Self-referential theories

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

June 26, 2013

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 [7], Benacerraf [3], Reinhardt [10], Penrose [8], and Putnam [9]. 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 five 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. There are true theories $(T_i)_{i\in\omega}$ such that T_i expresses a Gödel number of T_j ($j \prec i$), truth of T_j ($j \preceq i$), a weakened form of the truth of T_j (all i, j), and that T_j has some Gödel number ($j \preceq i$).

4. 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 1–3 fail.

5. Finally, if we do not extend the base language as in 4, then there do exist ill-founded r.e. partial orders $\prec$ such that 1–3 hold.

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

Our proofs of 1–3 are constructive, but the proof of 5 is nonconstructive. In short, if 5 were false, any of 1–3 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.

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 $\mathscr{L}_{\mathrm{PA}}$ contains a function symbol for this bijection. Similarly we may act as if $\mathscr{L}_{\mathrm{PA}}$ contains 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.

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] $\sigma_1 \rightarrow \cdots \rightarrow \sigma_n \rightarrow \phi$ is valid.

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

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 $\mathbf{T}_i{\models}$ for all $i \in I$.

In case I is a singleton, $\mathscr{L}_{\mathrm{PA}}(I)$ is a form of Shapiro's [11] 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 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 $\mathbf{T}_i{\models}$ $(i \in I)$ as follows:

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

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

Lemma 8. Let $\mathbf{T} = (T_i)_{i \in I}$ be a family of $\mathscr{L}_{\mathrm{PA}}(I)$-theories. For any $\mathscr{L}_{\mathrm{PA}}(I)$-formula ϕ and assignment s, $\mathscr{M}_{\mathbf{T}} \models \phi[s]$ if and only if $\mathscr{M}_{\mathbf{T}} \models \phi^s$.

Proof. By induction. $\square$

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. Any $\mathscr{L}_{\mathrm{PA}}(I)$-structure with standard first-order part and satisfying the conclusion of Lemma 8 satisfies the axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(I)$.

Proof. Let $\mathscr{M}$ be any $\mathscr{L}_{\mathrm{PA}}(I)$-structure with standard first-order part and satisfying the conclusion of Lemma 8. 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 $\mathrm{ucl}(\phi(x|0) \rightarrow \forall x(\phi \rightarrow \phi(x|S(x))) \rightarrow \forall x\phi)$. To see $\mathscr{M} \models \sigma$, let s be an arbitrary assignment and assume $\mathscr{M} \models \phi(x|0)[s]$ and $\mathscr{M} \models \forall x(\phi \rightarrow \phi(x|S(x)))[s]$. By Lemma 8, $\mathscr{M} \models \phi^{s(x|0)}$ 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 Lemma 8, $\mathscr{M} \models \forall x\phi[s]$. $\square$

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 $\mathbf{T}_i{\models}$*-closed* if $\mathbf{T}_i{\models}\phi \in T_i$ whenever $\phi \in T_i$. We say $\mathbf{T}$ is *closed* if each T_i is $\mathbf{T}_i{\models}$-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.

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

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 $\mathrm{ucl}(\mathbf{T}_i{\vDash}(\phi \rightarrow \psi) \rightarrow \mathbf{T}_i{\vDash}\phi \rightarrow \mathbf{T}_i{\vDash}\psi)$. 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 $\mathscr{M}_{\mathbf{T}'} \models \mathbf{T}$ for every closed r.e. family $\mathbf{T}' \supseteq \mathbf{T}$ 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. $\square$

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

The following lemma provides building blocks that can be combined in diverse ways, via Lemma 15, to form background theories of provability.

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

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

   (j-Introspection) the schema $\mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \rightarrow \mathbf{T}_j{\vDash}\mathbf{T}_j{\vDash}\phi)$.

5. $[S]_i$ where S is the set of axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(\omega)$.
6. $[S]_i$ where S is any r.e. set of true arithmetic sentences.
7. $[S]_i$ where S is: (j-SMT) (See [5] and [10]) $\mathrm{ucl}(\exists e \forall x(\mathbf{T}_j{\vDash}\phi \leftrightarrow x \in W_e))$, $e \notin \mathrm{FV}(\phi)$.
8. $\mathbf{T} \cup [S]_i$ where $\mathbf{T} = (T_k)_{k\in\omega}$ is closed-r.e.-generic and S is the schema $\mathbf{T}_j{\vDash}\phi$ ($\phi \in T_j$).

Proof. Straightforward. We prove 3 and 4 to show why sometimes one schema comes packaged with others.
(3) By Theorem 3, $[\text{Assigned Validity}]_j \cup [j\text{-Validity}]_i$ is r.e. Let $\mathbf{T}' = (T'_k)_{k\in\omega}$ be any closed r.e. family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories such that T'_j contains Assigned Validity and T'_i contains j-Validity. We must show $\mathscr{M}_{\mathbf{T}'}$ satisfies Assigned Validity and j-Validity. For Assigned Validity, let ϕ be valid and s an assignment. Since ϕ is valid, $\mathscr{M}_{\mathbf{T}'} \models \phi[s]$, so by Lemma 8, $\mathscr{M}_{\mathbf{T}'} \models \phi^s$ as desired. For j-Validity, let ϕ be valid and s an assignment. Since T'_j contains Assigned Validity, $T'_j \models \phi^s$, so by definition of $\mathscr{M}_{\mathbf{T}'}$, $\mathscr{M}_{\mathbf{T}'} \models \mathbf{T}_j{\vDash}\phi[s]$.
(4) Recursive enumerability is by Theorem 3. Let $\mathbf{T}' = (T'_k)_{k\in\omega}$ be any closed r.e. family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories such that T'_j contains Assigned Validity, j-Validity and j-Deduction, and T'_i contains j-Introspection. That $\mathscr{M}_{\mathbf{T}'}$ satisfies Assigned Validity and j-Validity is as in (3). That $\mathscr{M}_{\mathbf{T}'}$ satisfies j-Deduction is straightforward. For j-Introspection, let s be an assignment and assume $\mathscr{M}_{\mathbf{T}'} \models \mathbf{T}_j{\vDash}\phi[s]$, we will show $\mathscr{M}_{\mathbf{T}'} \models \mathbf{T}_j{\vDash}\mathbf{T}_j{\vDash}\phi[s]$. Since $\mathscr{M}_{\mathbf{T}'} \models \mathbf{T}_j{\vDash}\phi[s]$, $T'_j \models \phi^s$. By Theorem 3, there are $\sigma_1, \ldots, \sigma_n \in T'_j$ such that $\sigma_1 \rightarrow \cdots \rightarrow \sigma_n \rightarrow \phi^s$ is valid. Since T'_j contains j-Validity, $T'_j \models \mathbf{T}_j{\vDash}(\sigma_1 \rightarrow \cdots \rightarrow \sigma_n \rightarrow \phi^s)$. By repeated applications of j-Deduction contained in T'_j, $T'_j \models \mathbf{T}_j{\vDash}\sigma_1 \rightarrow \cdots \rightarrow \mathbf{T}_j{\vDash}\sigma_n \rightarrow \mathbf{T}_j{\vDash}(\phi^s)$. Since $\mathbf{T}'$ is closed, T'_j is $\mathbf{T}_j{\vDash}$-closed and so contains $\mathbf{T}_j{\vDash}\sigma_1, \ldots, \mathbf{T}_j{\vDash}\sigma_n$. So $T'_j \models (\mathbf{T}_j{\vDash}\phi)^s$ and $\mathscr{M}_{\mathbf{T}'} \models \mathbf{T}_j{\vDash}\mathbf{T}_j{\vDash}\phi[s]$. $\square$

4 First Consistency Result: Prioritizing Exact Codes

The following theorem fulfils the first promise from the introduction.

Theorem 18. Suppose $\prec$ is an r.e. well-founded partial order on ω and $\mathbf{T}^0 = (T_i^0)_{i\in\omega}$ 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:

- The axioms in T_i^0.
- $\forall x(\mathbf{T}_i\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ whenever $j \in \omega$, $\mathrm{FV}(\phi) \subseteq \{x\}$.
- $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi)$ 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 $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$. Fix $i \in \omega$. 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 $\forall x(\mathbf{T}_i\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ for some $j \in \omega$, $\mathrm{FV}(\phi) \subseteq \{x\}$. Let s be an assignment, $m \in \mathbb{N}$. The following are equivalent.

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

Case 3: σ is $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi)$ 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 $\prec$-induction, $\mathscr{M}_{\mathbf{T}} \models T_j$, 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 to fulfil the second and third promises 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 18 was used to exhibit a machine that knows its own code.

