HOKKAIDO UNIVERSITY

Title	Formal Truth as Provability : Proof-Theoretic Studies
Author(s)	林，大智
Degree Grantor	北海道大学
Degree Name	博士(文学)
Dissertation Number	甲第16167号
Issue Date	2024-09-25
DOI	https://doi.org/10.14943/doctoral.k16167
Doc URL	https://hdl.handle.net/2115/95971
Type	doctoral thesis
File Information	Daichi_Hayashi.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Doctoral Dissertation

Formal Truth as Provability: Proof-Theoretic Studies

Daichi Hayashi

Supervisor: Katsuhiko Sano

Graduate School of Humanities and Human Sciences

Hokkaido University

Acknowledgments

Foremost, I would like to express my gratitude to my PhD supervisor, Professor Katsuhiko Sano, for the continuous support and constructive suggestions. In particular, I learned the basics of proof theory, the main method used in this thesis, through his education and discussions with him. I would like to thank Associate Professor Hidenori Kurokawa, Associate Professor Kengo Miyazono, and Professor Tomomi Sato for being the referees of this thesis and for their helpful and insightful comments and suggestions. I am deeply grateful to Associate Professor Graham Emil Leigh for the opportunity of two short research visits in 2019 and 2022. Without his generous support and kind encouragement, our co-authored paper [33] would not have materialised. I would like to thank him for his permission to include the paper in Chapter 4 of this thesis. I must also mention that most of my research, including this thesis, is heavily inspired by and based on his excellent results on axiomatic theory of truth. I would like to thank the members of the Department of Philosophy and Ethics at Hokkaido University. I particularly thank Doctor Takahiro Sawasaki for valuable academic advice, including information on the template of this thesis. Additionally, I would like to thank Associate Professor Shigaku Nogami, a former graduate student at the University of Tokyo, for introducing me to the world of formal theories of truth. Finally, I would like to thank my family –Hidekazu Hayashi, Kazumi Hayashi, Takeo Mizuochi, and Masayuki Mizuochi– for their warm encouragement. This thesis is dedicated to the memory of Yoneko Mizuochi.

Sapporo, Japan — Daichi Hayashi
August 2024

Contents

Chapter 1
Introduction

1.1 Background

1.1.1 Truth as Provability

Tarski's formal definition of truth [61] was the first step towards various mathematical studies of consistent truth definitions and their axiomatisations. As is well known, Tarski's method is to define the truth of a sentence inductively with respect to its logical complexity. In particular, a negated sentence $\neg A$ is defined to be true if and only if A is not true. Over classical logic, this definition means the *principle of bivalence*, i.e., that every (declarative) sentence is either true or false. This principle has often been associated with the concept of truth, and it has been criticised from the viewpoint of anti-realism or verificationism (see, e.g., [16]).

On the other hand, some literature (e.g., [12, 20, 46]) on formal truth gives interpretations of truth as provability in some sense. In general, such an interpretation comprises rules R of the following form:

$$\frac{B_0 \text{ is true} \quad \cdots \quad B_n \text{ is true} \quad \cdots}{A \text{ is true}} \, (R),$$

which says that if the truths of B_0 to B_n (and more) are established, then the truth of A is established. Of course, this does not necessarily mean that the concept of truth characterised by such rules escapes the criticism of anti-realism. Similarly, perhaps such rules can be too complicated for us to learn. Nevertheless, such an understanding of truth as provability can provide an explanation, albeit in an idealised sense, of how we grasp the notion of truth and come to know the truth of a sentence in a bottom-up manner. The general aim of this thesis is therefore to consider to what extent and in what sense various theories of truth, including Tarskian classical truth, can be understood as provability.

1.1.2 Proof-Theoretic Strength of Axiomatic Truth

Just as Tarski's truth definition provides an extension of the truth predicate, set-theoretic definitions of truth are essential for the construction of a consistent notion of truth. On the other hand, if we consider truth to be a simple and basic notion that can be understood without depending on set-theoretic knowledge, then it is also important to study how such truth definitions can be axiomatised as formal systems for the truth predicate $\mathrm{T}(x)$. Therefore, the principal focus of this thesis is not on the set-theoretic interpretations of truth as provability itself, but rather on the investigation of axiomatic theories of truth that have such interpretations.

Furthermore, while axiomatic truth theories can be given for various formal systems that are expressively rich enough to describe their own syntax, this thesis will only consider truth theories over the theory of first-order arithmetic PA (Peano arithmetic) or its equivalent. This is because there is a standard model for PA and our intuition about what arithmetical statements are true works relatively well. It is also because arithmetic is itself one of the main subjects of mathematics, and because various basic techniques have been established for formulating axiomatic truth theories and studying their properties (see, e.g., [27]). In addition, for many consistent theories that contain PA, research on *ordinal analysis*, which aims to express their consistency strength in terms of ordinal numbers (called *proof-theoretic ordinals*), has been developed (for an overview, see [54]).

Let us now give a few more explanations about ordinal analysis. For small ordinal numbers, it is well known that they can be coded by natural numbers, and the basic relations and operations on those ordinals are then expressible as (primitive recursive) predicates and functions on natural numbers. In fact, such a coding can be given in a theory that contains or interprets PA. However, as long as the theory is recursively axiomatisable and consistent, it cannot derive all the important facts about those ordinal numbers because of Gödel's incompleteness theorem. In particular, it is not clear up to how large an ordinal number the theory can derive its well-foundedness. More generally, for a given primitive recursive relation $\prec$, the theory can express its well-foundedness by a suitable axiom or a schema $TI(\prec)$. Then, we can define the *proof-theoretic ordinal* of the theory as the supremum of the order-type of primitive recursive relations such that the theory can derive their well-foundedness. For example, as Gentzen [25] proved, the proof-theoretic ordinal of PA is ε_0, the least epsilon ordinal (see Chapter 2 for further details). Starting from PA, studies have been carried out on various extensions of PA to identify their proof-theoretic ordinals (for more details, see, e.g., [53]).

In the field of axiomatic truth theory, too, identifying the proof-theoretic ordinal of a given truth theory has been one of the most important tasks (cf. [30]). Therefore, for the various axiomatic truth theories introduced in this thesis, our main results are to determine their proof-theoretic ordinals. We also remark that, as we will see in this thesis, the interpretation of truth as provability is, in many cases, useful to determine the upper bound of the proof-theoretic ordinal of a given truth theory.

The ordinal analysis of axiomatic truth theories is also an important study for the elucidation of the notion of truth. First, again by Gödel's incompleteness theorem, there exist arithmetic statements that should be true but cannot be shown from PA. So, if it is possible to derive some such statements in axiomatic truth theory, then the notion of truth is useful for justifying arithmetical statements to some extent. Ordinal analysis is important for this purpose because, in many cases, it also reveals how many new arithmetical statements the theory can derive. Second, the proof-theoretic strength of axiomatic truth theories is also related to more philosophical motivations concerning truth. *Deflationism* on truth can be roughly summarised as a position that does not recognise any substantial content or value in truth, and one formulation of this requires the following of deflationists: axiomatic truth theories should be conservative extensions over the base theory, i.e., new statements of the language of the base theory should not be derivable (cf. [58]). It is therefore an important question for deflationism to elucidate whether a given theory is a conservation extension and, if not, why not.

1.2 Outline of Thesis

Based on the above motivation, we explain what results are achieved in this thesis.

1.2.1 Interpreting Truth as Recursive Provability

Friedman and Sheard [20] formulated nine axiomatic truth theories A to I and showed their consistency model-theoretically. After that, Leigh and Rathjen [50] gave a cut-elimination argument for determining the proof-theoretic strength of each of $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I. The key idea there is that when $\mathrm{T}\ulcorner A \urcorner$ (the truth of A) is proved, we obtain the *recursive* provability of A in some infinite sequent calculus. Here, *recursive* means that the proof of A is regarded as a labelled infinite branching tree, and the labels of each node can be computed by some (partial) recursive function. Since recursive derivations can be identified with the code of the recursive function, they claim that this interpretation can be formalised in a weak extension of PA, thereby providing an upper bound of the proof-theoretic ordinal.

In Chapter 3, we revisit the proof by Leigh and Rathjen [50] and show that it contains a flaw. They claim that the truth of a sentence can be interpreted as the recursive *provability* of the sentence itself, but a mistake here is that they interpret it in terms of mere *provability*. Instead, we prove that proof-theoretic strength can be obtained by giving a *specific procedure* that computes a recursive proof of the sentence. As a result, truth-as-recursive-provability interpretations for $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I succeed, and thus it is shown that the proof-theoretic ordinals given by Leigh and Rathjen are indeed correct. Furthermore, we show that the upper bound for $\mathsf{FS}(= \mathsf{D})$, another one of the Friedman–Sheard systems, can be obtained in a similar manner. However, it is also observed that the truth as recursive provability for FS is more restricted than for $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I because FS is fully in line with the Tarskian truth definition.

1.2.2 Interpreting Truth as Classical Realisability

Whilst a fully Tarskian theory, such as FS, is quite natural as a theory of truth, it is often incompatible with the idea of truth as provability, as we see at the end of Chapter 3. In particular, as we explained above, the axiom for negation states that any sentence is either true or false, but it is clear that this leads to contradictions in many cases, given the liar paradox. The aim of Chapter 4, therefore, is to give a formal theory that relies on realisability as a weakened notion of proof. Computational interpretations of classical theories are given in several studies (e.g., [2, 3, 4, 7, 8, 9, 47, 48, 49]), but in particular, we consider the interpretation of classical realisability given by Krivine [47, 48, 49]. The advantage of Krivine's interpretation is that the notion of arithmetical truth in the standard model can be regarded as a special case of classical realisability, whence Krivine's framework is a generalisation of Tarski's truth. Therefore, the formalisation of classical realisability is expected to be a generalised framework of axiomatic truth theory based on the traditional Tarskian truth. In Chapter 4, as a first step towards a formalisation of classical realisability, we first formulate a Tarskian non-self-referential formal theory CR and consider its proof-theoretic properties. In particular, we prove that CT itself is conservative over PA. We also show that it can be regarded as Tarskian truth theory CT under certain assumptions, and thus our theory can be regarded as a generalisation of axiomatic truth. We further generalise this to a self-referential theory FSR and consider its proof-theoretic properties. Similarly, FSR itself is shown to be conservative over PA.

1.2.3 Interpreting Truth as Frege Structure with Universes

Truths interpreted as provability and their axiomatisation are suited to the goal of justifying the unproved true sentences in a bottom-up way. The question that arises is: how proof-theoretically strong an axiomatic truth theory can be obtained in this way? As one way of strengthening a given truth theory, the idea of iterating truth predicates has been proposed in the literature [23, 37]. This is based on the following intuition. For a base theory B of an object language $\mathcal{L}$, its truth theory S_0 can be constructed using the truth predicate $\mathrm{T}_0(x)$. Now, regarding this new theory S_0 of language $\mathcal{L} \cup \{\mathrm{T}_0\}$ as the base theory, a new truth theory S_1 can be constructed by using a new truth predicate T_1 for it. This process can be repeated a transfinite number of times, yielding a strong truth theory according to the number of times it is iterated. To the author's knowledge, the theory $\mathsf{Aut}(\mathsf{VF})$ by Fujimoto [23] is currently the strongest theory obtained by this method. The aim of Chapter 5 is therefore to give even stronger results based on this approach. To this end, we adopt a framework called the *Frege structure* by Aczel [1]. As a formal system, this can be regarded as an axiomatic truth theory over a base theory of TON, which is a formal theory expressing combinatory logic. The Frege structure is superior to PA as a base theory in that it can easily express set-theoretic concepts. However, as is well known, TON and PA are relatively interpretable to each other, so the difference between the Frege structure and axiomatic truth theory over PA can be understood simply as a way of representation. Corresponding to iterations of truth predicates is the idea of *universes* in Frege structure. Each universe represents a pseudo-truth predicate, and by a principle called the *limit* axiom [14], a stronger universe can be generated from an existing one. The following results are obtained: The extension of KF, the Kripke–Feferman system, by the limit axiom was given by Cantini [14] and Kahle [41]. Then, we give its two extensions. For a system PT, the system based on Aczel's Frege structure, we give the same results as for KF. Finally, we extend VF, the supervaluation-style Frege structure, by the limit axiom. All of these five systems are shown to be proof-theoretically equivalent to each other and stronger than Aut(VF). To the best of the author's knowledge, these systems are the strongest among theories that are obtained in a bottom-up manner.

Chapter 2
Formal Theories of Truth

This chapter introduces famous axiomatic theories of truth that are treated in this thesis. After introducing necessary notation and conventions in the next section, we first define a non-self-referential theory CT that is based on Tarski's well-known truth definition (Section 2.2). Section 2.3 introduces a self-referential approach by Friedman and Sheard [20]. In Section 2.4, we introduce several self-referential theories related to Kripke's truth theory [46].

2.1 Notation and Conventions

We introduce abbreviations for common formal concepts concerning coding and recursive functions. We denote by $\mathcal{L}$ the first-order language of PA. The logical symbols of $\mathcal{L}$ are $\neg, \wedge, \vee, \rightarrow$, $\forall, \exists$, and $=$. We introduce a convention:

- $\forall x A$ means that $\forall x$ and A are strongly connected. Thus, $\forall x A \wedge B$ should be understood as $(\forall x A) \wedge B$.
- In contrast, $\forall x.\ A$ means that $\forall x$ and A are separated. Thus, $\forall x.\ A \wedge B$ means $\forall x(A \wedge B)$.
- Similarly, $A \rightarrow .\ B$ means that A and B are separated.

