathcal{O}$ as $\mathcal{O}$, and interpreting $\mathbf{T}_i\vDash$ $(i \in I)$ as in Definition 7. For any $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-structure $\mathscr{N}$, we write $\mathscr{N} \models \mathbf{T}$ if $\forall i \in I$, $\mathscr{N} \models T_i$. We say $\mathbf{T}$ is *true* if $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$.

Definition 75. If I is an index set and $\mathbf{T} = (T_i)_{i \in I}$ is a family of $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-theories, then for any $i \in I$ such that $\mathscr{M}_{\mathbf{T}} \models T_i$, we define the ordinal $\|T_i\| = \sup\{|m| + 1 : T_i \models \mathcal{O}(\overline{m})\}$.

The above definition makes sense: since $\mathscr{M}_{\mathbf{T}} \models T_i$ and $\mathcal{O}^{\mathscr{M}_{\mathbf{T}}} = \mathcal{O}$, the supremands are defined.

Definition 76. The *basic axioms of* $\mathcal{O}$ are the following $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}$-axioms.

1. $\mathcal{O}(0)$.
2. $\mathcal{O}(\overline{n}) \to \mathcal{O}(\overline{2^n})$, for every $n \in \mathbb{N}$.
3. $\forall x(\varphi_{\overline{n}}(x)\!\downarrow \,\&\, \mathcal{O}(\varphi_{\overline{n}}(x))) \to \mathcal{O}(\overline{3 \cdot 5^n})$, for every $n \in \mathbb{N}$.

We have written the last two lines using infinite schemata to strengthen the following result.

Theorem 77. Let I be an index set, $\prec$ a binary relation on I. Suppose $\mathbf{T} = (T_i)_{i \in I}$ is a family of $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-theories with the following properties:

1. $\forall i \in I$, T_i contains the axioms of Peano arithmetic.
2. $\forall i \in I$, T_i contains the basic axioms of $\mathcal{O}$.
3. $\forall i \in I$, $\forall j \prec i$, $\exists n \in \mathbb{N}$ such that $T_i \models \forall x(\mathbf{T}_j\vDash\mathcal{O}(x) \leftrightarrow x \in W_{\overline{n}})$.
4. $\forall i \in I$, $\forall j \prec i$, $T_i \models \forall x(\mathbf{T}_j\vDash\mathcal{O}(x) \to \mathcal{O}(x))$.

If $\mathscr{M}_{\mathbf{T}} \models T_i \cup T_j$ (in particular if $\mathbf{T}$ is true) and $j \prec i$, then $\|T_j\| < \|T_i\|$.

Proof. Assume $\mathscr{M}_{\mathbf{T}} \models T_i \cup T_j$ and $j \prec i$. By hypothesis there is some $n \in \mathbb{N}$ such that $T_i \models \forall x(\mathbf{T}_j\vDash\mathcal{O}(x) \leftrightarrow x \in W_{\overline{n}})$ and $T_i \models \forall x(\mathbf{T}_j\vDash\mathcal{O}(x) \to \mathcal{O}(x))$. From these, $T_i \models \forall x(x \in W_{\overline{n}} \to \mathcal{O}(x))$.

Since $\mathscr{M}_{\mathbf{T}} \models T_i$, in particular $\mathscr{M}_{\mathbf{T}} \models \forall x(\mathbf{T}_j\vDash\mathcal{O}(x) \leftrightarrow x \in W_{\overline{n}})$. This means $W_n = \{m \in \mathbb{N} : T_j \models \mathcal{O}(\overline{m})\}$. Since T_j includes the axiom $\mathcal{O}(0)$, $W_n \neq \emptyset$.

Since $W_n \neq \emptyset$, by computability theory there is some $k \in \mathbb{N}$ such that

$$\mathrm{PA} \models (\mathrm{domain}(\varphi_{\overline{k}}) = \mathbb{N}) \wedge (\mathrm{range}(\varphi_{\overline{k}}) = W_{\overline{n}}).$$

Since T_i includes PA, T_i also implies as much. Combined with $T_i \models \forall x(x \in W_{\overline{n}} \to \mathcal{O}(x))$, it follows that $T_i \models \forall x(\varphi_{\overline{k}}(x)\!\downarrow \,\&\, \mathcal{O}(\varphi_{\overline{k}}(x)))$. Since T_i contains the basic axiom $\forall x(\varphi_{\overline{k}}(x)\!\downarrow \,\&\, \mathcal{O}(\varphi_{\overline{k}}(x))) \to \mathcal{O}(\overline{3 \cdot 5^k})$, $T_i \models \mathcal{O}(\overline{3 \cdot 5^k})$.