5 Stratification

Definition 19. We define a binary relation $\leq_1$ on Ord by transfinite recursion so that for all $\alpha, \beta \in \mathrm{Ord}$, $\alpha \leq_1 \beta$ 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 20.

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. (See Figure 1) For any $\alpha \leq \beta \in \mathrm{Ord}$, $\alpha \leq_1 \beta$ 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 $X \cup \widetilde{Y} \cong_{(\leq, \leq_1)} X \cup Y$.

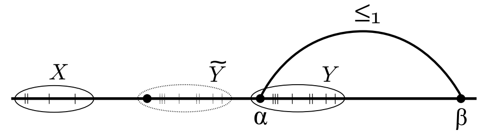

Figure 1:
Illustration of Theorem 20 part 3.

The usefulness of Theorem 20 will first appear in Theorem 31, but first we need some machinery.

Definition 21. Let $\mathcal{I} = ((\epsilon_0 \cdot \omega) \times \omega) \sqcup \omega$. Thus $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$ 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 *exponent.*

Definition 22. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, $\mathrm{On}(\phi) \subseteq \epsilon_0 \cdot \omega$ denotes the set of exponents appearing in ϕ.

Definition 23. 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 $\neg\phi_0$, $\phi_1 \rightarrow \phi_2$, 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:

- $K_7^\omega K_7^5(1=0) \rightarrow K_8(1=0)$ is 7-stratified but not 6- or 8-stratified.
- $K_7^5 K_7^\omega(1=0)$ is not 7-stratified, nor is $K_7^5 K_7(1=0)$.
- $K_7^5 K_8 K_7(1=0)$ is 7-stratified but $K_7^5 K_8 K_7^4(1=0)$ is not.

Definition 24. Suppose $X \subseteq \epsilon_0 \cdot \omega$ and $h : X \rightarrow \epsilon_0 \cdot \omega$ is order preserving. For each $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ, let $h(\phi)$ be the $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula obtained by 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

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

Definition 25. Suppose $X \subseteq \epsilon_0 \cdot \omega$ and $h : X \rightarrow \epsilon_0 \cdot \omega$ 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 26. 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 $\mathscr{N} \models h(\phi)[s]$.

$$\begin{tikzcd} \mathscr{L}_{\mathrm{PA}}(\mathcal{I}) \arrow[rr, "h"] \arrow[dr, "h(\mathscr{N})\models"'] & & \mathscr{L}_{\mathrm{PA}}(\mathcal{I}) \arrow[dl, "\mathscr{N}\models"] \\ & \{\mathrm{T}, \mathrm{F}\} & \end{tikzcd}$$

Proof. By induction. □

Corollary 27. 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 26. □

Definition 28. 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 29. Suppose $i \in \omega$. A set T of i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentences is *$i$-uniform* provided that 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 30. 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 31. (The Collapse Theorem) Suppose T is an i-uniform i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-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. We 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 20 part 2, $\epsilon_0 \cdot n \leq_1 \epsilon_0 \cdot n'$. By Theorem 20 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 27,

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

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

Definition 32. For every $i \in \omega$ we define the following $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-schema:

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

Definition 33. 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 $\mathbf{T}_5^{\omega}\models(1=0) \to \mathbf{T}_5^{\omega+1}\models\mathbf{T}_5^{\omega}\models(1=0)$, then ϕ^- is $\mathbf{T}_5\models(1=0) \to \mathbf{T}_5\models\mathbf{T}_5\models(1=0)$.

Lemma 34. If T is i-uniform 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 injective and order preserving; by Theorem 20 part 2, h is a covering. Since T is i-uniform, T contains $\psi \equiv h(\phi)$. Clearly ψ is very i-stratified and $\psi^- \equiv \phi^-$. □

Definition 35. 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 36. 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 $\mathscr{N} \models \phi^-[s]$.

$$\begin{tikzcd}
\mathscr{L}_{\mathrm{PA}}(\mathcal{I}) \arrow[rr, two heads, "-"] \arrow[dr, "\mathscr{N}^-\models"'] & & \mathscr{L}_{\mathrm{PA}}(\omega) \arrow[dl, "\mathscr{N}\models"] \\
& \{\mathrm{T}, \mathrm{F}\} &
\end{tikzcd}$$

Proof. By induction. $\square$

Corollary 37. 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 27. $\square$

A converse-like statement holds for Corollary 37 as well.

Lemma 38. 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 $\psi^- \equiv \phi$.

We postpone the proof of Lemma 38 until Section 6.

Definition 39. Let $i \in \omega$. We define the following $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-schemas.

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

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

Theorem 40. (Proof Stratification) Suppose T is an i-uniform i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory such that T includes i-Stratiavalidity, i-Stratideduction and i-Collapse. Assume that whenever $\phi \in T$ and $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified, $\mathbf{T}_i^\alpha\vDash\phi \in T$. 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.

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-Stratiavalidity, $T \models \mathbf{T}_i^\alpha\vDash((\rho \leftrightarrow \sigma) \rightarrow (\rho \rightarrow \sigma))$. By i-Stratideduction,

$$\begin{aligned} T &\models \mathbf{T}_i^\alpha\vDash((\rho \leftrightarrow \sigma) \rightarrow (\rho \rightarrow \sigma)) \rightarrow \mathbf{T}_i^\alpha\vDash(\rho \leftrightarrow \sigma) \rightarrow \mathbf{T}_i^\alpha\vDash(\rho \rightarrow \sigma) \\ \text{and } T &\models \mathbf{T}_i^\alpha\vDash(\rho \rightarrow \sigma) \rightarrow \mathbf{T}_i^\alpha\vDash\rho \rightarrow \mathbf{T}_i^\alpha\vDash\sigma. \end{aligned}$$

It follows that $T \models \mathbf{T}_i^\alpha\vDash\rho \rightarrow \mathbf{T}_i^\alpha\vDash\sigma$. 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 $\sigma_1 \rightarrow \cdots \rightarrow \sigma_n \rightarrow \phi$ is valid. By instances of i-Stratiavalidity and i-Stratideduction contained in $T \cap \beta$, and by the last hypothesis of the theorem, $T \cap \beta \models \mathbf{T}_i^\alpha\vDash\phi$.

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 20 part 2).

Claim 3: If ϕ is a very 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 37, 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 38, 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 34, 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 41. Suppose $\mathbf{T} = (T_i)_{i \in \omega}$ is a family of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories. We say $\mathbf{T}$ is *stratifiable* if $\forall i \in \omega$, T_i has the following properties.

1. T_i is i-stratified.
2. Whenever $\phi \in T_i$ and $\mathbf{T}_i^\beta \vDash \phi$ is i-stratified, $\mathbf{T}_i^\beta \vDash \phi \in T_i$.
3. T_i is i-uniform.
4. T_i includes i-Strativalidity, i-Stratideduction, and i-Collapse.

Definition 42. If $\mathbf{T} = (T_i)_{i \in \omega}$ is a stratifiable 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 43. (The Stratification Theorem) Suppose $\mathbf{T} = (T_i)_{i \in \omega}$ is a stratifiable 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.

$$\begin{aligned}
\mathscr{M}_{\mathrm{Str}(\mathbf{T})} &\models \mathbf{T}_i^{\epsilon_0 \cdot n} \vDash \psi[s] && \\
T_i \cap (\epsilon_0 \cdot n) &\models \psi^s && (\text{Definition of } \mathscr{M}_{\mathrm{Str}(\mathbf{T})})\\
T_i &\models \psi^s && (\text{Theorem 31})\\
T_i^- &\models (\psi^s)^- && (\text{Theorem 40})\\
T_i^- &\models (\psi^-)^s && (\text{Clearly } (\psi^s)^- \equiv (\psi^-)^s)\\
\mathscr{M}_{\mathrm{Str}(\mathbf{T})} &\models \mathbf{T}_i \vDash \psi^-[s]. && (\text{Definition of } \mathscr{M}_{\mathrm{Str}(\mathbf{T})})
\end{aligned}$$

□

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 $\mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \rightarrow \phi)$ to T_i, because this is not necessarily i-stratified: for example, the particular instance $\mathbf{T}_j{\vDash}\mathbf{T}_i{\vDash}(1=0) \rightarrow \mathbf{T}_i{\vDash}(1=0)$ is not i-stratified. But neither is, say, $\mathbf{T}_j{\vDash}\mathbf{T}_i^\alpha{\vDash}(1=0) \rightarrow \mathbf{T}_i^\alpha{\vDash}(1=0)$, where $\mathbf{T}_i^\alpha{\vDash}$ occurs within the scope of $\mathbf{T}_j{\vDash}$. We will use a schema $\mathrm{ucl}(\mathbf{T}_j{\vDash}\phi \rightarrow \phi^+)$, where $\bullet^+$ varies over what we call i-stratifiers.