The non-logical symbols of $\mathcal{L}$ are a constant symbol $\underline{0}$, the constant zero function $Z(x)$, and the successor function $x+1$. The system PA has the following defining axioms of these symbols:

- $Z(x) = \underline{0}$.
- $x + 1 = y + 1 \rightarrow x = y$.
- $x \neq \underline{0} \leftrightarrow \exists y(x = y + 1)$.

In addition, $\mathcal{L}$ has the following function symbols for primitive recursive functions:

Projection P^k_i is a k-ary function symbol of $\mathcal{L}$ for each $1 \leq k, i$.

Composition If g is an m-ary function symbol of $\mathcal{L}$ and $h_1, \dots, h_m$ are k-ary function symbols of $\mathcal{L}$, then $C_{g,h_1,\dots,h_m}$ is a k-ary function symbol of $\mathcal{L}$.

Primitive Recursion If g is a k-ary function symbol of $\mathcal{L}$ and h is a $k+2$-ary function symbol of $\mathcal{L}$, then $R_{g,h}$ is a $k+1$-ary function symbol of $\mathcal{L}$.

The system PA has the following defining axioms of these symbols:

- $P^k_i(x_1, \dots, x_k) = x_i$.
- $C_{g,h_1,\dots,h_m}(x_1, \dots, x_k) = g(h_1(x_1, \dots, x_k), \dots, h_m(x_1, \dots, x_k))$.
- $R_{g,h}(\underline{0}, x_1, \dots, x_k) = g(x_1, \dots, x_k)$.
- $R_{g,h}(y+1, x_1, \dots, x_k) = h(y, R_{g,h}(y, x_1, \dots, x_k), x_1, \dots, x_k)$.

Finally, PA has the schema of mathematical induction:

$$A(\underline{0}) \wedge \forall x(A(x) \rightarrow A(x+1)) \rightarrow .\ \forall x A(x), \text{ for each } \mathcal{L}\text{-formula } A(x).$$

In this thesis, when we are considering some expanded language $\mathcal{L}' \supseteq \mathcal{L}$, we assume that PA formulated for $\mathcal{L}'$ has the schema of mathematical induction for all $\mathcal{L}'$-formulas.

Using the successor function $x+1$, we can define *numerals* $\underline{0}, \underline{1}, \underline{2}, \dots$. Thus, we identify natural numbers n with the corresponding numeral $\underline{n}$.

The primitive recursive pairing function is denoted $\langle \cdot, \cdot \rangle$ with projection functions $(\cdot)_0$ and $(\cdot)_1$ satisfying $(\langle x, y \rangle)_0 = x$ and $(\langle x, y \rangle)_1 = y$. Sequences are treated as iterated pairing: $\langle x_0, x_1, \dots, x_k \rangle := \langle x_0, \langle x_1, \dots, x_k \rangle \rangle$. Based on these constructors, each finite extension of $\mathcal{L}$ is associated a fixed Gödel coding (denoted $\ulcorner e \urcorner$, where e is a finite string of symbols of the extended language) for which the basic syntactic constructions are primitive recursive. In particular, $\mathcal{L}$ has a unary function symbol num expressing the mapping $n \mapsto \ulcorner \underline{n} \urcorner$. In addition, $\mathcal{L}$ contains a binary function symbol $sub(x, y)$ representing the mapping $\ulcorner A(x) \urcorner, n \mapsto \ulcorner A(\underline{n}) \urcorner$ in the case that x is the only free variable of $A(x)$; the binary function symbols $x \dot{=} y$, $x \dot{\rightarrow} y$, and $\dot{\forall} xy$; and a unary function symbol $\dot{\mathrm{T}}x$ representing, respectively, the operations $\ulcorner s \urcorner, \ulcorner t \urcorner \mapsto \ulcorner s = t \urcorner$, $\ulcorner A \urcorner, \ulcorner B \urcorner \mapsto \ulcorner A \rightarrow B \urcorner$, $\ulcorner x \urcorner, \ulcorner A \urcorner \mapsto \ulcorner \forall x A \urcorner$, and $\ulcorner s \urcorner \mapsto \ulcorner \mathrm{T}(s) \urcorner$. These operations will sometimes be omitted, and we write $\ulcorner s = t \urcorner$ and $\ulcorner A \rightarrow B \urcorner$ for $\dot{=}(\ulcorner s \urcorner, \ulcorner t \urcorner)$ and $\dot{\rightarrow}(\ulcorner A \urcorner, \ulcorner B \urcorner)$, etc. The function symbols $x \dot{\wedge} y, x \dot{\vee} y, \dot{\neg} x$, and $\dot{\exists} xy$ are similarly defined.

We introduce a number of abbreviations for $\mathcal{L}$-expressions corresponding to common properties or operations on Gödel codes. The property of being the code of a variable is expressed by the formula $Var(x)$, $ClTerm(x)$ denotes the formula expressing that x is a code of a *closed* $\mathcal{L}$-term, and for a fixed expansion $\mathcal{L}'$ of

$\mathcal{L}$, $Sent_{\mathcal{L}'}(x)$ expresses that x is the code of a sentence. The concepts above are primitive recursively definable, meaning that the representing $\mathcal{L}$-formula is simply an equation between $\mathcal{L}$-terms. Given a formula $A(x)$ with at most x being free, $\ulcorner A(\dot{x})\urcorner$ abbreviates the term $sub(\ulcorner A(x)\urcorner, x)$ expressing the code of the formula $A(\underline{n})$, where n is the value of x. For sequences $\vec{x}$ of variables, $\ulcorner A(\dot{\vec{x}})\urcorner$ is defined similarly.

Quantification over codes is associated with similar abbreviations:

- $\forall\ulcorner A\urcorner \in Sent_{\mathcal{L}'}.\ B(\ulcorner A\urcorner)$ abbreviates $\forall x(Sent_{\mathcal{L}'}(x) \to B(x))$.
- $\forall\ulcorner A\urcorner \in Sent_{\mathcal{L}'}.\ B(\ulcorner \mathrm{T}\ulcorner A\urcorner\urcorner)$ abbreviates $\forall x(Sent_{\mathcal{L}'}(x) \to B(\dot{\mathrm{T}}(num(x))))$.
- $\forall\ulcorner s\urcorner.\ B(\ulcorner s\urcorner)$ abbreviates $\forall x(ClTerm(x) \to B(x))$.
- $\forall\ulcorner A_v\urcorner \in Sent_{\mathcal{L}'}.\ B(v, \ulcorner A\urcorner)$ abbr. $\forall x\forall v(Var(v) \wedge Sent_{\mathcal{L}'}(\dot{\forall} vx) \to B(v,x))$, namely quantification relative to codes of formulas with at most one distinguished variable free.

The set of *partial recursive* functions is defined to be the least class that contains the zero constant, the successor, and the projection functions, and it is closed under composition, primitive recursion, and *minimisation*: if a total function $f : \mathbb{N}^{k+1} \to \mathbb{N}$ is a partial recursive function, then the function $g : \mathbb{N}^k \to \mathbb{N}$ defined by the following is a partial recursive function: $g(x_1, \ldots, x_k) = y$ if and only if y is the least such that $f(x_1, \ldots, x_k, y) = 0$ and for all $y' < y$, $f(x_1, \ldots, x_k, y')$ is defined and is > 0.

Although $\mathcal{L}$ does not have all the symbols for partial recursive functions, they can be expressed in $\mathcal{L}$ via the Kleene 'T predicate' method. The ternary relation $x \cdot y \simeq z$ expresses that the result of evaluating the x-th partial recursive function on input y terminates with output z. Note that this relation has a Σ^0_1 definition in PA as the formula $\exists w(T(x, y, w) \wedge (w)_0 = z)$, where T is some primitive recursive predicate. [1] It will be notationally convenient to use $x \cdot y$ in place of a *term* (with the obvious interpretation), though the use of this abbreviation will be constrained to contexts in which the potential for confusion is minimal.

With the above ternary relation, we can express the property of two closed terms having equal value via a Σ^0_1-formula $Eq(y, z)$. That is, $Eq(y, z)$ expresses that y and z are codes of closed $\mathcal{L}$-terms s and t, respectively, such that $s = t$ is a true equation.

To measure proof-theoretic ordinals, we require an ordinal notation system OT for predicative ordinal numbers. We use several facts about OT (for the proof, see, e.g., [54, Chapter 3]). A formula $x \in OT$ is defined as meaning that x is a

[1] Here, an $\mathcal{L}$-formula A is Σ^0_1 if A is of the form $\exists xB$ for some quantifier-free formula B. Also, a predicate is *primitive recursive*, when its characteristic function is primitive recursive. In $\mathcal{L}$, each primitive recursive predicate can be expressed as some equation.

representation of an ordinal number in OT. Let α, β, and γ range over the ordinal numbers in OT. Thus, $\forall\alpha A(\alpha)$ abbreviates $\forall x(x \in OT \to A(x))$. By the standard method, we can define relations and operations on OT. Let $<$ be the less-than relation, 0 is zero as the ordinal number, $\alpha \in Suc$ means that α is a successor ordinal, $\alpha \in Lim$ says that α is a limit ordinal, $+$ is the ordinal addition, φxy is a binary ordinal function, called the Veblen function, and the binary primitive recursive function $[\alpha]_x$ returns the x-th element of the fundamental sequence for α. For these symbols, the following notations and facts are used in this thesis:

- Let $\omega^\alpha := \varphi 0\alpha$ and $1 := \omega^0$.
- $[\omega^\alpha]_n = \begin{cases} \omega^{\alpha-1} \times n & \text{if } \alpha \in Suc, \\ \omega^{[\alpha]_n} & \text{if } \alpha \in Lim. \end{cases}$
- Let $\omega_n(\alpha) := \begin{cases} \alpha & \text{if } n = 0, \\ \omega^{\omega_{n-1}(\alpha)} & \text{if } n > 0. \end{cases}$
- Let $\varepsilon_\alpha := \varphi 1\alpha$.
- $[\varepsilon_\alpha]_n = \begin{cases} \omega_n(1) & \text{if } \alpha = 0, \\ \omega_n(\varepsilon_{\alpha-1} + 1) & \text{if } \alpha \in Suc, \\ \varepsilon_{[\alpha]_n} & \text{if } \alpha \in Lim. \end{cases}$
- If $\alpha_0 \le \alpha_1$ and $\beta_0 \le \beta_1$, then it holds that $\varphi\alpha_0\beta_0 \le \varphi\alpha_1\beta_1$.
- If $\alpha < \beta$, then it holds that $\varphi\alpha(\varphi\beta\gamma) = \varphi\beta\gamma$. Thus, $\varphi\beta\gamma$ is a fixed point of the function $\varphi\alpha(x)$.
- We sometimes use *natural sum* $x\#y$, a binary function on ordinals such that
  - $\alpha\#\beta = \beta\#\alpha$;
  - if $\alpha_0 < \alpha$, then $\alpha_0\#\beta < \alpha\#\beta$.

The T-free consequences of many axiomatic truth theories can be expressed using some transfinite induction schema. For an $\mathcal{L}$-formula $A(x)$ and an ordinal α, we define the formula $\mathsf{TI}(A, \alpha) = \mathsf{Prog}\lambda x A(x) \to A(\alpha)$, where $\mathsf{Prog}\lambda x A(x) = \forall\alpha(\forall\beta < \alpha(A(\beta)) \to A(\alpha))$. Then, we define the schema: $\mathsf{TI}(\alpha) = \{\mathsf{TI}(A, \alpha) \mid A \text{ is an } \mathcal{L}\text{-formula}\}$. In addition, let $\mathsf{TI}(<\alpha) := \bigcup_{\beta<\alpha} \mathsf{TI}(\beta)$. Finally, let the $\mathcal{L}$-theory $\mathsf{PA} + \mathsf{TI}(<\alpha)$ be the extension of PA with the schema $\mathsf{TI}(<\alpha)$.

2.2 Compositional Truth CT

Tarski's well-known solution to the liar paradox [61] is to say nothing about the truth of sentences containing the truth predicate. In our setting, this means that we only consider the truth of $\mathcal{L}$-sentences, thanks to which the extension of the truth predicate can be constructed inductively:

Definition 1. Let s and t be any closed terms; let A, B be any $\mathcal{L}$-sentences; and let $A(x)$ be any $\mathcal{L}$-formula in which only x is free. The set $\mathbb{T}_{CT} \subseteq \mathbb{N}$ is defined by induction on the logical complexity of sentences:

1. $\ulcorner s = t \urcorner \in \mathbb{T}_{CT}$ iff $s^{\mathbb{N}} = t^{\mathbb{N}}$.
2. $\ulcorner \neg A \urcorner \in \mathbb{T}_{CT}$ iff $\ulcorner A \urcorner \notin \mathbb{T}_{CT}$.
3. $\ulcorner A \wedge B \urcorner \in \mathbb{T}_{CT}$ iff $\ulcorner A \urcorner \in \mathbb{T}_{CT}$ and $\ulcorner B \urcorner \in \mathbb{T}_{CT}$.
4. $\ulcorner A \vee B \urcorner \in \mathbb{T}_{CT}$ iff $\ulcorner A \urcorner \in \mathbb{T}_{CT}$ or $\ulcorner B \urcorner \in \mathbb{T}_{CT}$.
5. $\ulcorner A \to B \urcorner \in \mathbb{T}_{CT}$ iff $\ulcorner A \urcorner \in \mathbb{T}_{CT}$ implies $\ulcorner B \urcorner \in \mathbb{T}_{CT}$.
6. $\ulcorner \forall x A(x) \urcorner \in \mathbb{T}_{CT}$ iff $\ulcorner A(n) \urcorner \in \mathbb{T}_{CT}$ for every $n \in \mathbb{N}$.
7. $\ulcorner \exists x A(x) \urcorner \in \mathbb{T}_{CT}$ iff $\ulcorner A(n) \urcorner \in \mathbb{T}_{CT}$ for some $n \in \mathbb{N}$.