To finish the proof, calculate

$$\begin{aligned}
\|T_j\| &= \sup\{|m|+1 : T_j \models \mathcal{O}(\overline{m})\} \\
&= \sup\{|m| : T_j \models \mathcal{O}(\overline{m})\} && \text{(Since } T_j \text{ contains } \mathcal{O}(\overline{n}) \to \mathcal{O}(\overline{2^n}) \text{ for all } n \in \mathbb{N}) \\
&= \sup\{|m| : m \in W_n\} && \text{(Since } W_n = \{m \in \mathbb{N} : T_j \models \mathcal{O}(\overline{m})\}) \\
&= \sup\{|\varphi_k(0)|, |\varphi_k(1)|, \ldots\} && \text{(By choice of } k) \\
&= |3 \cdot 5^k| && \text{(Definition 73)} \\
&< \sup\{|m|+1 : T_i \models \mathcal{O}(\overline{m})\} && \text{(Since } T_i \models \mathcal{O}(\overline{3 \cdot 5^k})) \\
&= \|T_i\|.
\end{aligned}$$

□

Corollary 78. (Well-Foundedness of True Self-Referential Theories) Let I, $\mathbf{T}$, $\prec$ be as in Theorem 77. If $\mathbf{T}$ is true then $\prec$ is well founded, by which we mean there is no infinite descending sequence $i_0 \succ i_1 \succ \cdots$.

In particular Corollary 78 says that if I, $\mathbf{T}$, $\prec$ are as in Theorem 77 and $\mathbf{T}$ is true then $\prec$ is strict: there is no i with $i \prec i$. This gives a new form (under the additional new assumption of containing/knowing basic rudiments of computable ordinals) of the Lucas–Penrose–Reinhardt argument that a truthful theory (or machine) cannot state (or know) its own truth and its own Gödel number.

We could remove Peano arithmetic from Theorem 77 if we further departed from Kleene and changed line 3 of Definition 73 to read:

3. If $W_e \subseteq \mathcal{O}$, then $3 \cdot 5^e \in \mathcal{O}$ (and $|3 \cdot 5^e| = \sup\{|n| : n \in W_e\}$, or $|3 \cdot 5^e| = 0$ if $W_e = \emptyset$)

(and altered Definition 76 accordingly). The previous paragraph would still stand, in fact giving a version of the Lucas–Penrose–Reinhardt argument in which the theory (machine) is not required to contain (know) arithmetic.

We close the paper by showing that Corollary 78 fails without $\mathcal{O}$. Let WF be the set of all r.e. well-founded partial orders on ω and let Tr be the set of all true $\mathscr{L}_{\mathrm{PA}}$-sentences. It is well-known that WF is computability theoretically Π^1_1-complete and Tr is Δ^1_1, so WF cannot be defined in $\mathscr{L}_{\mathrm{PA}} \cup \{\mathrm{Tr}\}$.

Theorem 79. (Ill-Foundedness of True Self-Referential Theories)

1. There exists an r.e., ill-founded partial order $\prec$ on ω such that for every closed-r.e.-generic $\mathbf{T}^0 = (T^0_i)_{i \in \omega}$ there is an $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true, where $\mathbf{T}(n)$ is as in Theorem 24.

2. There exists an r.e., ill-founded partial order $\prec$ on ω such that for every $\prec$-stratifiable-r.e.-generic $\mathbf{T}^0 = (T^0_i)_{i \in \omega}$ there is an $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true, where $\mathbf{T}(n)$ is as in Theorem 72.

Proof. We prove (1), (2) is similar. Assume ¬(1). For each r.e. partial order $\prec$ on ω, let $S(\prec)$ be the statement of Theorem 24 for $\prec$, minus the requirement that $\prec$ be well founded. Combining ¬(1) with Theorem 24, $\prec$ is well founded if and only if $S(\prec)$ is true. We will argue that $S(\prec)$ is expressible in $\mathscr{L}_{\mathrm{PA}} \cup \{\mathrm{Tr}\}$, which is absurd because that would mean it is possible to define WF in $\mathscr{L}_{\mathrm{PA}} \cup \{\mathrm{Tr}\}$.

$S(\prec)$ is equivalent to the following:

- For any (Gödel number of an) r.e. family $\mathbf{T}^0 = (T^0_i)_{i \in \omega}$ of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories, if $\mathbf{T}^0$ is closed-r.e.-generic (i.e., if $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}^0$ for every closed r.e. family $\mathbf{U} \supseteq \mathbf{T}^0$ of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories), then there is some $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true (i.e., such that $\mathscr{M}_{\mathbf{T}(n)} \models \mathbf{T}(n)$), where $\mathbf{T}(n) = (T_i(n))_{i \in \omega}$, where each $T_i(n)$ is the smallest $\mathbf{T}_i\vDash$-closed theory containing the following:

  1. The axioms in T^0_i.
  2. $\forall x(\mathbf{T}_j \vDash \phi \leftrightarrow \langle \overline{\ulcorner \phi \urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ whenever $j \in \omega$, $\mathrm{FV}(\phi) \subseteq \{x\}$.
  3. $\mathrm{ucl}(\mathbf{T}_j \vDash \phi \to \phi)$ whenever $j \prec i$.