Definition 44. 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 $\neg\psi$, $\psi\rightarrow\rho$, or $\forall x\psi$, then ϕ^+ is $\neg\psi^+$, $\psi^+\rightarrow\rho^+$ 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

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

Lemma 45. Suppose $Z \subseteq \epsilon_0\cdot\omega$, $h : Z \rightarrow \epsilon_0\cdot\omega$ 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 46. Suppose $\mathscr{N}$ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure and $\bullet^+$ is an i-stratifier. We define an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{N}^+$ as follows. The universe and interpretation of arithmetic of $\mathscr{N}^+$ agree with those of $\mathscr{N}$, as do the interpretations of $\mathbf{T}_j{\vDash}$ ($j \neq i$) and $\mathbf{T}_j^\alpha{\vDash}$ (any α, j). As for $\mathbf{T}_i{\vDash}$, for any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s,

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

Lemma 47. (Compare Lemma 36) 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 $\mathscr{N} \models \phi^+[s]$.

$$\begin{tikzcd} \mathscr{L}_{\mathrm{PA}}(\omega) \arrow[rr, hook, "+"] \arrow[dr, "\mathscr{N}^+\vDash"'] & & \mathscr{L}_{\mathrm{PA}}(\mathcal{I}) \arrow[dl, "\mathscr{N}\vDash"] \\ & \{\mathrm{T},\mathrm{F}\} & \end{tikzcd}$$

Proof. By induction. □

Lemma 48. 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 47.

($\Leftarrow$) By Corollary 37. □

Given Lemma 48, Lemma 38 (which we promised to prove) is trivial.

Proof of Lemma 38. By Lemma 48 with $\bullet^+$ taken to be the i-veristratifier. □

For the remainder of the section, fix a strict r.e. well-founded partial-order $\prec$ on ω.

Definition 49. Suppose $\mathscr{M}$, $\mathscr{M}'$ are $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structures, $i \in \omega$. We say $\mathscr{M} \equiv_i \mathscr{M}'$ if there is a sequence

$$i_1, \bullet^1, i_2, \bullet^2, \ldots, i_n, \bullet^n,$$

each $i_k \succ i$, each $\bullet^k$ a computable i_k-stratifier, such that $\mathscr{M}' = (\cdots(\mathscr{M}^1)^2)\cdots)^n$.

Lemma 50. Suppose $\mathscr{M}, \mathscr{M}', \mathscr{M}''$ are $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structures and $i, j, k \in \omega$.

1. $\mathscr{M} \equiv_i \mathscr{M}$.
2. If $\mathscr{M} \equiv_i \mathscr{M}'$ then $\mathscr{M}$ and $\mathscr{M}'$ have the same universe and agree on all symbols of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$ except possibly for some symbols $\mathbf{T}_j\vDash$ where $j \succ i$.
3. If $\mathscr{M} \equiv_i \mathscr{M}'$ then for any $j \in \omega$, either $\mathscr{M}$ and $\mathscr{M}'$ agree on $\mathbf{T}_j\vDash$, or there is some j-stratifier $\bullet^+$ such that $\mathscr{M}^+$ and $\mathscr{M}'$ agree on $\mathbf{T}_j\vDash$.
4. If $\mathscr{M} \equiv_i \mathscr{M}'$, $\mathscr{M}' \equiv_j \mathscr{M}''$, and $k \preceq i, j$, then $\mathscr{M} \equiv_k \mathscr{M}''$.
5. If $\mathscr{M} \equiv_i \mathscr{M}'$ and $j \prec i$ then $\mathscr{M} \equiv_j \mathscr{M}'$.
6. If $j \prec i$ and $\bullet^+$ is a computable i-stratifier, then $\mathscr{M}^+ \equiv_j \mathscr{M}$.

Proof. Straightforward. □

Definition 51. If $\bullet^+$ is an i-stratifier, we extend the definition of ϕ^+ to all $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formulas ϕ by adding the clause: $(\mathbf{T}^\alpha_j\vDash\phi_0)^+ \equiv \mathbf{T}^\alpha_j\vDash\phi_0$ for every $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ_0. It is easy to check Lemma 47 extends to all $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formulas.

Lemma 52. Suppose $\mathscr{M}, \mathscr{M}'$ are $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structures and $\mathscr{M} \equiv_i \mathscr{M}'$ for some $i \in \omega$. Further suppose $\mathscr{M}$ has the property that for every $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^s$. Then for all such ϕ and s, $\mathscr{M}' \models \phi[s]$ if and only if $\mathscr{M}' \models \phi^s$.

Proof. By induction on sequence length, we may assume $\mathscr{M}' = \mathscr{M}^+$ for some computable j-stratifier $\bullet^+$, $j \succ i$. For any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, the following are equivalent.

$$\begin{aligned}
\mathscr{M}^+ &\models \phi[s] && \\
\mathscr{M} &\models \phi^+[s] && \text{(Lemma 47 and Definition 51)}\\
\mathscr{M} &\models (\phi^+)^s && \text{(Hypothesis)}\\
\mathscr{M} &\models (\phi^s)^+ && \text{(Clearly } (\phi^+)^s \equiv (\phi^s)^+\text{)}\\
\mathscr{M}^+ &\models \phi^s. && \text{(Lemma 47 and Definition 51)}
\end{aligned}$$

□

Lemma 53. Suppose the $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-structure $\mathscr{M}$ is an instance of Definition 7. For any $i \in \omega$, $\mathscr{M}' \equiv_i \mathscr{M}$, $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{M}' \models \phi[s]$ if and only if $\mathscr{M}' \models \phi^s$.

Proof. By Lemmas 8 and 52. □

7 Second Consistency Result: Prioritizing Recursive Enumerability

In this section we fulfil the second promise from the introduction. Throughout, $\prec$ is an r.e. well-founded partial-order of ω.

Definition 54. (Compare Definition 14) Suppose $\mathbf{T} = (T_i)_{i\in\omega}$ is an r.e. family of i-uniform i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories. We say $\mathbf{T}$ is *stratified-r.e.-generic* if for every stratifiable r.e. family $\mathbf{U} \supseteq \mathbf{T}$, every $i \in \omega$, and every $\mathscr{M} \equiv_i \mathscr{M}_{\mathrm{Str}(\mathbf{U})}$, $\mathscr{M} \models T_i$.

Lemma 55. 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 stratified-r.e.-generic families, then $\mathbf{T}$ is stratified-r.e.-generic.

Proof. Straightforward. $\square$

Lemma 56. (Compare Lemma 17) For any $i, j \in \mathbb{N}$, each of the following families is stratified-r.e.-generic.

1. $[i\text{-Stratideduction}]_i$.
2. $[j\text{-Deduction}]_i$ (if $j \not\succeq i$).
3. $[S]_i$ (if $j \neq i$) where S is: (Weak j-Deduction) $\mathrm{ucl}(\mathbf{T}_j{\vDash}(\phi \rightarrow \psi) \rightarrow \mathbf{T}_j{\vDash}\phi \rightarrow \mathbf{T}_j{\vDash}(\psi \wedge (\phi \vee \neg\phi)))$.
4. $[S]_i$ where S is: (i-Assigned Strativalidity) the schema ϕ^s (ϕ valid and i-stratified, s an assignment).
5. $[i\text{-Assigned Strativalidity}]_i \cup [i\text{-Strativalidity}]_i$.
6. $[i\text{-Assigned Strativalidity}]_i \cup [i\text{-Validity}]_j$ (if $j \neq i$).
7. $[i\text{-Assigned Strativalidity}]_i \cup [i\text{-Strativalidity}]_i \cup [i\text{-Stratideduction}]_i \cup [i\text{-Introspection}]_j$ ($j \neq i$).
8. $[i\text{-Assigned Strativalidity}]_i \cup [i\text{-Strativalidity}]_i \cup [i\text{-Stratideduction}]_i \cup [S]_i$ where S is:

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

9. $[S]_i$ where S is the set of those axioms of Peano arithmetic for $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$ that are i-stratified.
10. $[S]_i$ where S is any r.e. set of true arithmetic sentences.
11. $[j\text{-SMT}]_i$ ($j \neq i$).
12. $[S]_i$, where S is: (i-Strati-SMT) $\mathrm{ucl}(\exists e \forall x(\mathbf{T}_i^\alpha{\vDash}\phi \leftrightarrow x \in W_e))$ when this is i-stratified, $e \notin \mathrm{FV}(\phi)$.
13. $\mathbf{T} \cup [S]_i$ where $\mathbf{T} = (T_k)_{k\in\omega}$ is stratified-r.e.-generic and S is the schema $\mathbf{T}_i^\alpha{\vDash}\phi$ ($\phi \in T_i$ such that this is i-stratified).