Based on the construction of $\mathbb{T}_{CT}$, the corresponding formal $\mathcal{L}_{\mathrm{T}}$-theory CT is defined such that $\langle \mathbb{N}, \mathbb{T}_{CT} \rangle \models \mathsf{CT}$ holds.

Definition 2. (CT) The $\mathcal{L}_{\mathrm{T}}$-theory CT (compositional truth) consists of PA formulated for the language $\mathcal{L}_{\mathrm{T}}$ plus the following axioms.

($\mathsf{CT}_{=}$) $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(\ulcorner s \urcorner, \ulcorner t \urcorner)$.

($\mathsf{CT}_{\neg}$) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner \neg A \urcorner \leftrightarrow \neg \mathrm{T}\ulcorner A \urcorner$.

($\mathsf{CT}_{\wedge}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner A \wedge B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{CT}_{\vee}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner A \vee B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{CT}_{\to}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner A \to B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{CT}_{\forall}$) $\forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner \forall x A(x) \urcorner \leftrightarrow \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$.

($\mathsf{CT}_{\exists}$) $\forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner \exists x A(x) \urcorner \leftrightarrow \exists v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$.

We mention that CT satisfies Tarski's Convention T [61, pp. 187–188] for formulas in $\mathcal{L}$:

Fact 3. The Tarski-biconditional is derivable in CT for every formula of $\mathcal{L}$. That is, for each formula $A(x_1, \ldots, x_k)$ of $\mathcal{L}$ in which only the distinguished variables occur free, we have

$$\mathsf{CT} \vdash \forall x_1, \ldots, x_k (\mathrm{T}\ulcorner A(\dot{x}_1, \ldots, \dot{x}_k) \urcorner \leftrightarrow A(x_1, \ldots, x_k)).$$

2.3 Friedman–Sheard Truths

Friedman and Sheard's approach to the liar paradox in [20] is to list the natural axioms and rules that the truth predicate should satisfy, and then to identify the (maximally) consistent combinations among them. First, an $\mathcal{L}_{\mathrm{T}}$-theory $\mathsf{Base}_{\mathrm{T}}$ is defined as the common base theory.

Definition 4. Let an $\mathcal{L}_{\mathrm{T}}$-theory PA_{T} be Peano arithmetic with the induction schema for all $\mathcal{L}_{\mathrm{T}}$-sentences, and $\mathsf{Base}_{\mathrm{T}}$ be PA_{T} augmented with the following axioms:

(T-Imp) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner A \to B \urcorner \to \mathrm{T}\ulcorner B \urcorner$,

(T-Val) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{val}(\ulcorner A \urcorner) \to \mathrm{T}\ulcorner A \urcorner$,

(T-PRE) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{Ax}_{\mathsf{PRE}}(\ulcorner A \urcorner) \to \mathrm{T}\ulcorner A \urcorner$,

where $\mathrm{val}(x)$ denotes that x is a logically valid sentence of $\mathcal{L}_{\mathrm{T}}$, and $\mathrm{Ax}_{\mathsf{PRE}}(x)$ is a defined relation symbol denoting that x is a true equation of $\mathcal{L}$.

Over $\mathsf{Base}_{\mathrm{T}}$, the additional principles displayed in Table 2.1 are considered. Note that the axiom schemata (T-In) and (T-Out) allow any $\mathcal{L}_{\mathrm{T}}$-sentence as A. The same applies to the rules (T-Intro) to ($\neg$T-Elim). In addition, the axioms from (T-Rep) to ($\exists$-Inf) are shorthand for universal statements. For instance, the axiom (T-Rep) has precisely the form: $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner$. Finally, the axiom (T-Cons) usually has the following form (cf. [20, 50]):

$$\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \neg(\mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner \neg A \urcorner),$$

which is equivalent to our version over $\mathsf{Base}_{\mathrm{T}}$.

As a result, Friedman and Sheard proved that there are nine maximally consistent systems A to I. Thus, each system would be inconsistent if it were extended by any one of the remaining principles. In other words, any consistent combination is a subtheory of one of A to I.

Definition 5. The systems A to I consist of $\mathsf{Base}_{\mathrm{T}}$ and the following principles, respectively:

A. (T-In), (T-Rep), (T-Del), (T-Comp), ($\forall$-Inf), ($\exists$-Inf), (T-Intro), ($\neg$T-Elim)

B. (T-Rep), (T-Cons), (T-Comp), ($\forall$-Inf), ($\exists$-Inf)

C. (T-Del), (T-Cons), (T-Comp), ($\forall$-Inf), ($\exists$-Inf)

D. (T-Cons), (T-Comp), ($\forall$-Inf), ($\exists$-Inf), (T-Intro), (T-Elim), ($\neg$T-Intro), ($\neg$T-Elim)

Table 2.1: Principles of truth

Name	Axiom Schema
(T-In)	$A \to \mathrm{T}\ulcorner A \urcorner$
(T-Out)	$\mathrm{T}\ulcorner A \urcorner \to A$
Name	Axiom
(T-Rep)	$\mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner$
(T-Del)	$\mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner \to \mathrm{T}\ulcorner A \urcorner$
(T-Cons)	$\neg \mathrm{T}\ulcorner \bot \urcorner$
(T-Comp)	$\mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner \neg A \urcorner$
($\forall$-Inf)	$\forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner \to \mathrm{T}\ulcorner \forall x A(x) \urcorner$
($\exists$-Inf)	$\mathrm{T}\ulcorner \exists x A(x) \urcorner \to \exists v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$
Name	Rule of Inference
(T-Intro)	From A infer $\mathrm{T}\ulcorner A \urcorner$
(T-Elim)	From $\mathrm{T}\ulcorner A \urcorner$ infer A
($\neg$T-Intro)	From $\neg A$ infer $\neg \mathrm{T}\ulcorner A \urcorner$
($\neg$T-Elim)	From $\neg \mathrm{T}\ulcorner A \urcorner$ infer $\neg A$

E. (T-Del), (T-Cons), ($\forall$-Inf), (T-Intro), (T-Elim), ($\neg$T-Intro)

F. (T-Del), ($\forall$-Inf), (T-Intro), (T-Elim), ($\neg$T-Elim)

G. (T-Rep), ($\forall$-Inf), (T-Intro), (T-Elim), ($\neg$T-Elim)

H. (T-Out), (T-Rep), (T-Del), (T-Cons), ($\forall$-Inf), (T-Elim), ($\neg$T-Intro)

I. (T-Rep), (T-Del), ($\forall$-Inf), (T-Elim), ($\neg$T-Elim)

The consistency proof of H is given by interpreting truth as provability in an infinitary Hilbert system (see Section 2.4.3). Cantini [12] determined the proof-theoretic strength of H by interpreting truth as provability in an infinitary sequent calculus (see also Section 5.6 of Chapter 5). A proof-theoretic analysis of D is given by Halbach [28]. The other systems are proof-theoretically analysed by Leigh and Rathjen [50]. In Chapter 3, the upper-bound proofs of D, E, F, G, and I are reconsidered.

2.3.1 Truth as Satisfaction: FS

As Halbach [28] proves, the system D is deductively equivalent to a fully compositional theory, called FS. As Chapter 4 considers a formal theory of classical realisability that is motivated by FS, we present FS here.

Definition 6. (GCT) The $\mathcal{L}_{\mathrm{T}}$-theory GCT (compositional truth) consists of PA formulated for the language $\mathcal{L}_{\mathrm{T}}$ plus the following axioms.

($\mathsf{GCT}_{=}$) $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(\ulcorner s \urcorner, \ulcorner t \urcorner)$.

($\mathsf{GCT}_{\neg}$) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \neg A \urcorner \leftrightarrow \neg \mathrm{T}\ulcorner A \urcorner$.

($\mathsf{GCT}_{\wedge}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \wedge B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{GCT}_{\vee}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \vee B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{GCT}_{\to}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \to B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{GCT}_{\forall}$) $\forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \forall x A(x) \urcorner \leftrightarrow \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$.

($\mathsf{GCT}_{\exists}$) $\forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \exists x A(x) \urcorner \leftrightarrow \exists v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$.

Definition 7. (FS) The $\mathcal{L}_{\mathrm{T}}$-theory FS is obtained by extending GCT by the rules $\mathsf{T\text{-}Intro}$ and $\mathsf{T\text{-}Elim}$:

$$\frac{A}{\mathrm{T}\ulcorner A \urcorner}\ \mathsf{T\text{-}Intro}, \qquad \frac{\mathrm{T}\ulcorner A \urcorner}{A}\ \mathsf{T\text{-}Elim}.$$

Finally, we introduce Friedman and Sheard's model of FS, called the *revision model*. The revision operator $\Gamma : \mathbb{N} \to \mathbb{N}$ is defined as follows:

$$\Gamma(S) := \{\ulcorner A \urcorner \in \mathbb{N} \mid \langle \mathbb{N}, S \rangle \models A\}.$$

We define a series of operators $\Gamma^n(S)$ inductively:

$$\Gamma^0(S) := S,$$
$$\Gamma^{n+1}(S) := \Gamma(\Gamma^n(S)).$$

Then, by induction on the length of derivation in FS, we can prove the following:

Proposition 8. Take any $S \subseteq \mathbb{N}$ and assume $\mathsf{FS} \vdash A$ for an $\mathcal{L}_{\mathrm{T}}$-formula A. Then, there exists a number n such that for all $m \geq n$, it holds that $\langle \mathbb{N}, \Gamma^m(S) \rangle \models A^{\forall}$ where $A^{\forall}$ is a universal closure of A.

The consistency of FS follows by this proposition: we assume $\mathsf{FS} \vdash \bot$; then, there exists a number n such that for all $m \geq n$, it holds that $\langle \mathbb{N}, \Gamma^m(S) \rangle \models \bot$. Since $\bot$ is an $\mathcal{L}$-sentence, we get $\mathbb{N} \models \bot$, a contradiction. Thus, we have $\mathsf{FS} \nvdash \bot$.

2.4 Kripke-Style Truths

In this section, we introduce some well-known semantic definitions of truth and their axiomatisations that are based on three-valued logics. The main feature of these approaches is that, in addition to true/false, a third value (*undetermined*), is allowed. This makes problematic sentences such as λ undetermined, and thus the paradox is avoided. Extensions of these systems are studied in Chapter 5.

2.4.1 Strong Kleene Schema: KF

For a self-referential definition of truth, Kripke [46] proposes to separate the inside from the outside of the truth predicate. The extension of truth is then constructed as (the least) fixed point of the monotone operator based on a three-valued logic, while the outer logic obeys classical logic. In particular, Kripke preferred Strong Kleene three-valued logic, under which a complex sentence is evaluated from its subsentences according to the displayed truth table.

$\neg$	
T	F
U	U
F	T

$\vee$	T	U	F
T	T	T	T
U	T	U	U
F	T	U	F

$\wedge$	T	U	F
T	T	U	F
U	U	U	F
F	F	F	F

$\rightarrow$	T	U	F
T	T	U	F
U	T	U	U
F	T	T	T

In the table, falsity is defined as truth of negation, so a sentence A is false (F) if and only if $\neg A$ is true (T). The evaluation of conjunctive sentences is as follows: a sentence $A \vee B$ is true (T) when either A or B is true; $A \vee B$ is false (F) when both A and B are false; $A \vee B$ is undetermined (U) in the other cases. As for universal sentences, $\forall x A(x)$ is true when $A(n)$ is true for every $n \in \mathbb{N}$; $\forall x A(x)$ is false when $A(n)$ is false for some $n \in \mathbb{N}$; $\forall x A(x)$ is undetermined in the other cases. Note that the disjunction, the conditional, and the existential quantifier are definable: $A \vee B :\equiv \neg(\neg A \wedge \neg B)$; $A \rightarrow B :\equiv \neg A \vee B$; $\exists x A(x) :\equiv \neg \forall x \neg A(x)$. Finally, the evaluation of the truth predicate is very natural: $\mathrm{T}\ulcorner A \urcorner$ is true when A is true; $\mathrm{T}\ulcorner A \urcorner$ is false when A is false; $\mathrm{T}\ulcorner A \urcorner$ is undetermined in the other cases.[2]

To see Kripke's definition of truth in terms of truth-as-provability, let us rephrase the above truth table in the form of an infinitary proof system.

Definition 9. Kripke's truth set $\mathbb{T}_{SK}$ is defined as the set of all codes $\ulcorner A \urcorner$ such that A is derived in the system in Table 2.2.

It is clear that the converse direction of each rule also holds in $\mathbb{T}_{SK}$. For example, if $\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner \in \mathbb{T}_{SK}$, then so does $\ulcorner A \urcorner \in \mathbb{T}_{SK}$. Therefore, we can easily

[2] Many studies (e.g., [30]) also add an axiom saying that $\mathrm{T}(t)$ is false if t does not denote any sentence A. However, it does not increase the proof-theoretic strength of our version.