This is manifestly expressible in $\mathscr{L}_{\mathrm{PA}}$ except for the clauses $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}^0$ and $\mathscr{M}_{\mathbf{T}(n)} \models \mathbf{T}(n)$. We will show that $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}^0$ is expressible in $\mathscr{L}_{\mathrm{PA}} \cup \{\mathrm{Tr}\}$; the expressibility of $\mathscr{M}_{\mathbf{T}(n)} \models \mathbf{T}(n)$ is similar.

Define an operator $F_{\mathbf{U}}$ which takes an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ and outputs an $\mathscr{L}_{\mathrm{PA}}$-formula $F_{\mathbf{U}}(\phi)$ as follows:

- If ϕ is atomic, let $F_{\mathbf{U}}(\phi) \equiv \phi$.
- If ϕ is $\neg\phi_0$, $\phi_1 \rightarrow \phi_2$, or $\forall x\phi_0$, let $F_{\mathbf{U}}(\phi)$ be $\neg F_{\mathbf{U}}(\phi_0)$, $F_{\mathbf{U}}(\phi_1) \rightarrow F_{\mathbf{U}}(\phi_2)$, or $\forall x F_{\mathbf{U}}(\phi_0)$, respectively.
- Suppose ϕ is $\mathbf{T}_i{\vDash}\psi$ and $\mathrm{FV}(\psi) = \{x_1, \ldots, x_k\}$. Let $f : \mathbb{N}^k \rightarrow \mathbb{N}$ be the computable function such that for all $m_1, \ldots, m_k \in \mathbb{N}$, $f(m_1, \ldots, m_k) = \ulcorner\psi(x_1|\overline{m_1}) \cdots (x_k|\overline{m_k})\urcorner$. Let $F_{\mathbf{U}}(\phi)$ be: "U_i proves the sentence with Gödel number $f(x_1, \ldots, x_k)$" (so $\mathrm{FV}(F_{\mathbf{U}}(\phi)) = \{x_1, \ldots, x_k\}$).

It is easy to check that for every $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ and assignment s, $\mathscr{M}_{\mathbf{U}} \models \phi[s]$ if and only if $\mathbb{N} \models F_{\mathbf{U}}(\phi)[s]$. In particular, for every $\mathscr{L}_{\mathrm{PA}}(\omega)$-sentence ϕ, $\mathscr{M}_{\mathbf{U}} \models \phi$ if and only if $\mathbb{N} \models F_{\mathbf{U}}(\phi)$. Thus, the clause $\mathscr{M}_{\mathbf{U}} \models \mathbf{T}^0$ can be expressed in $\mathscr{L}_{\mathrm{PA}} \cup \{\mathrm{Tr}\}$ as follows: $\forall i \forall x(x \in T_i^0 \rightarrow \mathrm{Tr}(\ulcorner F_{\mathbf{U}}(x)\urcorner))$. □

The way we prove Theorem 79 by referring to the computability theoretical complexity of WF is similar to a recent argument by Kripke [7].

References

[1] Alexander, S. (2013). The Theory of Several Knowing Machines. Doctoral dissertation, the Ohio State University.

[2] Alexander, S. (2014). A machine that knows its own code. *Studia Logica*, **102**, 567–576.

[3] Benacerraf, P. (1967). God, the Devil, and Gödel. *The Monist*, **51**, 9–32.

[4] Carlson, T.J. (1999). Ordinal arithmetic and Σ_1-elementarity. *Archive for Mathematical Logic*, **38**, 449–460.

[5] Carlson, T.J. (2000). Knowledge, machines, and the consistency of Reinhardt's strong mechanistic thesis. *Annals of Pure and Applied Logic*, **105**, 51–82.

[6] Carlson, T.J. (2001). Elementary patterns of resemblance. *Annals of Pure and Applied Logic*, **108**, 19–77.

[7] Kripke, S.A. (2019). Ungroundedness in Tarskian Languages. *The Journal of Philosophical Logic*, **48**, 603–609.

[8] Lucas, J.R. (1961). Minds, machines, and Gödel. *Philosophy*, **36**, 112–127.

[9] Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics.* Oxford University Press.

[10] Putnam, H. (2006). After Gödel. *Logic Journal of the IGPL*, **14**, 745–754.

[11] Reinhardt, W. (1985). Absolute versions of incompleteness theorems. *Nous*, **19**, 317–346.

[12] Shapiro, S. (1985). Epistemic and Intuitionistic Arithmetic. In: S. Shapiro (ed.), *Intensional Mathematics* (North-Holland, Amsterdam), pp. 11–46.