Proof. Mostly straightforward. For uniformity of 4–8, use Lemma 27. Uniformity of the other families is clear. Stratification is immediate in all cases. Recursive enumerability follows from the fact that $\prec$ is r.e. We sketch 2, 3, and 6 to highlight subtle points. In each case, let $\mathbf{U} = (U_k)_{k\in\omega}$ be a stratifiable r.e. family extending the family in question. For brevity let $\hat{\mathbf{U}} = \mathrm{Str}(\mathbf{U})$.

(2) Let $\mathscr{M} \equiv_i \mathscr{M}_{\hat{\mathbf{U}}}$, we must show $\mathscr{M} \models \mathrm{ucl}(\mathbf{T}_j{\vDash}(\phi \rightarrow \psi) \rightarrow \mathbf{T}_j{\vDash}\phi \rightarrow \mathbf{T}_j{\vDash}\psi)$. Let s be an assignment and assume $\mathscr{M} \models \mathbf{T}_j{\vDash}(\phi \rightarrow \psi)[s]$ and $\mathscr{M} \models \mathbf{T}_j{\vDash}\phi[s]$. Since $j \not\succeq i$, Lemma 50 says $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}}$ agree on $\mathbf{T}_j{\vDash}$. By definition of $\mathscr{M}_{\hat{\mathbf{U}}}$, $U_j^- \models \phi^s \rightarrow \psi^s$ and $U_j^- \models \phi^s$, thus $U_j^- \models \psi^s$, so $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_j{\vDash}\psi[s]$ and so does $\mathscr{M}$.

(3) Let $\mathscr{M} \equiv_i \mathscr{M}_{\hat{\mathbf{U}}}$, we must show $\mathscr{M} \models \mathrm{ucl}(\mathbf{T}_j{\vDash}(\phi \rightarrow \psi) \rightarrow \mathbf{T}_j{\vDash}\phi \rightarrow \mathbf{T}_j{\vDash}(\psi \wedge (\phi \vee \neg\phi)))$. Let s be an assignment and assume $\mathscr{M} \models \mathbf{T}_j{\vDash}(\phi \rightarrow \psi)[s]$ and $\mathscr{M} \models \mathbf{T}_j{\vDash}\phi[s]$. If $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}}$ agree on $\mathbf{T}_j{\vDash}$, reason as in (2) above. If not, Lemma 50 says there is some j-stratifier $\bullet^+$ such that $\mathscr{M}$ and $\mathscr{M}^+_{\hat{\mathbf{U}}}$ agree on $\mathbf{T}_j{\vDash}$. By definition of $\mathscr{M}^+_{\hat{\mathbf{U}}}$, $\mathscr{M}_{\hat{\mathbf{U}}} \models (\mathbf{T}_j{\vDash}(\phi \rightarrow \psi))^+[s]$ and $\mathscr{M}_{\hat{\mathbf{U}}} \models (\mathbf{T}_j{\vDash}\phi)^+[s]$. Let $\alpha, \beta \in \epsilon_0 \cdot \omega$ be such that $(\mathbf{T}_j{\vDash}(\phi \rightarrow \psi))^+ \equiv \mathbf{T}_j^\alpha{\vDash}(\phi^+ \rightarrow \psi^+)$ and $(\mathbf{T}_j{\vDash}\phi)^+ \equiv \mathbf{T}_j^\beta{\vDash}\phi^+$. Then $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_j^\alpha{\vDash}(\phi^+ \rightarrow \psi^+)[s]$ and $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_j^\beta{\vDash}\phi^+[s]$. This means $U_j \cap \alpha \models (\phi^+ \rightarrow \psi^+)^s$ and $U_j \cap \beta \models (\phi^+)^s$. Since ϕ is a subformula of $\phi \rightarrow \psi$, 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 $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_j^\alpha{\vDash}(\psi^+ \wedge (\phi^+ \vee \neg\phi^+))[s]$. By Definition 44,

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

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

(6) By (4), $\widetilde{\mathscr{M}} \models i$-Assigned Strativalidity whenever $\widetilde{\mathscr{M}} \equiv_i \mathscr{M}_{\hat{\mathbf{U}}}$. Let $\mathscr{M} \equiv_j \mathscr{M}_{\hat{\mathbf{U}}}$, we will show $\mathscr{M} \models$ i-Validity. Let ϕ be a valid $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula, s an assignment.

Case 1: $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}}$ agree on $\mathbf{T}_i\vDash$. Let $\bullet^+$ be an i-stratifier. Since ϕ is valid, so is ϕ^+ (by Lemma 48), so $(\phi^+)^s \in U_i$ (since [i-Assigned Strativalidity]$_i$ is part of line 6). Clearly $((\phi^+)^s)^- \equiv \phi^s$, so $\phi^s \in U_i^-$, thus $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_i\vDash\phi[s]$, and so does $\mathscr{M}$.

Case 2: $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}}$ disagree on $\mathbf{T}_i\vDash$. By Lemma 50 there is an i-stratifier $\bullet^+$ such that $\mathscr{M}$ and $\mathscr{M}^+_{\hat{\mathbf{U}}}$ 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, $(\phi^+)^s \in U_i$. In fact by choice of α, $(\phi^+)^s \in U_i \cap \alpha$, so $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_i^\alpha\vDash\phi^+[s]$, that is, $\mathscr{M}_{\hat{\mathbf{U}}} \models (\mathbf{T}_i\vDash\phi)^+[s]$. By Lemma 47, $\mathscr{M}^+_{\hat{\mathbf{U}}} \models \mathbf{T}_i\vDash\phi[s]$. ☐

Definition 57. 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 *stratifiable-r.e.-generic* if there is some stratified-r.e.-generic family $\mathbf{T} = (T_i)_{i\in\omega}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories such that each $T_i^- = T_i^0$.

Theorem 58. Let $\mathbf{T}^0 = (T_i^0)_{i\in\omega}$ 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.

- The axioms contained in T_i^0.
- Assigned Validity, i-Validity and i-Deduction.
- $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi)$ whenever $j \preceq i$.
- $\forall x(\mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}})$ 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 $W_n = W_{f(n)}$. Fix this n for the rest of the proof and write $\mathbf{T}$ for $\mathbf{T}(n)$, T_i for $T_i(n)$.

Since $\mathbf{T}^0$ is stratifiable-r.e.-generic, there is a stratified-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.

- U_i contains V_i.
- U_i contains i-Assigned Strativalidity, i-Strativalidity, i-Stratideduction and i-Collapse.
- U_i contains $\mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi \rightarrow \phi)$ whenever $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified.
- U_i contains $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi^+)$ for every $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, $j \prec i$, and i-stratifier $\bullet^+$.
- 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.
- 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 stratifiable r.e. (to see $\mathbf{U}$ is uniform, use Lemma 45, to see $\mathbf{U}$ is r.e., use Theorem 20 part 1); $\mathbf{U} \supseteq \mathbf{V}$; and for each $i \in \omega$, $U_i^- = T_i$.

Let $\mathbf{S} = (S_i)_{i\in\mathcal{I}} = \mathrm{Str}(\mathbf{U})$. By definition this means that for all $i \in \omega$ and $\alpha \in \epsilon_0 \cdot \omega$,

$$S_i = U_i^- = T_i \text{ and } S_{(\alpha,i)} = U_i \cap \alpha.$$

In order to show $\mathscr{M}_{\mathbf{T}} \models \mathbf{T}$ and thus prove the theorem, we will show $\mathscr{M}_{\mathbf{S}} \models \mathbf{S}$, which is more than sufficient, since $\mathscr{M}_{\mathbf{S}}$ and $\mathscr{M}_{\mathbf{T}}$ agree on $\mathscr{L}_{\mathrm{PA}}(\omega)$. But for sake of a stronger induction hypothesis, we will prove more. We will prove that for every $i \in \omega$, every $j \preceq i$, every $\mathscr{M} \equiv_i \mathscr{M}_{\mathbf{S}}$, and every $\alpha \in \epsilon_0 \cdot \omega$, $\mathscr{M} \models S_j \cup S_{(\alpha,j)}$.