Table 2.2: Kripke's truth set $\mathbb{T}_{SK}$

$$\frac{s^{\mathbb{N}} = t^{\mathbb{N}}}{s = t}\ (=), \qquad \frac{s^{\mathbb{N}} \neq t^{\mathbb{N}}}{s \neq t}\ (\neq), \qquad \frac{A}{\neg\neg A}\ (\neg\neg),$$

$$\frac{A}{\mathrm{T}\ulcorner A \urcorner}\ (\mathrm{T}), \qquad \frac{\neg A}{\neg \mathrm{T}\ulcorner A \urcorner}\ (\neg\mathrm{T}), \qquad \frac{A_0 \quad A_1}{A_0 \wedge A_1}\ (\wedge),$$

$$\frac{\neg A_i\ (i \leq 1)}{\neg(A_0 \wedge A_1)}\ (\neg\wedge), \qquad \frac{\cdots A(n) \cdots \ (n \in \mathbb{N})}{\forall x A(x)}\ (\forall), \qquad \frac{\neg A(n)}{\neg \forall x A(x)}\ (\neg\forall).$$

The rules for $\vee$, $\rightarrow$, and $\exists$ are similarly defined.

show that the classical model $\langle \mathbb{N}, \mathbb{T}_{SK} \rangle$ satisfies the following $\mathcal{L}_{\mathrm{T}}$-theory KF, which was first introduced by Feferman [18].

Definition 10. (cf. [11, 18]) The theory KF is given by PA and the following axioms: [3]

$\mathsf{K}_{=}$ $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(\ulcorner s \urcorner, \ulcorner t \urcorner)$,

$\mathsf{K}_{\neq}$ $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s \neq t \urcorner \leftrightarrow \neg Eq(\ulcorner s \urcorner, \ulcorner t \urcorner)$,

$\mathsf{K}_{\neg\neg}$ $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \neg\neg A \urcorner \leftrightarrow \mathrm{T}\ulcorner A \urcorner$,

$\mathsf{K}_{\wedge}$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \wedge B \urcorner \leftrightarrow [\mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner B \urcorner]$,

$\mathsf{K}_{\neg\wedge}$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \neg(A \wedge B) \urcorner \leftrightarrow [\mathrm{T}\ulcorner \neg A \urcorner \vee \mathrm{T}\ulcorner \neg B \urcorner]$,

$\mathsf{K}_{\forall}$ $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \forall x A(x) \urcorner \leftrightarrow \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$,

$\mathsf{K}_{\neg\forall}$ $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \neg \forall x A(x) \urcorner \leftrightarrow \exists v \mathrm{T}\ulcorner \neg A(\dot{v}) \urcorner$,

K_{T} $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner \leftrightarrow \mathrm{T}\ulcorner A \urcorner$,

$\mathsf{K}_{\neg\mathrm{T}}$ $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \neg \mathrm{T}\ulcorner A \urcorner \urcorner \leftrightarrow \mathrm{T}\ulcorner \neg A \urcorner$.

[3] The axioms for $\vee, \exists$ are similarly defined. Instead, we could omit and derive these axioms from $\mathsf{K3}$–$\mathsf{K7}$ by defining $(x \mathbin{\underset{\cdot}{\vee}} y)$ and $(\underset{\cdot}{\exists} vx)$ as abbreviations of $\underset{\cdot}{\neg}((\underset{\cdot}{\neg} x) \mathbin{\underset{\cdot}{\wedge}} (\underset{\cdot}{\neg} y))$ and $\underset{\cdot}{\neg}(\underset{\cdot}{\forall} v(\underset{\cdot}{\neg} x))$, respectively.

2.4.2 Aczel–Feferman Schema: DT

In [19], Feferman introduces another Kripke-style theory DT. Motivated by Russell [55], Feferman argues that the truth predicate should be applied only to *determinate* (meaningful) sentences, and that determinateness should be characterised inductively:

- Every equation $s = t$ is determinate.
- $\mathrm{T}\ulcorner A \urcorner$ is determinate iff A is determinate.
- $\neg A$ is determinate iff A is determinate.
- For $\bullet \in \{\wedge, \vee\}$, $A \bullet B$ is determinate iff both A and B are determinate.
- For $Q \in \{\forall, \exists\}$, $QxA(x)$ is determinate iff $A(n)$ is determinate for all $n \in \mathbb{N}$.
- $A \to B$ is determinate iff A is determinate and truth of A implies determinateness of B.

In the above, only the conditional has a special definition, for the determinateness of $A \to B$ is not defined simply as that of $\neg A \vee B$. This is necessary to be able to validate natural universal statements, such as 'Every $\mathcal{L}$-consequence of PA is true' [19, p. 208].

In Feferman's definition, the predicate $\mathrm{D}(x)$ for determinateness is definable as $\mathrm{T}(x) \vee \mathrm{T}(\underset{\cdot}{\neg} x)$, which means that x is determined to be true or false. Then, the above definition is formalisable by using the predicate $\mathrm{D}(x)$. Furthermore, by adding the compositional axioms restricted to determinate sentences, an $\mathcal{L}_{\mathrm{T}}$-theory DT is obtained:

Definition 11. Let $\mathrm{D}(x) :\equiv \mathrm{T}(x) \vee \mathrm{T}(\underset{\cdot}{\neg} x)$. The theory DT is given by PA augmented with the following axioms:

$\mathsf{D}_=$ $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{D}\ulcorner s = t \urcorner$,

$\mathsf{D}_\neg$ $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner \neg A \urcorner \leftrightarrow \mathrm{D}\ulcorner A \urcorner$,

$\mathsf{D}_\wedge$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner A \wedge B \urcorner \leftrightarrow [\mathrm{D}\ulcorner A \urcorner \wedge \mathrm{D}\ulcorner B \urcorner]$,

$\mathsf{D}_\to$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner A \to B \urcorner \leftrightarrow [\mathrm{D}\ulcorner A \urcorner \wedge \{\mathrm{T}\ulcorner A \urcorner \to \mathrm{D}\ulcorner B \urcorner\}]$,

$\mathsf{D}_\forall$ $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner \forall x A(x) \urcorner \leftrightarrow \forall v \mathrm{D}\ulcorner A(\dot{v}) \urcorner$,

D_T $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner \leftrightarrow \mathrm{D}\ulcorner A \urcorner$,

$\mathsf{DT}_=$ $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(\ulcorner s \urcorner, \ulcorner t \urcorner)$,

$\mathsf{DT}_\neg$ $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner \neg A \urcorner \to .\ \mathrm{T}\ulcorner \neg A \urcorner \leftrightarrow \neg \mathrm{T}\ulcorner A \urcorner$,

$\mathsf{DT}_\wedge$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner A \wedge B \urcorner \to .\ \mathrm{T}\ulcorner A \wedge B \urcorner \leftrightarrow [\mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner B \urcorner]$,

$\mathsf{DT}_\to$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner A \to B \urcorner \to .\ \mathrm{T}\ulcorner A \to B \urcorner \leftrightarrow [\mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner B \urcorner]$,

$\mathsf{DT}_\forall$ $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner \forall x A(x) \urcorner \to .\ \mathrm{T}\ulcorner \forall x A(x) \urcorner \leftrightarrow \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$,

DT_T $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{D}\ulcorner A \urcorner \to .\ \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner \leftrightarrow \mathrm{T}\ulcorner A \urcorner$.

The axioms for $\vee$ and $\exists$ are similar.

As Feferman requires, the Tarski-schema restricted to determinate sentences is indeed derivable:

Corollary 12. ([19, Theorem 1]) In DT, the following is derivable for every $\mathcal{L}_\mathrm{T}$-sentence A:

$$\mathrm{D}\ulcorner A \urcorner \to .\ \mathrm{T}\ulcorner A \urcorner \leftrightarrow A.$$

Proof. The proof is by induction on A. We only consider the conditional case $A :\equiv B \to C$. Thus, we assume $\mathrm{D}\ulcorner B \to C \urcorner$. Then, by the axiom ($\mathsf{D}_\to$), we have $\mathrm{D}\ulcorner B \urcorner$ and $\mathrm{T}\ulcorner B \urcorner \to \mathrm{D}\ulcorner C \urcorner$. Thus, by the induction hypothesis, we get $\mathrm{T}\ulcorner B \urcorner \leftrightarrow B$ and $\mathrm{T}\ulcorner B \urcorner \to .\ \mathrm{T}\ulcorner C \urcorner \leftrightarrow C$. Combining them with the axiom ($\mathsf{DT}_\to$) yields $\mathrm{T}\ulcorner B \to C \urcorner \leftrightarrow .\ B \to C$, as required. ■

Just as KF is based on Strong Kleene logic, DT can be understood as based on *Aczel–Feferman* logic, a variant of Weak Kleene logic that has the truth table below. Thus, similar to $\mathbb{T}_{SK}$, an extension $\mathbb{T}_{AF}$ of the truth predicate of DT can be constructed.

$\neg$	
T	F
U	U
F	T

$\vee$	T	U	F
T	T	U	T
U	U	U	U
F	T	U	F

$\wedge$	T	U	F
T	T	U	F
U	U	U	U
F	F	U	F

$\to$	T	U	F
T	T	U	F
U	U	U	U
F	T	T	T

Finally, it should be noted that before [19], Aczel [1] defines a (semantic) system similar to DT. The axiomatisation of Aczel's system is discussed in Section 5.4 of Chapter 5.

2.4.3 Supervaluation Schema: VF

Kripke [46, p. 711] also suggest a truth theory based on van Fraassen's supervaluation interpretation [65]. This idea is comprehensively investigated by Cantini [12], who provides proof theory and model theory of an axiomatisation VF. As Cantini remarks [12, p. 258], Cantini's VF is an $\mathcal{L}$-conservative extension of H, one of the Friedman–Sheard systems (Definition 5). For simplicity, we define VF just as H here.

Definition 13. The $\mathcal{L}_{\mathrm{T}}$-theory VF consists of $\mathsf{Base}_{\mathrm{T}}$ and the following axioms:

($\mathsf{T\text{-}Out}$) $\mathrm{T}\ulcorner A \urcorner \to A$;

(Rep) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner$;

(Del) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner \to \mathrm{T}\ulcorner A \urcorner$;

($\mathsf{T\text{-}Cons}$) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \neg(\mathrm{T}\ulcorner A \urcorner \wedge \mathrm{T}\ulcorner \neg A \urcorner)$;

($\forall\mathsf{\text{-}Inf}$) $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \forall n \mathrm{T}\ulcorner A(\dot{n}) \urcorner \to \mathrm{T}\ulcorner \forall x A(x) \urcorner$.

Note that ($\mathsf{T\text{-}Elim}$) and ($\neg\mathsf{T\text{-}Intro}$) are derivable from VF, so they are omitted.

As a model for $\mathsf{VF}(= \mathsf{H})$, Friedman and Sheard [20, p. 13] interpret the truth predicate as provability in an infinitary Hilbert system. Let Th_∞ be PA closed under the rule ($\mathsf{T\text{-}Intro}$) and the omega rule (if $A(n)$ is derived for every $n \in \mathbb{N}$, then $\forall x A(x)$ is derivable). We let $\mathsf{Th}_\infty \vdash A$ mean that A is derivable in Th_∞. Then, we can prove the following:

Proposition 14. ([20, p. 13]) 1. Th_∞ is consistent.

2. If $\mathsf{Th}_\infty \vdash A$, then $\langle \mathbb{N}, \mathsf{Th}_\infty \rangle \models A$.

From this proposition, we easily obtain $\langle \mathbb{N}, \mathsf{Th}_\infty \rangle \models \mathsf{VF}$. For example, the axiom ($\mathsf{T\text{-}Cons}$) intuitively states that Th_∞ does not derive a contradiction, which is exactly what item 1 of the above proposition says. Thus, the truth predicate of VF can be understood as the provability in Th_∞.

To determinate the proof-theoretic strength of a system, as is defined for VF in [12, Section 4], it is often more useful to give a truth-as-provability interpretation in the form of a sequent calculus than a Hilbert system. Sequent calculi for various truth theories are given in Chapter 3. Moreover, an extension of Cantini's sequent calculus for VF as a Frege structure is presented in Section 5.6 of Chapter 5.

Chapter 3

Truth as Recursive Provability for Friedman–Sheard Theories

In Chapter 2, we introduced various consistent truth theories. For the proof-theoretic purpose, however, it is important to know how strong a metatheory is required to provide such a consistency proof. In particular, Leigh and Rathjen [50] interpret truth as *recursive* provability in an infinitary sequent calculus and give cut elimination arguments, in order to determine the proof-theoretic ordinal of the Friedman–Sheard systems $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I. In Section 3.1, we explain their proof in detail and point out a flaw therein. In Section 3.2, we modify their proof and verify that their proof-theoretic ordinals are indeed correct. In Section 3.3, we show that it is possible to give a truth-as-provability interpretation to FS in a restricted sense.

The content of Section 3.3 is based on [31].

3.1 Ordinal Analysis by Cut Elimination

The ordinal analysis for some Friedman–Sheard systems in [50] is based on the cut elimination argument, which is an extension of that for PA. Therefore, in Section 3.1.1, in order to explain Leigh and Rathjen's proof, we will first introduce the well-known cut elimination argument for PA. In Section 3.1.2, we examine in detail the proof by Leigh and Rathjen and observe that this is insufficient to give the proof-theoretic upper bound.

3.1.1 Cut Elimination for PA

Following the standard literature (cf. [50, 54, 57]), we give a cut elimination argument for PA. For that purpose, we first define an infinitary proof system PA^∞ in which PA can be embedded. The system PA^∞ is formulated as an ordinary Tait calculus, and hence, each formula is identified with its negation normal form. For any $\mathcal{L}_{\mathrm{T}}$-formula A, let $|A|$ be the number of the logical symbols except $\neg$ occurring in A. In particular, $|\neg\neg A| = |\neg A| = |A|$ holds for every formula A. An ($\mathcal{L}_{\mathrm{T}}$)-*sequent* Γ is a finite set of $\mathcal{L}_{\mathrm{T}}$-sentences. [1] For the sequents Γ, Δ and a sentence A, we let Γ, Δ and Γ, A represent the sequents $\Gamma \cup \Delta$ and $\Gamma \cup \{A\}$, respectively.