Fix $i \in \omega$. Since $\prec$ is well-founded, we may assume the following:

(∗) For every $k \preceq j \prec i$, every $\mathscr{M} \equiv_j \mathscr{M}_{\mathbf{S}}$, and every $\alpha \in \epsilon_0 \cdot \omega$, $\mathscr{M} \models S_k \cup S_{(\alpha,k)}$.

Fix $\mathscr{M} \equiv_i \mathscr{M}_{\mathbf{S}}$. For all $j \prec i$, Lemma 50 says $\mathscr{M} \equiv_j \mathscr{M}_{\mathbf{S}}$ and therefore by (∗) we already have $\mathscr{M} \models S_j \cup S_{(\alpha,j)}$. It remains to show $\forall\alpha \in \epsilon_0 \cdot \omega$, $\mathscr{M} \models S_i \cup S_{(\alpha,i)}$.

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

By induction on α. Let $\sigma \in S_{(\alpha,i)}$. This means $\sigma \in U_i \cap \alpha$.

Case 1: $\sigma \in V_i$. Then $\mathscr{M} \models \sigma$ because $\mathbf{V}$ is stratified-r.e.-generic and $\mathbf{U} \supseteq \mathbf{V}$ is stratifiable r.e.

Case 2: σ is an instance of i-Assigned Strativalidity, i-Strativalidity, or i-Stratideduction. Then $\mathscr{M} \models \sigma$ by Lemma 56.

Case 3: σ is $\mathrm{ucl}(\mathbf{T}_i^{\alpha_0}\vDash\phi \rightarrow \phi)$ 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 $\mathscr{M} \models \mathbf{T}_i^{\alpha_0}\vDash\phi[s]$, then:

$$\begin{aligned}
\mathscr{M} &\models \mathbf{T}_i^{\alpha_0}\vDash\phi[s] && \text{(Assumption)}\\
\mathscr{M}_{\mathbf{S}} &\models \mathbf{T}_i^{\alpha_0}\vDash\phi[s] && \text{($\mathscr{M}$ and $\mathscr{M}_{\mathbf{S}}$ agree on $\mathbf{T}_i^{\alpha_0}\vDash$ by Lemma 50)}\\
S_{(\alpha_0,i)} &\models \phi^s && \text{(Definition of $\mathscr{M}_{\mathbf{S}}$)}\\
\mathscr{M} &\models \phi^s && \text{(By induction, $\mathscr{M} \models S_{(\alpha_0,i)}$)}\\
\mathscr{M} &\models \phi[s]. && \text{(By Lemma 53)}
\end{aligned}$$

Case 4: σ is $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi^+)$ for some $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, $j \prec i$, and i-stratifier $\bullet^+$. By Lemma 45 we may assume $\bullet^+$ is computable. Let s be an assignment and assume $\mathscr{M} \models \mathbf{T}_j\vDash\phi[s]$, then:

$$\begin{aligned}
\mathscr{M} &\models \mathbf{T}_j\vDash\phi[s] && \text{(Assumption)}\\
\mathscr{M}_{\mathbf{S}} &\models \mathbf{T}_j\vDash\phi[s] && \text{(Since $j \prec i$, $\mathscr{M}$ and $\mathscr{M}_{\mathbf{S}}$ agree on $\mathbf{T}_j\vDash$ by Lemma 50)}\\
T_j &\models \phi^s && \text{(Definition of $\mathscr{M}_{\mathbf{S}}$)}\\
\mathscr{M}^+ &\models \phi^s && \text{(By Lemma 50, $\mathscr{M}^+ \equiv_j \mathscr{M}_{\mathbf{S}}$, so by $(*)$, $\mathscr{M}^+ \models T_j$)}\\
\mathscr{M} &\models (\phi^s)^+ && \text{(Lemma 47)}\\
\mathscr{M} &\models (\phi^+)^s && \text{(Clearly $(\phi^s)^+ \equiv (\phi^+)^s$)}\\
\mathscr{M} &\models \phi^+[s]. && \text{(Lemma 53)}
\end{aligned}$$

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:

$$\begin{aligned}
\mathscr{M} &\models \mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}}[s] && \\
\mathscr{M}_{\mathbf{S}} &\models \mathbf{T}_j\vDash\phi \leftrightarrow \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, x\rangle \in W_{\overline{n}}[s] && \text{($\mathscr{M}$ and $\mathscr{M}_{\mathbf{S}}$ agree on the symbols in question)}\\
\mathscr{M}_{\mathbf{S}} &\models \mathbf{T}_j\vDash\phi[s] \text{ iff } \mathscr{M}_{\mathbf{S}} \models \langle\overline{\ulcorner\phi\urcorner}, \overline{j}, \overline{m}\rangle \in W_{\overline{n}} && \text{(Lemma 53)}\\
\mathscr{M}_{\mathbf{S}} &\models \mathbf{T}_j\vDash\phi[s] \text{ iff } \langle\ulcorner\phi\urcorner, j, m\rangle \in W_n && \text{($\mathscr{M}_{\mathbf{S}}$ has standard first-order part)}\\
T_j &\models \phi^s \text{ iff } \langle\ulcorner\phi\urcorner, j, m\rangle \in W_n && \text{(Definition of $\mathscr{M}_{\mathbf{S}}$)}\\
T_j &\models \phi(x|\overline{m}) \text{ iff } \langle\ulcorner\phi\urcorner, j, m\rangle \in W_n. && \text{(Since $\mathrm{FV}(\phi) \subseteq \{x\}$)}
\end{aligned}$$

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 $\mathscr{M}$ and $\mathscr{M}_{\mathbf{S}}$ agree on $\mathbf{T}_i^{\beta}\vDash$ and $\mathbf{T}_i^{\gamma}\vDash$, we need only show $\mathscr{M}_{\mathbf{S}} \models \mathbf{T}_i^{\beta}\vDash\phi \leftrightarrow \mathbf{T}_i^{\gamma}\vDash\phi[s]$. 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 31.

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 $\mathscr{M}_{\mathbf{S}} \models \mathbf{T}_i^{\alpha_0}\vDash\phi$, so $\mathscr{M} \models \mathbf{T}_i^{\alpha_0}\vDash\phi$ since $\mathscr{M}$ and $\mathscr{M}_{\mathbf{S}}$ agree on $\mathbf{T}_i^{\alpha_0}\vDash$.

Cases 1–7 establish $\mathscr{M} \models S_{(\alpha,i)}$. 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, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^-[s]$.

By induction on ϕ. 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 is $\mathbf{T}_j^{\alpha}\vDash\psi$ for some $j \neq i$. Impossible, this is not i-stratified.

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

$$\begin{aligned}
\mathscr{M} &\models (\mathbf{T}_i^\alpha \vDash \psi)^-[s] & & \\
\mathscr{M}_\mathbf{S} &\models (\mathbf{T}_i^\alpha \vDash \psi)^-[s] & & (\mathscr{M} \text{ and } \mathscr{M}_\mathbf{S} \text{ agree on } \mathbf{T}_i \vDash) \\
\mathscr{M}_\mathbf{S} &\models \mathbf{T}_i^\alpha \vDash \psi[s] & & (\text{Theorem 43}) \\
\mathscr{M} &\models \mathbf{T}_i^\alpha \vDash \psi[s]. & & (\mathscr{M} \text{ and } \mathscr{M}_S \text{ agree on } \mathbf{T}_i^\alpha \vDash)
\end{aligned}$$

Claim 3: $\mathscr{M} \models S_i$.

For any $\sigma \in S_i$, there is some $\tau \in U_i$ such that $\tau^- \equiv \sigma$; since U_i is i-uniform, we may take τ to be very i-stratified (Lemma 34). By Claim 1, $\mathscr{M} \models U_i$, so $\mathscr{M} \models \tau$. By Claim 2, $\mathscr{M} \models \sigma$. □

This satisfies the second promise from the introduction: given a well-founded r.e. partial order $\prec$ on ω, we have exhibited 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 56 parts 11–12.

8 Third Consistency Result: Prioritizing Truth

In this section, $\prec$ remains a well-founded, r.e. partial order on ω.