Definition 15. (System PA^∞) By induction on $\alpha \in \mathrm{OT}$ and $k \in \mathbb{N}$, we define a derivation relation $(\mathrm{PA}^\infty) \vdash^{\alpha}_{k} \Gamma$ for an $\mathcal{L}_{\mathrm{T}}$-sequent Γ. The relation $\vdash^{\alpha}_{k} \Gamma$ informally means that the sequent Γ is derived with the length α and with the cut-rank k. The system PA^∞ comprises Basic axioms and Basic rules displayed in Table 3.1. Note that in each of the axioms and rules, the displayed sentences in the conclusion are called the *principal* sentences of the axiom or rule. The other part Γ is called the *context* of the axiom or rule.

Table 3.1: System PA^∞

Basic axioms

($\mathsf{Ax.1}$) $\vdash^{\alpha}_{k} \Gamma, A$, if A is a true $\mathcal{L}$-literal.
($\mathsf{Ax.2}$) $\vdash^{\alpha}_{k} \Gamma, \neg\mathrm{T}s, \mathrm{T}t$, if $s^{\mathbb{N}} = t^{\mathbb{N}}$.

Basic rules (assume: $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, A_0, A_1}{\vdash^{\alpha}_{k} \Gamma, A_0 \vee A_1}\ (\vee) \qquad \frac{\vdash^{\alpha_0}_{k} \Gamma, A_0 \quad \vdash^{\alpha_1}_{k} \Gamma, A_1}{\vdash^{\alpha}_{k} \Gamma, A_0 \wedge A_1}\ (\wedge)$$

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, A(i)}{\vdash^{\alpha}_{k} \Gamma, \exists x A}\ (\exists) \qquad \frac{\dots \vdash^{\alpha_i}_{k} \Gamma, A(i) \dots \ (i \in \mathbb{N})}{\vdash^{\alpha}_{k} \Gamma, \forall x A}\ (\forall)$$

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, A \quad \vdash^{\alpha_1}_{k} \Gamma, \neg A}{\vdash^{\alpha}_{k} \Gamma}\ (\mathsf{cut}), \text{ where } |A| < k.$$

The following lemma is proved almost in the same way as the standard sequent calculus LK of classical first-order logic.

[1] As all of the theories considered in the later subsections are formulated in $\mathcal{L}_{\mathcal{L}_{\mathrm{T}}}$, we consider sequents with the truth predicate here.

Lemma 16. The following basic properties hold in PA^∞.

1. (Weakening) If $\vdash^{\alpha}_{k} \Gamma$, then $\vdash^{\beta}_{l} \Delta$ for any $\alpha \leq \beta$, $\Gamma \subseteq \Delta$, and $k \leq l$.
2. ($\wedge$-Inversion) If $\vdash^{\alpha}_{k} \Gamma, A_0 \wedge A_1$, then $\vdash^{\alpha}_{k} \Gamma, A_i$ for every $i \leq 1$.
3. ($\forall$-Inversion) If $\vdash^{\alpha}_{k} \Gamma, \forall x A(x)$, then $\vdash^{\alpha}_{k} \Gamma, A(\bar{n})$ for every $n \in \mathbb{N}$.
4. ($\vee$-exportation) If $\vdash^{\alpha}_{k} \Gamma, A \vee B$, then $\vdash^{\alpha}_{k} \Gamma, A, B$.
5. (False Literal) If $\vdash^{\alpha}_{k} \Gamma, A$ for a false $\mathcal{L}$-literal A, then $\vdash^{\alpha}_{k} \Gamma$.
6. (Substitution) If $\vdash^{\alpha}_{k} \Gamma, A(s)$ and $s^{\mathbb{N}} = t^{\mathbb{N}}$, then $\vdash^{\alpha}_{k} \Gamma, A(t)$.

For an $\mathcal{L}_{\mathrm{T}}$-formula A, let $A^\forall$ be its universal closure.

Lemma 17. (Embedding Lemma) Let A be any $\mathcal{L}_{\mathrm{T}}$-formula. If $\mathsf{PA} \vdash A$, then $\mathsf{PA}^\infty \vdash^{\alpha}_{k} A^\forall$ for some $k \in \mathbb{N}$ and $\alpha < \omega + \omega$.

Proof. The proof is by induction on the length of the derivation of A in PA. We can prove the claim essentially in the same way as the translation of the Hilbert system of classical first-order logic into the sequent calculus LK.

As the base case, we consider the induction axiom $[A(0) \wedge \forall x(A(x) \to A(x+1))] \to \forall x A(x)$. First, by induction on $|A|$, we can derive the law of excluded middle (LEM) with a finite length: $\vdash^{<\omega}_{0} A, \neg A$. To derive the induction axiom, we deduce $\neg A(0), \neg\forall x(A(x) \to A(x+1)), A(n)$ by induction on n. The case $n = 0$ is obtained by (LEM) and Weakening. The inductive step is derived as follows:

$$\dfrac{\dfrac{\dfrac{\text{ind. hyp.}}{\vdash^{<\omega}_{0} \neg A(0), \neg\forall x(A(x) \to A(x+1)), A(n)} \quad \dfrac{\mathsf{LEM}}{\vdash^{<\omega}_{0} \neg A(n+1), A(n+1)}}{\vdash^{<\omega}_{0} \neg A(0), \neg\forall x(A(x) \to A(x+1)), A(n) \wedge \neg A(n+1), A(n+1)}\,(\wedge)}{\vdash^{<\omega}_{0} \neg A(0), \neg\forall x(A(x) \to A(x+1)), A(n+1)}\,(\exists)$$

Here, we used Weakening implicitly. Finally, the induction axiom is inferred by ($\forall$):

$$\dfrac{\dfrac{\cdots \vdash^{<\omega}_{0} \neg A(0), \neg\forall x(A(x) \to A(x+1)), A(n) \cdots}{\vdash^{\omega}_{0} \neg A(0), \neg\forall x(A(x) \to A(x+1)), \forall x A(x)}\,(\forall)}{\vdash^{\omega+2}_{0} [A(0) \wedge \forall x(A(x) \to A(x+1))] \to \forall x A(x)}\,(\vee)^{\times 2}$$

The other axioms and the inductive step are similarly treated. ■

For $\alpha, \beta \in \mathrm{OT}$, let $\alpha \# \beta$ be the *natural sum* of α and β.

Lemma 18. (Reduction Lemma) In PA^{∞}, the following holds:

$$\text{If } \vdash^{\alpha}_{k} \Gamma, A \text{ and } \vdash^{\beta}_{k} \Delta, \neg A \text{ hold with } |A| = k, \text{ then } \vdash^{\alpha\#\beta}_{k} \Gamma, \Delta.$$

Proof. Again, the proof is similar to the standard cut elimination proof in LK (e.g., [57, p. 874]). We prove the claim by induction on $\alpha\#\beta$. We divide the cases by whether A and $\neg A$ are active in the last inference of the derivation, respectively.

Case I: either A or $\neg A$ is not principal. By symmetry, we may assume that A is not active. For example, we consider the case where Γ, A is derived by a rule R from the unique premise Γ', A:

$$\frac{\vdash^{\alpha_0}_{k_0} \Gamma', A}{\vdash^{\alpha}_{k} \Gamma, A}\ (\mathrm{R}) .$$

Since $\alpha_0\#\beta < \alpha\#\beta$, we can apply the induction hypothesis to $\vdash^{\alpha_0}_{k_0} \Gamma', A$ and $\vdash^{\beta}_{k} \Delta, \neg A$. Thus, we obtain $\vdash^{\alpha_0\#\beta}_{k} \Gamma', \Delta$. Therefore, the conclusion $\vdash^{\alpha\#\beta}_{k} \Gamma, \Delta$ follows by (R) and Weakening.

Case II: both A and $\neg A$ are principal.

Case IIa $|A| > 0$: For example, we consider the case $A \equiv \forall x B(x)$ and the following derivations:

$$\frac{\cdots \vdash^{\alpha_i}_{k} \Gamma, B(i) \cdots (i \in \mathbb{N})}{\vdash^{\alpha}_{k} \Gamma, \forall x B(x)}\ (\forall) , \qquad \frac{\vdash^{\beta_n}_{k} \Delta, \neg\forall x B(x), \neg B(n)}{\vdash^{\beta}_{k} \Delta, \neg\forall x B(x)}\ (\exists) .$$

By the induction hypothesis, we deduce $\vdash^{\alpha\#\beta_n}_{k} \Gamma, \Delta, \neg B(n)$. Since $|B(n)| < |\forall x B(x)| = k$, we can apply (cut) to $\vdash^{\alpha\#\beta_n}_{k} \Gamma, \Delta, \neg B(n)$ and $\vdash^{\alpha_n}_{k} \Gamma, B(n)$. Thus, it follows that $\vdash^{\max(\alpha_n, \alpha\#\beta_n)+1}_{k} \Gamma, \Delta$. As $\max(\alpha_n, \alpha\#\beta_n) + 1 \leq \alpha\#\beta$, the conclusion $\vdash^{\alpha\#\beta}_{k} \Gamma, \Delta$ is obtained by Weakening.

Case IIb $|A| = 0$: If A is a literal of $\mathcal{L}$, then the conclusion is clear by False Literal and Weakening of Lemma 16. Thus, we may assume $A \equiv \mathrm{T}(s)$. Then, by the definition of PA^{∞}, $\neg A \equiv \neg\mathrm{T}(s)$ can only be active in the axiom $(\mathsf{Ax}.2)$, hence $\Delta, \neg\mathrm{T}(s)$ is an instance of $(\mathsf{Ax}.2)$. This means that $\mathrm{T}(t) \in \Delta$ for some $t^{\mathbb{N}} = s^{\mathbb{N}}$. Therefore, the conclusion $\vdash^{\alpha\#\beta}_{k} \Gamma, \Delta$ is derivable from $\vdash^{\alpha}_{k} \Gamma, A$ by Substitution and Weakening of Lemma 16.

■

From Reduction Lemma, we obtain cut elimination for PA^{∞}.

Lemma 19. (Cut Elimination) In PA^∞, the following holds:

$$\text{If} \mid\!\frac{\alpha}{k+1}\, \Gamma, \text{ then} \mid\!\frac{\omega^\alpha}{k}\, \Gamma.$$

Lemma 20. (Consistency of PA) Every $\mathcal{L}$-theorem of PA is true in $\mathbb{N}$.

Proof. Taking any $\mathcal{L}$-sentence A, we assume $\mathsf{PA} \vdash A$. Then, by Embedding Lemma (Lemma 17), we have $\mathsf{PA}^\infty \mid\!\frac{\alpha}{k}\, A$ for some $k \in \mathbb{N}$ and $\alpha < \omega + \omega$. Thus, Cut Elimination (Lemma 19) yields $\mathsf{PA}^\infty \mid\!\frac{\omega_k(\alpha)}{0}\, A$. By transfinite induction up to α, we can easily show that if $\mid\!\frac{\alpha}{0}\, \Gamma$, then Γ is not empty and at least one of Γ is true in $\mathbb{N}$. Therefore, it follows that A is true in $\mathbb{N}$. ■

While the consistency of PA follows more directly from the definition of $\mathbb{N}$, the above cut elimination argument is more suitable for formalisation in a weak arithmetic, such as PA. For that purpose, we consider *recursive* derivations, for PA is not expressively rich enough to express arbitrary derivations of PA^∞.

Definition 21. (PA^∞_{rec}) We define a *code* d of PA^∞_{rec} inductively.

($\mathsf{Ax}.1$) If A is a true $\mathcal{L}$-literal, then we say that the sequence $d = \langle \ulcorner \mathsf{Ax}.1 \urcorner, \alpha, k, \Gamma, \{A\}\rangle$ is a *code* of PA^∞_{rec}, where $\ulcorner \mathsf{Ax}.1 \urcorner$ is the name of the rule ($\mathsf{Ax}.1$) (called the *main rule* of d); α is any ordinal number (called the *length* of d); k is any natural number (called the *cut rank* of d); Γ is any sequent (called the *context* of d); $\{A\}$ is a single sequent (A is called the *principal* sentence of d). In addition, the sequent Γ, A is called the *end sequent* of d.

($\mathsf{Ax}.2$) If $s^{\mathbb{N}} = t^{\mathbb{N}}$, then the sequence $d = \langle \ulcorner \mathsf{Ax}.2 \urcorner, \alpha, k, \Gamma, \{\neg \mathrm{T}s, \mathrm{T}t\}\rangle$ is a code of PA^∞_{rec}. Note that the principal sentences of d are $\neg \mathrm{T}s$ and $\mathrm{T}t$. Also, the end sequent of d is $\Gamma, \neg \mathrm{T}s, \mathrm{T}t$.

($\wedge$) Let d_0 and d_1 be codes of PA^∞_{rec}. Then, the sequence $d = \langle \ulcorner \wedge \urcorner, \alpha, k, \Gamma, \{A_0 \wedge A_1\}, d_0, d_1\rangle$ is a code of PA^∞_{rec}, if the following hold:

- The length of d_0 and d_1 are less than α.
- The cut rank of d_0 and d_1 is k.
- The end sequent of d_0 is Γ, A_0.
- The end sequent of d_1 is Γ, A_1.

Note that the principal sentence of d is $A_0 \wedge A_1$. We say that d_0 and d_1 are the *subderivations* of d.