Definition 59. For any index set I and $\mathscr{L}_{\mathrm{PA}}(I)$-structure $\mathscr{M}$, let $\mathscr{M} \cap \mathrm{Tr}$ be the $\mathscr{L}_{\mathrm{PA}}(I)$-structure that has the same first-order part as $\mathscr{M}$ and is otherwise defined recursively so that for every $i \in I$, every $\mathscr{L}_{\mathrm{PA}}(I)$-formula ϕ, and every assignment s,

$$\mathscr{M} \cap \mathrm{Tr} \models \mathbf{T}_i \vDash \phi[s] \text{ if and only if } \mathscr{M} \models \mathbf{T}_i \vDash \phi[s] \text{ and } \mathscr{M} \cap \mathrm{Tr} \models \phi[s].$$

Lemma 60. Suppose I is an index set and $\mathscr{M}$ is an $\mathscr{L}_{\mathrm{PA}}(I)$-structure such that for every $\mathscr{L}_{\mathrm{PA}}(I)$-formula ϕ and assignment s, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^s$. Then for all such ϕ and s, $\mathscr{M} \cap \mathrm{Tr} \models \phi[s]$ if and only if $\mathscr{M} \cap \mathrm{Tr} \models \phi^s$.

Proof. By induction on complexity of ϕ. The interesting case is when ϕ is $\mathbf{T}_i \vDash \psi$ for some $i \in I$. The following are equivalent.

$$\begin{aligned}
\mathscr{M} \cap \mathrm{Tr} &\models \mathbf{T}_i \vDash \psi[s] & & \\
\mathscr{M} \models \mathbf{T}_i \vDash \psi[s] \;\&\; \mathscr{M} \cap \mathrm{Tr} &\models \psi[s] & & (\text{Definition of } \mathscr{M} \cap \mathrm{Tr}) \\
\mathscr{M} \models \mathbf{T}_i \vDash \psi^s \;\&\; \mathscr{M} \cap \mathrm{Tr} &\models \psi^s & & (\text{Left: Hypothesis, Right: Induction}) \\
\mathscr{M} \models \mathbf{T}_i \vDash \psi^s[s] \;\&\; \mathscr{M} \cap \mathrm{Tr} &\models \psi^s[s] & & (\psi^s \text{ is a sentence}) \\
\mathscr{M} \cap \mathrm{Tr} &\models \mathbf{T}_i \vDash \psi^s[s] & & (\text{Definition of } \mathscr{M} \cap \mathrm{Tr}) \\
\mathscr{M} \cap \mathrm{Tr} &\models \mathbf{T}_i \vDash \psi^s. & & (\psi^s \text{ is a sentence})
\end{aligned}$$

□

Lemma 61. Suppose I is an index set, $\mathbf{T} = (T_i)_{i\in I}$ is a family of $\mathscr{L}_{\mathrm{PA}}(I)$-theories, ϕ is an $\mathscr{L}_{\mathrm{PA}}(I)$-formula, s is an assignment, and $i \in I$. Further, suppose $\mathscr{M}_\mathbf{T} \cap \mathrm{Tr} \models T_i$. Then $\mathscr{M}_\mathbf{T} \cap \mathrm{Tr} \models \mathbf{T}_i \vDash \phi[s]$ if and only if $\mathscr{M}_\mathbf{T} \models \mathbf{T}_i \vDash \phi[s]$.

Proof. Assume the hypotheses. The following are equivalent.

$$\begin{aligned}
\mathscr{M}_\mathbf{T} \cap \mathrm{Tr} &\models \mathbf{T}_i \vDash \phi[s] & & \\
\mathscr{M}_\mathbf{T} \models \mathbf{T}_i \vDash \phi[s] \;\&\; \mathscr{M}_\mathbf{T} \cap \mathrm{Tr} &\models \phi[s] & & (\text{Definition of } \mathscr{M}_\mathbf{T} \cap \mathrm{Tr}) \\
T_i \models \phi^s \;\&\; \mathscr{M}_\mathbf{T} \cap \mathrm{Tr} &\models \phi[s] & & (\text{Definition of } \mathscr{M}_\mathbf{T}) \\
T_i \models \phi^s \;\&\; \mathscr{M}_\mathbf{T} \cap \mathrm{Tr} &\models \phi^s & & (\text{Lemmas 8 and 60}) \\
T_i &\models \phi^s & & (\text{By hypothesis, } \mathscr{M}_\mathbf{T} \cap \mathrm{Tr} \models T_i) \\
\mathscr{M}_\mathbf{T} &\models \mathbf{T}_i \vDash \phi[s]. & & (\text{Definition of } \mathscr{M}_\mathbf{T})
\end{aligned}$$

□

Definition 62. (Compare Definition 54) Suppose $\mathbf{T} = (T_i)_{i\in\omega}$ is an r.e. family of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories such that each T_i is an i-uniform set of i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-sentences. We say $\mathbf{T}$ is *stratified-true-generic* if for every stratifiable r.e. family $\mathbf{U} \supseteq \mathbf{T}$, every $i \in \omega$, and every $\mathscr{M} \equiv_i \mathscr{M}_{\mathrm{Str}(\mathbf{U})} \cap \mathrm{Tr}$, $\mathscr{M} \models T_i$.

Lemma 63. If the family $\mathbf{T} = (T_i)_{i\in\omega}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories is r.e. and is a union of stratified-true-generic families, then $\mathbf{T}$ is stratified-true-generic.

Proof. Straightforward. $\square$

Definition 64. Suppose ϕ is an $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula. By the *exposed indices of* ϕ we mean the set of those $i \in \omega$ such that there is an occurrence of $\mathbf{T}_i\vDash$ in ϕ not within the scope of $\mathbf{T}_j\vDash$ or $\mathbf{T}_j^\alpha\vDash$ for any j, α.

For example, the exposed indices of $\mathbf{T}_5\vDash(1=0) \wedge \mathbf{T}_6\vDash\mathbf{T}_7\vDash(1=0) \wedge \mathbf{T}_8^\omega\vDash\mathbf{T}_9\vDash(1=0)$ are 5 and 6.

Lemma 65. (Compare Lemmas 17 and 56) For any $i, j \in \omega$, each of the following families is statified-true-generic.

1. Any of the families from lines 1–10 of Lemma 56.
2. $[S]_i$ $(j \neq i)$ where S is: (i-Weak j-Truth) $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi)$ for any $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ such that for every exposed index k of ϕ, $k \prec i$.
3. $[S]_i$ where S is: (i-Weak i-Stratotruth) $\mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi \rightarrow \phi)$ for any $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ such that $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified and for every exposed index k of ϕ, $k \prec i$.
4. $\mathbf{T} \cup [S]_i$ where $\mathbf{T} = (T_k)_{k\in\omega}$ is stratified-true-generic and S is the schema $\mathbf{T}_i^\alpha\vDash\phi$ ($\phi \in T_i$ such that $\mathbf{T}_i^\alpha\vDash\phi$ is i-stratified).

Proof. Mostly straightforward and similar to Lemma 56. We prove (2) for illustrative purposes.

Let $\mathbf{U}$ be a stratifiable r.e. family extending $[i\text{-Weak } j\text{-Truth}]_i$, $j \neq i$. For brevity let $\hat{\mathbf{U}} = \mathrm{Str}(\mathbf{U})$. Let ϕ be an $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula such that all exposed indices of ϕ are $\prec i$. Assume $\mathscr{M} \equiv_i \mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr}$, we will show $\mathscr{M} \models \mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi)$. Let s be an assignment and assume $\mathscr{M} \models \mathbf{T}_j\vDash\phi[s]$.

Case 1: $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr}$ agree on $\mathbf{T}_j\vDash$. Then $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models \mathbf{T}_j\vDash\phi[s]$. This means $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_j\vDash\phi[s]$ and $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models \phi[s]$. Since $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr}$ agree on all $\mathbf{T}_k\vDash$ with $k \prec i$, $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr}$ agree on $\mathbf{T}_k\vDash$ whenever k is an exposed index of ϕ, and it follows that, since $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models \phi[s]$, $\mathscr{M} \models \phi[s]$.