($\forall$) Let e be the code of a recursive function that enumerates codes $d_0, d_1, \cdots, d_n, \cdots$ of PA^∞_{rec}. Then, the sequence $d = \langle \ulcorner \forall \urcorner, \alpha, k, \Gamma, \{\forall x A(x)\}, e \rangle$ is a code of PA^∞_{rec}, if the following hold for every $n \in \mathbb{N}$:

- The length of d_n is less than α.
- The cut rank of d_n is k.
- The end sequent of d_n is $\Gamma, A(n)$.

Here, the *subderivations* of d are $d_0, d_1, \cdots, d_n, \cdots$.

The definitions for the other rules ($\vee$), ($\exists$), and (cut) are similar.

By the above definition, each code d of PA^∞_{rec} can be seen as expressing some derivation of PA^∞. Moreover, the premises of each omega rule are recursively enumerated; thus, they are expressible as (the code of) a recursive function. Thanks to this requirement, the above definition is formalisable in PA. In particular, we can define an $\mathcal{L}$-formula $d \,{\mid}\!\frac{\alpha}{k}\, \Gamma$ expressing that d is a code of PA^∞_{rec} such that the length is α, the cut rank is k, and the end sequent is Γ. We mean by $\mathsf{PA}^\infty_{rec} \,{\mid}\!\frac{\alpha}{k}\, \Gamma$ that there exists a code d that satisfies $d \,{\mid}\!\frac{\alpha}{k}\, \Gamma$.

Under the above preparation on recursive derivations, the formalisation of the consistency proof proceeds as follows:

1. Taking any $\mathcal{L}$-sentence A, we assume $\mathsf{PA} \vdash A$. We take the code d of this finite derivation. Thus, $Bew_{\mathsf{PA}}(d, \ulcorner A \urcorner)$ is true in $\mathbb{N}$.
2. A close look at the proof of Embedding Lemma (Lemma 17) shows that there exists a (partial) recursive function $F_{embd}(x)$ such that $F_{embd}(d) \,{\mid}\!\frac{\alpha}{k}\, A$ is true, where the length α and the cut rank k are computable from $F_{embd}(d)$.
3. Similarly, there exists a (partial) recursive function $F_{cut}(x)$ that satisfies:

$$F_{cut}(F_{embd}(d)) \,{\mid}\!\frac{\omega_k(\alpha)}{0}\, A.$$

4. It is well known that PA can, for each $k \in \mathbb{N}$, define a partial truth predicate $T_k(x)$, meaning that x is (the code of) a true $\mathcal{L}$-sentence whose logical complexity is less than k. Now, we can formalise the consistency proof of PA (Lemma 20) as follows. We take any ordinal number β and a natural number k. Then, the following is obtained:

$$\mathsf{PA} + \mathsf{TI}(\beta) \vdash \forall \beta_0 < \beta. \forall \ulcorner \Gamma \urcorner \in Seq^k_{\mathcal{L}}.\ (\mathsf{PA}^\infty_{rec} \,{\mid}\!\frac{\beta}{0}\, \Gamma) \to \exists \ulcorner B \in \Gamma \urcorner.\ T_k(\ulcorner B \urcorner),$$

where $Seq^k_{\mathcal{L}}$ denotes $\mathcal{L}$-sequents such that the logical complexity of each member is less than k; $\exists \ulcorner B \in \Gamma \urcorner$ means that there exists a sentence B of Γ.

5. Since $\omega_k(\alpha) < \varepsilon_0$ and transfinite induction up to ε_0 is available in PA, we have $\mathsf{PA} \vdash A$.

Note that for the above proof to be successful, each derivation of PA has to be transformed into that of $\mathsf{PA}^{\infty}_{rec}$ in an effective (computable) way. In fact, $F_{cut}(F_{embd}(x))$ is such a (partial) recursive function. In the next subsection 3.1.2, we will see that this is a serious problem in Leigh and Rathjen's upper-bound proofs of the Friedman–Sheard theories. A solution will then be presented.

3.1.2 Leigh and Rathjen's Analysis for Friedman–Sheard Systems

In [50], Leigh and Rathjen give cut elimination arguments to obtain upper bounds of the systems $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I of the Friedman–Sheard theories. For $\mathsf{S} \in \{\mathsf{E}, \mathsf{F}, \mathsf{G}, \mathsf{I}\}$, the proof generally proceeds as follows. Assume $\mathsf{S} \vdash A$ for an $\mathcal{L}$-sentence A. Then, similar to for PA in Section 3.1.1, the derivation is embedded into a corresponding sequent calculus S^{∞}; thus, we have $\mathsf{S}^{\infty} \vdash A$. Therefore, to show the soundness of S, it suffices to verify the soundness of S^{∞}. Finally, by running the argument in a recursive system $\mathsf{S}^{\infty}_{rec}$, the soundness proof is formalisable within a suitable extension T of Peano arithmetic. As a result, we obtain $\mathsf{T} \vdash A$; thus, T gives an upper bound of S. Leigh and Rathjen do not provide a complete explanation about how to conduct the proof in $\mathsf{S}^{\infty}_{rec}$, but that is exactly where the problem lies. In the present subsection, we take G as an example, in order to observe the problem in detail.

Here, let us repeat the definition of G.

Definition 22. (System G [50]) An $\mathcal{L}_{\mathrm{T}}$-theory G consists of $\mathsf{Base}_{\mathrm{T}}$ with the following axioms and rules:

(Rep) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner$;

($\mathrm{T}_{\forall}$) $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner \to \mathrm{T}\ulcorner \forall x A(x) \urcorner$;

(T-Intro) from A infer $\mathrm{T}\ulcorner A \urcorner$;

(T-Elim) from $\mathrm{T}\ulcorner A \urcorner$ infer A;

($\neg$T-Elim) from $\neg \mathrm{T}\ulcorner A \urcorner$ infer $\neg A$.

To obtain the upper bound of G, [50] embeds G into the corresponding infinitary system G^{∞}, in which the cut elimination theorem holds. Instead of explicitly having the rule T-Elim in G^{∞}, [50] proved the admissibility of T-Elim. To put it more precisely, they showed a kind of *soundness*: when a T-positive sequent Γ is derived, then Γ is satisfied under the interpretation of $\mathrm{T}\ulcorner A \urcorner$ as the derivability of A in G^{∞}. In particular, if $\mathrm{T}\ulcorner A \urcorner$ is derived, then A is also derivable by the soundness. This means the admissibility of T-Elim.

Definition 23. (System G^{∞} (cf. [50, Definition 3.16])) The derivation system G^{∞} consists of the basic axioms, the basic rules, and the truth rules in Table 3.2.

Table 3.2: System G^{∞}

Truth rules (assume: $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$$\frac{\vdash^{\alpha_0}_{k} A}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{T\text{-}Intro}) \qquad \frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner A \urcorner \quad \vdash^{\alpha_1}_{k} \Gamma, \mathrm{T}\ulcorner A \to B \urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner B \urcorner}\ (\mathsf{T}_{\to})$$

$$\frac{\ldots \vdash^{\alpha_i}_{k} \Gamma, \mathrm{T}\ulcorner A(i) \urcorner \ldots \ (i \in \mathbb{N})}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \forall x A \urcorner}\ (\mathsf{T}_{\forall}) \qquad \frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner A \urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner}\ (\mathsf{Rep})$$

Similarly to $\mathsf{PA}^{\infty}_{rec}$ (Definition 21), the recursive system $\mathsf{G}^{\infty}_{rec}$ is defined to be G^{∞}, in which only recursive derivations are allowed. When we do not care whether it is G^{∞} or $\mathsf{G}^{\infty}_{rec}$, we shall write $\mathsf{G}^{\infty}_{(rec)}$.

Basic properties and cut elimination for G^{∞} (and also for $\mathsf{G}^{\infty}_{rec}$) are shown by [50]:

Lemma 24. The following basic properties hold in $\mathrm{G}^{\infty}_{(rec)}$.

1. (Weakening) If $\vdash^{\alpha}_{k} \Gamma$, then $\vdash^{\beta}_{l} \Delta$ for any $\alpha \leq \beta$, $\Gamma \subseteq \Delta$, and $k \leq l$.
2. ($\wedge$-Inversion) If $\vdash^{\alpha}_{k} \Gamma, A_0 \wedge A_1$, then $\vdash^{\alpha}_{k} \Gamma, A_i$ for every $i \leq 1$.
3. ($\forall$-Inversion) If $\vdash^{\alpha}_{k} \Gamma, \forall x A(x)$, then $\vdash^{\alpha}_{k} \Gamma, A(\bar{n})$ for every $n \in \mathbb{N}$.
4. ($\vee$-exportation) If $\vdash^{\alpha}_{k} \Gamma, A \vee B$, then $\vdash^{\alpha}_{k} \Gamma, A, B$.
5. (False Literal) If $\vdash^{\alpha}_{k} \Gamma, A$ for a false $\mathcal{L}$-literal A, then $\vdash^{\alpha}_{k} \Gamma$.
6. (Substitution) If $\vdash^{\alpha}_{k} \Gamma, A(s)$ and $s^{\mathbb{N}} = t^{\mathbb{N}}$, then $\vdash^{\alpha}_{k} \Gamma, A(t)$.

Lemma 25. (Reduction Lemma for $\mathsf{G}^{\infty}_{(rec)}$ [50, Lemma 3.19]) In $\mathrm{G}^{\infty}_{(rec)}$, the following holds:

$$\text{If } \vdash^{\alpha}_{k} \Gamma, A \text{ and } \vdash^{\beta}_{k} \Delta, \neg A \text{ hold with } |A| = k, \text{ then } \vdash^{\alpha \# \beta}_{k} \Gamma, \Delta.$$

Proof. We extend the proof for $\mathsf{PA}^{\infty}_{(rec)}$ (Lemma 18). Thus, the claim is shown by induction on $\alpha\#\beta$, and we divide the cases by whether A and $\neg A$ are principal in the last inference rule of the derivation, respectively.

Case I: either A or $\neg A$ is not principal. Although the proof is almost the same as for Case I of Lemma 18, we have to be careful when A is derived by $\mathsf{T\text{-}Intro}$:

$$\dfrac{\vdash^{\alpha_0}_{k} B}{\vdash^{\alpha}_{k} \Gamma', \mathrm{T}\ulcorner B\urcorner, A}\ \mathsf{T\text{-}Intro}\ , \text{ where } \Gamma = \Gamma' \cup \{\mathrm{T}\ulcorner B\urcorner\}.$$

In this case, we can directly derive $\vdash^{\alpha\#\beta}_{k} \Gamma, \Delta$ from $\vdash^{\alpha_0}_{k} B$ by $\mathsf{T\text{-}Intro}$.

Case II: both A and $\neg A$ are principal.

Case IIa: $|A| > 0$. Since none of Truth rules of $\mathsf{G}^{\infty}_{(rec)}$ have the principal sentence that has the logical complexity > 0, the proof is exactly the same as for Case IIa of Lemma 18.

Case IIb: $|A| = 0$. It suffices to consider the case $A \equiv \mathrm{T}(s)$. Then, $\neg A \equiv \neg\mathrm{T}(s)$ is principal only when $\Delta, \neg\mathrm{T}(s)$ is an instance of $(\mathsf{Ax}.2)$, and hence $\mathrm{T}(t)$ is contained in Δ for some $t^{\mathbb{N}} = s^{\mathbb{N}}$. Thus, $\vdash^{\alpha\#\beta}_{k} \Gamma, \Delta$ is obtained from $\vdash^{\alpha}_{k} \Gamma, A$ by Substitution and Weakening for $\mathsf{G}^{\infty}_{(rec)}$.

■

Remark 26. For the proof of Reduction Lemma for $\mathsf{G}^{\infty}_{(rec)}$ to succeed, it is important that the additional rules of $\mathsf{G}^{\infty}_{(rec)}$ have only the *atomic* principal sentence. That is the reason why Leigh and Rathjen [50] do not add the rule $\mathsf{T\text{-}Elim}$ to $\mathsf{G}^{\infty}_{(rec)}$ explicitly.

Lemma 27. (Cut elimination for $\mathsf{G}^{\infty}_{(rec)}$) The following is admissible:

$$\dfrac{\vdash^{\alpha}_{0} \Gamma, A \qquad \vdash^{\beta}_{0} \Delta, \neg A}{\vdash^{\omega_{|A|}(\alpha\#\beta)}_{0} \Gamma, \Delta}\ .$$

As is explained in Remark 26, the system $\mathsf{G}^{\infty}_{(rec)}$ does not have the rule $\mathsf{T\text{-}Elim}$. Thus, to prove Embedding Lemma, we must show the admissibility of $\mathsf{T\text{-}Elim}$.

Definition 28. (Basic T-positive sequent) A *basic* T-*positive* sequent Γ is of the following form:

$$\mathrm{T}\ulcorner A_0\urcorner, \mathrm{T}\ulcorner A_1\urcorner, \cdots, \mathrm{T}\ulcorner A_m\urcorner.$$

Proposition 29. (Soundness of G^∞ (cf. [50, Lemma 3.23])) Let Γ be basic T-positive. If $\mathrel{|}\!\frac{\alpha}{0}\, \Gamma$, then it follows that $\mathrel{|}\!\frac{\varphi 1\alpha}{0}\, A$ for some $\mathrm{T}\ulcorner A \urcorner \in \Gamma$.

Proof. The proof is by induction on α. We divide the cases by the last inference rule of the derivation.