Case 2: $\mathscr{M}$ and $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr}$ disagree on $\mathbf{T}_j\vDash$. By Lemma 50, there is a j-stratifier $\bullet^+$ such that $\mathscr{M}$ and $(\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr})^+$ agree on $\mathbf{T}_j\vDash$. Thus $(\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr})^+ \models \mathbf{T}_j\vDash\phi[s]$. By Lemma 47, $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models (\mathbf{T}_j\vDash\phi)^+[s]$, so $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models \mathbf{T}_j^\alpha\vDash\phi^+[s]$ for some $\alpha \in \epsilon_0 \cdot \omega$. Since i is not an exposed index in ϕ, it follows that $\phi^+ \equiv \phi$, so $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models \mathbf{T}_j^\alpha\vDash\phi[s]$. By definition this means $\mathscr{M}_{\hat{\mathbf{U}}} \models \mathbf{T}_j^\alpha\vDash\phi[s]$ and $\mathscr{M}_{\hat{\mathbf{U}}} \cap \mathrm{Tr} \models \phi[s]$. From here the proof is as in Case 1. $\square$

Definition 66. If $\mathbf{T}^0 = (T_i^0)_{i\in\omega}$ is a family of $\mathscr{L}_{\mathrm{PA}}(\omega)$-theories, we say $\mathbf{T}^0$ is *stratifiable-true-generic* if there is some stratified-true-generic family $\mathbf{T} = (T_i)_{i\in\mathbb{N}}$ of $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theories such that each $T_i^- = T_i^0$.

Theorem 67. Let $\mathbf{T}^0 = (T_i^0)_{i\in\omega}$ be any stratifiable-true-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.

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

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. Let n be obtained the same way as in the proof of Theorem 58. Fix this n and write $\mathbf{T}$ for $\mathbf{T}(n)$, T_i for $T_i(n)$.

Since $\mathbf{T}^0$ is stratifiable-true-generic, there is a stratified-true-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 each $i \in \omega$, let U_i be the smallest $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-theory such that the following hold.

- U_i contains V_i.
- U_i contains i-Assigned Strativalidity, i-Strativalidity, i-Stratideduction and i-Collapse.
- U_i contains $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi^+)$ for every $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, $j \prec i$, and i-stratifier $\bullet^+$.
- U_i contains $\mathrm{ucl}(\mathbf{T}_i^\alpha\vDash\phi \rightarrow \phi)$ for every $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ such that this is i-stratified.
- 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.
- U_i contains $\mathrm{ucl}(\exists e\forall x(\mathbf{T}_i^\alpha\vDash\phi \leftrightarrow x \in W_e))$ whenever this is i-stratified and $e \notin \mathrm{FV}(\phi)$.
- U_i contains i-Collapse.
- 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 $\mathbf{U} = (U_i)_{i\in\omega}$, note $\mathbf{U}$ is stratifiable r.e., $\mathbf{U} \supseteq \mathbf{V}$, and each $U_i^- = T_i$.

Let $\mathbf{S} = (S_i)_{i\in\mathcal{I}} = \mathrm{Str}(\mathbf{U})$. This means that for all $i \in \omega$ and $\alpha \in \epsilon_0\cdot\omega$, $S_i = U_i^- = T_i$ and $S_{(\alpha,i)} = U_i \cap \alpha$.

To show $\mathscr{M}_\mathbf{T} \models \mathbf{T}$ (proving the theorem), we will first prove $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models \mathbf{S}$. Then we will argue that in fact $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} = \mathscr{M}_\mathbf{S}$. Having established this, we will be finished, since $\mathscr{M}_\mathbf{S}$ and $\mathscr{M}_\mathbf{T}$ agree on $\mathscr{L}_{\mathrm{PA}}(\omega)$. For sake of a strong induction hypothesis, we will prove more. We will prove that $\forall i \in \omega$, $\forall j \preceq i$, for every $\mathscr{M} \equiv_i \mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$ and every $\alpha \in \epsilon_0\cdot\omega$, $\mathscr{M} \models S_j \cup S_{(\alpha,j)}$.

Fix $i \in \omega$. Since $\prec$ is well-founded we may assume:

$$(*)\ \text{For all } k \preceq j \prec i,\ \mathscr{M} \equiv_j \mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \text{ and } \alpha \in \epsilon_0\cdot\omega,\ \mathscr{M} \models S_k \cup S_{(\alpha,k)}.$$

Fix $\mathscr{M} \equiv_i \mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$. For all $j \prec i$, $\mathscr{M} \equiv_j \mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$ and $(*)$ already gives $\mathscr{M} \models S_j \cup U_j$, it remains to show $\forall\alpha \in \epsilon_0\cdot\omega$, $\mathscr{M} \models S_i \cup S_{(\alpha,i)}$.

Claim 1: For any $j \prec i$, any $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, and any assignment s, $\mathscr{M}_\mathbf{S} \models \mathbf{T}_j\vDash\phi[s]$ if and only if $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models \mathbf{T}_j\vDash\phi[s]$.

By $(*)$, $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models S_j$. The claim now follows by Lemma 61.

Claim 2: $\forall\alpha \in \epsilon_0\cdot\omega$, $\mathscr{M} \models S_{(\alpha,i)}$.

By induction on α. Let $\sigma \in S_{(\alpha,i)}$. This means $\sigma \in U_i \cap \alpha$.

Case 1: $\sigma \in V_i$. Then $\mathscr{M} \models \sigma$ because $\mathbf{V}$ is stratified-true-generic, $\mathbf{U} \supseteq \mathbf{V}$ is stratifiable r.e., and $\mathscr{M} \equiv_i \mathscr{M}_{\mathrm{Str}(\mathbf{U})} \cap \mathrm{Tr}$.

Case 2: σ is an instance of i-Assigned Strativalidity, i-Strativalidity, or i-Stratideduction. Then $\mathscr{M} \models \sigma$ by Lemma 65.

Cases 3–4: σ is $\mathrm{ucl}(\mathbf{T}_j\vDash\phi \rightarrow \phi^+)$ for some $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ, $j \prec i$, and i-stratifier $\bullet^+$, or σ is $\mathrm{ucl}(\mathbf{T}_i^{\alpha_0}\vDash\phi \rightarrow \phi)$ for some $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ such that $\mathbf{T}_i^{\alpha_0}\vDash\phi \rightarrow \phi$ is i-stratified. Either case is similar to the corresponding case from the proof of Theorem 58.

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 $j \prec i$ and some $\mathscr{L}_{\mathrm{PA}}(\omega)$-formula ϕ with $\mathrm{FV}(\phi) \subseteq \{x\}$. Similar to the corresponding case from Theorem 58, using Claim 1 to equate the interpretation of $\mathbf{T}_j\vDash$ in $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$ with that in $\mathscr{M}_\mathbf{S}$.

Case 6: σ is $\mathrm{ucl}(\exists e\forall x(\mathbf{T}_i^\beta\vDash\phi \leftrightarrow x \in W_e))$ for some ϕ such that this formula is i-stratified, $e \notin \mathrm{FV}(\phi)$. Since $\sigma \in S_{(\alpha,i)}$, $\beta < \alpha$. Thus by induction, $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models S_{(\beta,i)}$, so by Lemma 61, $\mathscr{M}_\mathbf{S}$ and $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$ (and hence $\mathscr{M}$) agree on $\mathbf{T}_i^\beta\vDash$. Thus to see $\mathscr{M} \models \sigma$ it suffices to notice that for any assignment s, $\{m \in \mathbb{N} : U_i \cap \beta \models \phi(x|\overline{m})^s\}$ is r.e.

Case 7: σ is an instance $\mathrm{ucl}(\mathbf{T}_i^\beta{\vDash}\phi \leftrightarrow \mathbf{T}_i^\gamma{\vDash}\phi)$ of i-Collapse. Since $\sigma \in S_{(\alpha,i)}$, $\beta,\gamma < \alpha$, so by induction, $\mathscr{M} \models S_{(\beta,i)} \cup S_{(\gamma,i)}$. Lemma 61 therefore says that $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$ and $\mathscr{M}_\mathbf{S}$ agree on $\mathbf{T}_i^\beta{\vDash}$ and $\mathbf{T}_i^\gamma{\vDash}$. From here, the proof is similar to the corresponding case from Theorem 58.

Case 8: σ is $\mathbf{T}_i^{\alpha_0}{\vDash}\phi$ for some $\phi \in U_i$ such that $\mathbf{T}_i^{\alpha_0}{\vDash}\phi$ is i-stratified. Since $\sigma \in S_{(\alpha,i)}$, $\alpha_0 < \alpha$; since $\mathbf{T}_i^{\alpha_0}{\vDash}\phi$ is i-stratified, $\mathrm{On}(\phi) \subseteq \alpha_0$ and so $\phi \in S_{(\alpha_0,i)}$. By induction, $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models S_{(\alpha_0,i)}$, so $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models \phi$. And certainly $\mathscr{M}_\mathbf{S} \models \mathbf{T}_i^{\alpha_0}{\vDash}\phi$. Together these show $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models \mathbf{T}_i^{\alpha_0}{\vDash}\phi$. Since $\mathscr{M}$ and $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$ agree on $\mathbf{T}_i^{\alpha_0}{\vDash}$, $\mathscr{M} \models \mathbf{T}_i^{\alpha_0}{\vDash}\phi$.