$\mathsf{T\text{-}Intro}$. Assume that $\mathrel{|}\!\frac{\alpha}{0}\, \Gamma', \mathrm{T}\ulcorner A \urcorner$ is derived from $\mathrel{|}\!\frac{\alpha_0}{0}\, A$, where $\Gamma = \Gamma' \cup \{\mathrm{T}\ulcorner A \urcorner\}$. In this case, $\mathrel{|}\!\frac{\varphi 1\alpha}{0}\, A$ is obtained by Weakening.

Rep. For $\Gamma = \Gamma' \cup \{\mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner\}$, we assume that $\mathrel{|}\!\frac{\alpha}{0}\, \Gamma', \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner$ is derived from $\mathrel{|}\!\frac{\alpha_0}{0}\, \Gamma', \mathrm{T}\ulcorner A \urcorner$ by (Rep). Then, by the induction hypothesis, we have $\mathrel{|}\!\frac{\varphi 1\alpha_0}{0}\, B$ for some $\mathrm{T}\ulcorner B \urcorner \in \Gamma' \cup \{\mathrm{T}\ulcorner A \urcorner\}$. If $\mathrm{T}\ulcorner B \urcorner \in \Gamma' \subseteq \Gamma$, then $\mathrel{|}\!\frac{\varphi 1\alpha}{0}\, B$ follows by Weakening. If $\mathrm{T}\ulcorner B \urcorner = \mathrm{T}\ulcorner A \urcorner$, then we obtain $\mathrel{|}\!\frac{\varphi 1\alpha}{0}\, \mathrm{T}\ulcorner A \urcorner$ by $\mathsf{T\text{-}Intro}$.

The other cases are similar. Note that we do not have to consider the cut rule, for the cut rank is 0. ■

From Soundness of G^∞, we can immediately conclude the admissibility of ($\mathsf{T\text{-}Elim}$). Therefore, if $\mathsf{G} \vdash B$, then $\mathsf{G}^\infty \vdash B$. Moreover, if B does not contain the truth predicate, then the cut eliminability of G^∞ implies that B is derivable in PA^∞. Thus, B is proved to be arithmetically true.

As the final step of the upper-bound proof of G, [50, Corollary 3.27] claims that all the above arguments hold for G^∞_{rec}, and thus they are formalisable in Peano arithmetic expanded with a suitable transfinite induction schema. However, as far as their comments are literally understood, there is a counterexample to formalisability. As the consequence of the above Soundness, if $\mathsf{G}^\infty \vdash \mathrm{T}\ulcorner A \urcorner \lor \mathrm{T}\ulcorner B \urcorner$, then it follows that $\mathsf{G}^\infty \vdash \mathrm{T}\ulcorner A \urcorner$ or $\mathsf{G}^\infty \vdash \mathrm{T}\ulcorner B \urcorner$. This is called the *truth disjunction* property in [21]. If this property holds also for G^∞_{rec}, we can further show that the truth disjunction property holds for G itself. However, in the presence of the axiom ($\forall$-Inf), this property should fail for any recursively enumerable system ([21, Proposition 1.2]). Therefore, soundness is false for G^∞_{rec} though it is true for G^∞. I also give a concrete counterexample to the Soundness of G^∞_{rec}.

Proposition 30. (Counterexample to Soundness of G^∞_{rec})

1. For any $\mathcal{L}$-sentence A, there exists a proof of the sequent $\mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner \neg A \urcorner$ with length $\alpha < \omega$ in G^∞_{rec}.
2. There is an $\mathcal{L}$-sentence B such that neither B nor $\neg B$ is derivable in G^∞_{rec} with length $\varphi 1\omega$.
3. Thus, there is an $\mathcal{L}$-sentence B such that $\mathsf{G}^\infty_{rec} \mathrel{|}\!\frac{\omega}{0}\, \mathrm{T}\ulcorner B \urcorner, \mathrm{T}\ulcorner \neg B \urcorner$, whereas $\mathsf{G}^\infty_{rec} \mathrel{|}\!\not\frac{\varphi 1\omega}{0}\, B$ and $\mathsf{G}^\infty_{rec} \mathrel{|}\!\not\frac{\varphi 1\omega}{0}\, \neg B$.

Proof.

1 By induction on A. For example, the case $A \equiv \forall x B(x)$ is proved as follows:

$$\dfrac{\dfrac{\vdots}{\mathrm{T}\ulcorner B(n)\urcorner, \mathrm{T}\ulcorner \neg B(n)\urcorner}\ \mathrm{IH} \quad \dfrac{\dfrac{\vdots}{\neg B(n) \to \neg\forall x B(x)}\ \text{by logic}}{\mathrm{T}\ulcorner B(n)\urcorner, \mathrm{T}\ulcorner \neg B(n) \to \neg\forall x B(x)\urcorner}\ (\mathsf{T\text{-}Intro})}{\dfrac{\cdots \mathrm{T}\ulcorner B(n)\urcorner, \mathrm{T}\ulcorner \neg\forall x B(x)\urcorner \cdots}{\mathrm{T}\ulcorner \forall x B(x)\urcorner, \mathrm{T}\ulcorner \neg\forall x B(x)\urcorner}\ (\mathrm{T}_\forall)}\ (\mathrm{T}_\to)$$

2 Let $Bew_{\mathsf{G}^\infty_{rec}}(x, y)$ be a derivability predicate for G^∞_{rec}, meaning that x is derivable with length y in G^∞_{rec}. By Gödel's lemma, let B be an $\mathcal{L}$-sentence such that $B \leftrightarrow \neg Bew_{\mathsf{G}^\infty_{rec}}(\ulcorner B\urcorner, \varphi 1\omega)$; thus, B asserts that B is not derivable with the length $\varphi 1\omega$. If $\mathsf{G}^\infty_{rec} \left|\frac{\varphi 1\omega}{\quad}\right. B$, then $Bew_{\mathsf{G}^\infty_{rec}}(\ulcorner B\urcorner, \varphi 1\omega)$ is true, and thus, $\neg B$ is true. However, we can easily show that any $\mathcal{L}$-sentence cut freely derivable in $\mathsf{G}^\infty_{(rec)}$ is true in $\mathbb{N}$; thus, B should also be true, a contradiction. Therefore, $\mathsf{G}^\infty_{rec} \not\left|\frac{\varphi 1\omega}{\quad}\right.$ B. Next, if $\mathsf{G}^\infty_{rec} \left|\frac{\varphi 1\omega}{\quad}\right. \neg B$, then we also have $\mathsf{G}^\infty_{rec} \left|\frac{\varphi 1\omega+1}{\quad}\right. Bew_{\mathsf{G}^\infty_{rec}}(\ulcorner B\urcorner, \varphi 1\omega)$. Thus, again by the $\mathcal{L}$-soundness of G^∞_{rec}, it follows that $Bew_{\mathsf{G}^\infty_{rec}}(\ulcorner B\urcorner, \varphi 1\omega)$ is true, which implies $\mathsf{G}^\infty_{rec} \left|\frac{\varphi 1\omega}{\quad}\right. B$. This, combined with the current assumption $\mathsf{G}^\infty_{rec} \left|\frac{\varphi 1\omega}{\quad}\right. \neg B$, implies that G^∞_{rec} is inconsistent, which contradicts the consistency of $\mathsf{G}^\infty_{(rec)}$. Therefore, $\mathsf{G}^\infty_{rec} \not\left|\frac{\varphi 1\omega}{\quad}\right. \neg B$. In summary, neither B nor $\neg B$ is derivable with length $\varphi 1\omega$.

3 Item 3 is an immediate consequence of items 1 and 2.

■

The above counterexample means that Soundness (Proposition 29) fails for G^∞_{rec}, and hence, it is not clear whether the admissibility of $\mathsf{T\text{-}Elim}$ holds in G^∞_{rec}. However, as we have seen in Section 3.1.1, embedding into the recursive system is required for the upper-bound proof. In addition, the same problem occurs also in the other systems E, F, and I. In the next subsection, to solve this problem, we instead prove that a property weaker than Soundness is sufficient for the admissibility proof of $\mathsf{T\text{-}Elim}$.

3.2 Reconsidering Upper Bounds of $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I

In Section 3.2.1, by modifying Leigh and Rathjen's original proof, we verify that their ordinal bound $\varphi 20$ for G is correct. In Section 3.2.2, the system I is addressed. The analyses of F (Section 3.2.3) and E (Section 3.2.4) are more complicated than those of G and I. In particular, we need to present the derivation systems for them as labelled calculi.

In conclusion, the results of Section 3.2 are summarised as follows:

Theorem 34 (cf. [50, Corollary 3.27]) $\mathsf{G} \leq \mathsf{PA} + \mathsf{TI}(< \varphi 20)$.

Theorem 48 (cf. [50, Corollary 3.35]) $\mathsf{I} \leq \mathsf{PA} + \mathsf{TI}(< \varphi 20)$.

Theorem 63 (cf. [50, Theorem 3.15]) $\mathsf{F} \leq \mathsf{PA} + \mathsf{TI}(< \varphi\omega 0)$.

Theorem 73 (cf. [50, Theorem 3.45]) $\mathsf{E} \leq \mathsf{PA} + \mathsf{TI}(< \varphi\omega 0)$.

Note that Leigh and Rathjen [50] also give the lower-bound proofs for $\mathsf{G}, \mathsf{I}, \mathsf{F}$, and E. Thus, each of the above ordinal numbers is an exact proof-theoretic ordinal.

3.2.1 An Analysis of G

In the previous section, we showed that the truth disjunction property is too strong to determine the proof-theoretic upper bound. As a solution, I shall use a weaker property, *strong disquotability*. To illustrate this, let us reconsider what property must be established to show the admissibility of ($\mathsf{T\text{-}Elim}$). Thus, we assume that $\mathrm{T}\ulcorner A \urcorner$ is derived. Then, our purpose is to *disquote* it, i.e., to derive A. Now, the premises of $\mathrm{T}\ulcorner A \urcorner$ do not always consist of only one sentence. For instance, consider the following derivation of $A \equiv \mathrm{T}\ulcorner B \urcorner$:

$$\dfrac{\dfrac{\vdots}{\vdash \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner, \mathrm{T}\ulcorner B \urcorner}}{\vdash \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner} \ \mathsf{Rep}$$

To derive $\mathrm{T}\ulcorner B \urcorner$, we want to use the induction hypothesis for the premise $\vdash \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner, \mathrm{T}\ulcorner B \urcorner$. But how? Our idea is to disquote $\mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner$ only:

$$\vdash \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner, \mathrm{T}\ulcorner B \urcorner \quad \overset{\text{disquote only } \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner}{\Longrightarrow} \quad \vdash \mathrm{T}\ulcorner B \urcorner, \mathrm{T}\ulcorner B \urcorner$$

In general, for a basic T-positive sequent $\mathrm{T}\ulcorner A_0 \urcorner, \mathrm{T}\ulcorner A_1 \urcorner, \cdots, \mathrm{T}\ulcorner A_n \urcorner$, it is sufficient that *any* of its subsets is disquotable. If this is the case, then I shall say that this sequent is *strongly disquotable*. If all the derived basic T-positive sequents are strongly disquotable in a derivation system, then I shall say that the system is *strongly disquotable*. We can easily observe that strong disquotability is weaker than the truth disjunction property in the presence of $\mathsf{T\text{-}Intro}$. Thus, my aim is to show the strong disquotability of $\mathsf{G}^{\infty}_{rec}$, in a way formalisable in $\mathsf{PA} + \mathsf{TI}(< \varphi 20)$. Then, we can obtain the admissibility of T-Elim in $\mathsf{G}^{\infty}_{rec}$, provably in $\mathsf{PA}+\mathsf{TI}(< \varphi 20)$.

Definition 31. (Disquotation for G) Let Γ be a basic T-positive sequent of the following form:

$$\mathrm{T}\ulcorner A_0 \urcorner, \mathrm{T}\ulcorner A_1 \urcorner, \cdots, \mathrm{T}\ulcorner A_m \urcorner.$$

Then, let the *disquotation* $Disq(\Gamma)$ of Γ be the sequent:

$$A_0, A_1, \cdots A_m.$$

Lemma 32. (Strong disquotation for $\mathsf{G}^{\infty}_{(rec)}$) Let Γ, Δ be basic T-positive. If $\mathsf{G}^{\infty}_{rec} \vdash^{\alpha}_{0} \Gamma, \Delta$ then $\mathsf{G}^{\infty}_{rec} \vdash^{\varphi 1 \alpha}_{0} \Gamma, Disq(\Delta)$.

In particular, there exists a partial recursive function $F_{SD}(x, y)$ such that if $d \vdash^{\alpha}_{0} \Gamma, \Delta$, then $F_{SD}(d, \ulcorner \Delta \urcorner) \vdash^{\varphi 1 \alpha}_{0} \Gamma, Disq(\Delta)$ holds.

Proof. To construct a required function $F_{SD}(x, y)$, we divide the cases by the last rule of the derivation. The most non-trivial case is where there are infinitely many premises, so we concentrate on the rule $(\mathsf{T}_{\forall})$.

$(\mathsf{T}_{\forall})$ For $\Gamma, \Delta = \Sigma' \cup \{\mathrm{T}\ulcorner \forall x A(x) \urcorner\}$, we assume that the recursive derivation d is of the following form:

$$\dfrac{\cdots \overset{\vdots\, d_n}{\vdash^{\alpha_n}_{0} \Sigma', \mathrm{T}\ulcorner A(n) \urcorner} \cdots (n \in \mathbb{N})}{\vdash^{\alpha}_{0} \Sigma', \mathrm{T}\ulcorner \forall x A(x) \urcorner}\ (\mathsf{Rep}),$$

where d_n is a subderivation of d for each $n \in \mathbb{N}$. Note that since d is a recursive derivation, we have a recursive function that enumerates $d_0, d_1, \cdots, d_n, \cdots$.