Cases 1–8 show that $\mathscr{M} \models S_{(\alpha,i)}$, proving Claim 2.

Claim 3: For any very i-stratified $\mathscr{L}_{\mathrm{PA}}(\mathcal{I})$-formula ϕ and assignment s, $\mathscr{M} \models \phi[s]$ if and only if $\mathscr{M} \models \phi^-[s]$.

We'll prove more: for any $\mathscr{M}' \equiv_i \mathscr{M}_\mathbf{S} \cap \mathrm{Tr}$, $\mathscr{M}' \models \phi[s]$ if and only if $\mathscr{M}' \models \phi^-[s]$. The proof is by induction on ϕ. The 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 is $\mathbf{T}_j^\alpha{\vDash}\psi$ for some $j \neq i$, $\alpha \in \epsilon_0 \cdot \omega$. Impossible, this is not i-stratified.

Case 3: ϕ is $\mathbf{T}_i^\alpha{\vDash}\psi$ for some $\alpha \in \epsilon_0 \cdot \omega$. Then $\phi^- \equiv \mathbf{T}_i{\vDash}\psi^-$. Since ϕ is very i-stratified, $\alpha = \epsilon_0 \cdot n$ for some $n > 0$. By Condition 2 of Definition 44, $\mathrm{On}(\psi) \subseteq \epsilon_0 \cdot n$. The following are equivalent.

$$\begin{aligned}
\mathscr{M}' &\models \mathbf{T}_i^{\epsilon_0\cdot n}{\vDash}\psi[s] && \\
\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} &\models \mathbf{T}_i^{\epsilon_0\cdot n}{\vDash}\psi[s] && \text{(Since } \mathscr{M}' \equiv_i \mathscr{M}_\mathbf{S} \cap \mathrm{Tr},\ \mathscr{M}' \text{ and } \mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \text{ agree on } \mathbf{T}_i^{\epsilon_0\cdot n}{\vDash})\\
\mathscr{M}_\mathbf{S} \models \mathbf{T}_i^{\epsilon_0\cdot n}{\vDash}\psi[s] \;\&\; \mathscr{M}_\mathbf{S} \cap \mathrm{Tr} &\models \psi[s] && \text{(Definition of } \mathscr{M}_\mathbf{S} \cap \mathrm{Tr})\\
\mathscr{M}_\mathbf{S} \models \mathbf{T}_i^{\epsilon_0\cdot n}{\vDash}\psi[s] \;\&\; \mathscr{M}_\mathbf{S} \cap \mathrm{Tr} &\models \psi^-[s] && \text{(Induction)}\\
\mathscr{M}_\mathbf{S} \models \mathbf{T}_i{\vDash}\psi^-[s] \;\&\; \mathscr{M}_\mathbf{S} \cap \mathrm{Tr} &\models \psi^-[s] && \text{(Theorems 31 and 40)}\\
\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} &\models \mathbf{T}_i{\vDash}\psi^-[s] && \text{(Definition of } \mathscr{M}_\mathbf{S} \cap \mathrm{Tr})\\
\mathscr{M}' &\models \mathbf{T}_i{\vDash}\psi^-[s]. && (\mathscr{M}' \text{ and } \mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \text{ agree on } \mathbf{T}_i{\vDash})
\end{aligned}$$

Claim 4: $\mathscr{M} \models S_i$.

Identical to Claim 3 of the proof of Theorem 58.

By Claims 2 and 4, $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} \models \mathbf{S}$. By Lemma 61 it follows that in fact $\mathscr{M}_\mathbf{S} \cap \mathrm{Tr} = \mathscr{M}_\mathbf{S}$, proving the theorem. □

This fulfils the third promise from the introduction: there are true theories $(T_i)_{i\in\omega}$ such that T_i expresses the 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 ($j \preceq i$); that T_i can further be taken to express a weakened version of the truth of T_j (all i, j) is by Lemma 65 part 2.

9 Well-Foundation and Ill-Foundation

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

Definition 68. 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 69. $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}$ is the language of Peano arithmetic extended by a unary predicate $\mathcal{O}$. The following notions are defined by analogy with Section 2:

- For any assignment s and $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}(I)$-formula ϕ with $\mathrm{FV}(\phi)=\{x_1,\ldots,x_n\}$, $\phi^s \equiv \phi(x_1|\overline{s(x_1)})\cdots(x_n|\overline{s(x_n)})$.
- 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 70. 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 71. The *basic axioms of* $\mathcal{O}$ are the following $\mathscr{L}_{\mathrm{PA}}^{\mathcal{O}}$-axioms.

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

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

Theorem 72. 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. $\forall i\in I$, $\forall j\prec i$, $T_i\models\forall x(\mathbf{T}_j\vDash\mathcal{O}(x)\rightarrow\mathcal{O}(x))$.

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 $T_i\models\forall x(\mathbf{T}_j\vDash\mathcal{O}(x)\rightarrow\mathcal{O}(x))$. From these, $T_i\models\forall x(x\in W_{\overline{n}}\rightarrow\mathcal{O}(x))$.

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 $T_i\models\forall x(x\in W_{\overline{n}}\rightarrow\mathcal{O}(x))$, 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 $\forall x(\varphi_{\overline{k}}(x)\downarrow \,\&\, \mathcal{O}(\varphi_{\overline{k}}(x)))\rightarrow\mathcal{O}(\overline{3\cdot 5^k})$, $T_i\models\mathcal{O}(\overline{3\cdot 5^k})$.

To finish the proof, calculate

$$\begin{aligned}
\|T_j\| &= \sup\{|m|+1 : T_j\models\mathcal{O}(\overline{m})\} & \\
&= \sup\{|m| : T_j\models\mathcal{O}(\overline{m})\} & &\text{(Since } T_j \text{ contains } \mathcal{O}(\overline{n})\rightarrow\mathcal{O}(\overline{2^n}) \text{ for all } n\in\mathbb{N})\\
&= \sup\{|m| : m\in W_n\} & &\text{(Since } W_n=\{m\in\mathbb{N} : T_j\models O(\overline{m})\})\\
&= \sup\{|\varphi_k(0)|,|\varphi_k(1)|,\ldots\} & &\text{(By choice of } k)\\
&= |3\cdot 5^k| & &\text{(Definition 68)}\\
&< \sup\{|m|+1 : T_i\models\mathcal{O}(\overline{m})\} & &\text{(Since } T_i\models\mathcal{O}(\overline{3\cdot 5^k}))\\
&= \|T_i\|.
\end{aligned}$$

□

Corollary 73. (Well-Foundedness of True Self-Referential Theories) Let I, $\mathbf{T}$, $\prec$ be as in Theorem 72. 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 73 says that if I, $\mathbf{T}$, $\prec$ are as in Theorem 72 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 72 if we further departed from Kleene and changed line 3 of Definition 68 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 71 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 73 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 74. (Ill-Foundedness of True Self-Referential Theories)

1. For any closed-r.e.-generic $\mathbf{T}^0 = (T^0_i)_{i\in\omega}$, there is an r.e., ill-founded partial order $\prec$ on ω and an $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true, where $\mathbf{T}(n)$ is as in Theorem 18.

2. For any stratifiable-r.e.-generic $\mathbf{T}^0 = (T^0_i)_{i\in\omega}$, there is an r.e., ill-founded partial order $\prec$ on ω and an $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true, where $\mathbf{T}(n)$ is as in Theorem 58.

3. For any stratifiable-true-generic $\mathbf{T}^0 = (T^0_i)_{i\in\omega}$, there is an r.e., ill-founded partial order $\prec$ on ω and an $n \in \mathbb{N}$ such that $\mathbf{T}(n)$ is true, where $\mathbf{T}(n)$ is as in Theorem 67.

Proof. We prove (1), the others are similar. Assume $\neg(1)$. If $\prec$ is any r.e. partial order on ω, then, combining $\neg(1)$ with Theorem 18, $\prec$ is well founded if and only if the conclusion of Theorem 18 holds for $\prec$. Thus it is possible to define WF in $\mathscr{L}_{\mathrm{PA}} \cup \{\mathrm{Tr}\}$. Absurd. $\square$

References

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

[2] Alexander, S. (preprint). A machine that knows its own code. To appear in *Studia Logica*. arXiv: `http://arxiv.org/abs/1305.6080`

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

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

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

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

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