1. If $\mathrm{T}\ulcorner \forall x A(x) \urcorner \in \Delta$, then we have $\Delta = (\Sigma' \cap \Delta) \cup \{\mathrm{T}\ulcorner \forall x A(x) \urcorner\}$, and hence, $Disq(\Delta) = Disq(\Sigma' \cap \Delta) \cup \{\forall x A(x)\}$.
   By the induction hypothesis for d_n, we obtain:
   $$F_{SD}(d_n, \ulcorner (\Sigma' \cap \Delta) \cup \{\mathrm{T}\ulcorner A(n) \urcorner\} \urcorner) \vdash^{\varphi 1 \alpha_n}_{0} \Sigma' \cap \Gamma, A(n), Disq(\Sigma' \cap \Delta).$$
   By Weakening, it follows that:
   $$F_{SD}(d_n, \ulcorner (\Sigma' \cap \Delta) \cup \{\mathrm{T}\ulcorner A(n) \urcorner\} \urcorner) \vdash^{\varphi 1 \alpha_n}_{0} \Gamma, A(n), Disq(\Sigma' \cap \Delta).$$
   Now, we have a recursive function that returns the value $F_{SD}(d_n, \ulcorner (\Sigma' \cap \Delta) \cup \{\mathrm{T}\ulcorner A(n) \urcorner\} \urcorner)$ for each n. Thus, by the rule $(\forall)$ we can define the value of $F_{SD}(d, \ulcorner \Delta \urcorner)$ so that:
   $$F_{SD}(d, \ulcorner \Delta \urcorner) \vdash^{\varphi 1 \alpha}_{0} \Gamma, \forall x A(x), Disq(\Sigma' \cap \Delta).$$
   As $\Delta = (\Sigma' \cap \Delta) \cup \{\mathrm{T}\ulcorner \forall x A(x) \urcorner\}$, this is equivalent to:
   $$F_{SD}(d, \ulcorner \Delta \urcorner) \vdash^{\varphi 1 \alpha}_{0} \Gamma, Disq(\Delta).$$

2. If $\mathrm{T}\ulcorner\forall x A(x)\urcorner \notin \Delta$, then $\mathrm{T}\ulcorner\forall x A(x)\urcorner \in \Gamma$.
   By the induction hypothesis for d_n, we obtain:

$$F_{SD}(d_n, \ulcorner\Sigma' \cap \Delta\urcorner) \Big|\frac{\varphi 1\alpha_n}{0} \; \Sigma' \cap \Gamma, \mathrm{T}\ulcorner A(n)\urcorner, Disq(\Sigma' \cap \Delta).$$

   As above, we have a recursive function that returns the value $F_{SD}(d_n, \ulcorner\Sigma' \cap \Delta\urcorner)$ for each n. Therefore, by ($\mathsf{T}_\forall$), we have a recursive derivation e:

$$e \Big|\frac{\varphi 1\alpha}{0} \; \Sigma' \cap \Gamma, \mathrm{T}\ulcorner\forall x A(x)\urcorner, Disq(\Sigma' \cap \Delta).$$

   Thus, by Weakening, we can recursively define the value of $F_{SD}(d, \ulcorner\Delta\urcorner)$ such that:

$$F_{SD}(d, \ulcorner\Delta\urcorner) \Big|\frac{\varphi 1\alpha}{0} \; \Gamma, Disq(\Delta).$$

The other cases are easier. ■

From the strong disquotability of G^∞_{rec}, the admissibility of T-Elim directly follows, and so does Embedding Lemma of G:

Lemma 33. (Embedding of G) If $\mathsf{G} \vdash A$, then it follows that $\mathsf{G}^\infty_{rec} \Big|\frac{\alpha}{0} A$ for some $\alpha < \varphi 20$.

Theorem 34. (Upper bound of G) Let A be an $\mathcal{L}$-sentence.
If $\mathsf{G} \vdash A$, then $\mathsf{PA} + \mathsf{TI}(< \varphi 20) \vdash A$.

3.2.2 An Analysis of I

Definition 35. (System I) An $\mathcal{L}_\mathrm{T}$-theory I consists of Base_T with the following axioms and rules:

(Rep) $\forall\ulcorner A\urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{T}\ulcorner A\urcorner \to \mathrm{T}\ulcorner\mathrm{T}\ulcorner A\urcorner\urcorner$;

(Del) $\forall\ulcorner A\urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \mathrm{T}\ulcorner\mathrm{T}\ulcorner A\urcorner\urcorner \to \mathrm{T}\ulcorner A\urcorner$;

($\mathrm{T}_\forall$) $\forall\ulcorner A_x\urcorner \in \mathrm{Sent}_{\mathcal{L}_\mathrm{T}}.\ \forall v \mathrm{T}\ulcorner A(\dot{v})\urcorner \to \mathrm{T}\ulcorner\forall x A(x)\urcorner$;

(T-Elim) from $\mathrm{T}\ulcorner A\urcorner$ infer A.

($\neg$T-Elim) from $\neg\mathrm{T}\ulcorner A\urcorner$ infer $\neg A$.

Similar to G^∞, Leigh and Rathjen [50, Definition 3.28] define an infinitary system (which we denote as $\mathsf{I}^\infty_{L\&R}$) in which I is embeddable. First, let us briefly observe that $\mathsf{I}^\infty_{L\&R}$ has the same problem as G^∞.

Definition 36. (System $\mathsf{PA}^{\infty}_{\mathrm{T}}$ and $\mathsf{I}^{\infty}_{L\&R}$) The derivation system $\mathsf{PA}^{\infty}_{\mathrm{T}}$ consists of PA^{∞} (Definition 15) with the rule $\mathsf{T\text{-}Intro}$ of G^{∞} (Definition 23). The calculus $\mathsf{I}^{\infty}_{L\&R}$ is obtained from G^{∞} by eliminating the rule $\mathsf{T\text{-}Intro}$ and instead adding the axiom $(\mathsf{Ax}.3)$ and the rule (Del):

$$\frac{\mathsf{PA}^{\infty}_{\mathrm{T}} \left|\frac{\alpha}{0}\right. A}{\left|\frac{\alpha}{k}\right. \Gamma, \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{Ax}.3), \qquad \frac{\left|\frac{\alpha}{k}\right. \Gamma, \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner}{\left|\frac{\alpha}{k}\right. \Gamma, \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{Del}).$$

Note that $(\mathsf{Ax}.3)$ says that if A is derived in $\mathsf{PA}^{\infty}_{\mathrm{T}}$, then $\Gamma, \mathrm{T}\ulcorner A \urcorner$ is derived in $\mathsf{I}^{\infty}_{L\&R}$. In addition, it is easy to prove that $\mathsf{PA}^{\infty}_{\mathrm{T}}$ is a subsystem of $\mathsf{I}^{\infty}_{L\&R}$ [50, Proposition 3.29].

Similar to G^{∞}, it is shown in [50] that the basic properties (Lemma 24) and Cut Elimination (Lemma 27) hold in both $\mathsf{PA}^{\infty}_{\mathrm{T}}$ and $\mathsf{I}^{\infty}_{L\&R}$. Thus, the remaining task is to show the admissibility of $\mathsf{T\text{-}Elim}$ and $\neg\mathsf{T\text{-}Elim}$. To prove the former, they showed the following:

Proposition 37. (Soundness of $\mathsf{I}^{\infty}_{L\&R}$ [50, Lemma 3.32]) Assume that Γ is basic T-positive. If $\mathsf{I}^{\infty}_{L\&R} \left|\frac{\alpha}{0}\right. \Gamma$, then it follows that $\mathsf{PA}^{\infty}_{\mathrm{T}} \left|\frac{\varphi 1 \alpha}{0}\right. A$ for some $\mathrm{T}\ulcorner A \urcorner \in \Gamma$.

From this proposition, the admissibility of $\mathsf{T\text{-}Elim}$ follows, because $\mathsf{PA}^{\infty}_{\mathrm{T}}$ is a subsystem of $\mathsf{I}^{\infty}_{L\&R}$. But again, the problem is that its recursive version is false. In fact, exactly the same counterexample as for $\mathsf{G}^{\infty}_{rec}$ (Proposition 30) applies.

To solve this, we now define an infinitary system $\mathsf{I}^{\infty}_{(rec)}$ in which I is embeddable. Although the recursive version of Proposition 37 is false, our proof is essentially based on Leigh and Rathjen's original argument. To be more precise, the truth predicate in $\mathsf{I}^{\infty}_{(rec)}$ intuitively represents the sequent calculus $\mathsf{PA}^{\infty}_{\mathrm{T}}$ itself; thus, the truth predicate now applies not to sentences, but to sequents. Since a sequent is a finite set of sentences, it can be coded by an appropriate Gödel numbering.

As is the case with $\mathsf{G}^{\infty}_{rec}$, the basic properties, Cut Elimination, and Disquotation for the recursive version of both $\mathsf{PA}^{\infty}_{\mathrm{T}}$ and I^{∞} are provable in an effective manner.

Definition 38. (System $\mathsf{I}^{\infty}_{(rec)}$) The derivation system I^{∞} consists of PA^{∞} and the truth rules in Table 3.3. The calculus $\mathsf{I}^{\infty}_{rec}$ is a subsystem of I^{∞} in which only recursive derivations are allowed.

Lemma 39. The following basic properties hold in $\mathrm{I}^{\infty}_{(rec)}$.

1. (Weakening) If $\left|\frac{\alpha}{k}\right. \Gamma$, then $\left|\frac{\beta}{l}\right. \Delta$ for any $\alpha \leq \beta$, $\Gamma \subseteq \Delta$, and $k \leq l$.
2. ($\wedge$-Inversion) If $\left|\frac{\alpha}{k}\right. \Gamma, A_0 \wedge A_1$, then $\left|\frac{\alpha}{k}\right. \Gamma, A_i$ for every $i \leq 1$.

Table 3.3: System I^∞

Truth rules (assume: $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$(\mathrm{T}_{\mathsf{Ax.1}})$ $\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A\urcorner$, if A is a true arithmetical literal.
$(\mathrm{T}_{\mathsf{Ax.2}})$ $\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, \neg \mathrm{T}s, \mathrm{T}t\urcorner$, if $s^{\mathbb{N}} = t^{\mathbb{N}}$.

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A_0, A_1\urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A_0 \vee A_1\urcorner}\ (\mathrm{T}_\vee) \qquad \frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A_0\urcorner \quad \vdash^{\alpha_1}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A_1\urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A_0 \wedge A_1\urcorner}\ (\mathrm{T}_\wedge)$$

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A(i)\urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, \exists x A\urcorner}\ (\mathrm{T}_\exists) \qquad \frac{\ldots \vdash^{\alpha_i}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A(i)\urcorner \ldots\ (i \in \mathbb{N})}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, \forall x A\urcorner}\ (\mathrm{T}_\forall)$$

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A\urcorner \quad \vdash^{\alpha_1}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, \neg A\urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta\urcorner}\ (\mathrm{T}_{\mathsf{cut}})$$

$$\frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner A\urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, \mathrm{T}\ulcorner A\urcorner\urcorner}\ (\mathsf{Rep}) \qquad \frac{\vdash^{\alpha_0}_{k} \Gamma, \mathrm{T}\ulcorner \mathrm{T}\ulcorner A\urcorner\urcorner}{\vdash^{\alpha}_{k} \Gamma, \mathrm{T}\ulcorner \Delta, A\urcorner}\ (\mathsf{Del}).$$

3. ($\forall$-Inversion) If $\vdash^{\alpha}_{k} \Gamma, \forall x A(x)$, then $\vdash^{\alpha}_{k} \Gamma, A(\bar{n})$ for every $n \in \mathbb{N}$.
4. ($\vee$-Exportation) If $\vdash^{\alpha}_{k} \Gamma, A \vee B$, then $\vdash^{\alpha}_{k} \Gamma, A, B$.
5. (False Literal) If $\vdash^{\alpha}_{k} \Gamma, A$ for a false $\mathcal{L}$-literal A, then $\vdash^{\alpha}_{k} \Gamma$.
6. (Substitution) If $\vdash^{\alpha}_{k} \Gamma, A(s)$ and $s^{\mathbb{N}} = t^{\mathbb{N}}$, then $\vdash^{\alpha}_{k} \Gamma, A(t)$.

The system I^∞ is defined such that the principal sentence of each truth rule is atomic. Thus, as is noted in Remark 26, Cut elimination is provable in the same way as for $\mathsf{G}^\infty_{(rec)}$.

Lemma 40. (Cut Elimination for $\mathsf{I}^\infty_{(rec)}$) The following is admissible:

$$\frac{\vdash^{\alpha}_{0} \Gamma, A \quad \vdash^{\beta}_{0} \Delta, \neg A}{\vdash^{\omega_{|A|}(\alpha \# \beta)}_{0} \Gamma, \Delta}\ .$$

Next, we aim to show the admissibility of $\mathsf{T\text{-}Elim}$. If $\mathsf{I}^\infty_{(rec)}$ does not have the rule (Del), then the system can be roughly seen as a subsystem of $\mathsf{G}^\infty_{(rec)}$, and thus, we can show the strong disquotability, similar to Lemma 32. But, the elimination of (Del) intuitively means that the inside of the truth predicate, i.e., $\mathsf{PA}^{\infty}_{\mathrm{T}}$, admits the rule $\mathsf{T\text{-}Elim}$. Since $\mathsf{PA}^{\infty}_{\mathrm{T}}$ is a proper subsystem of G^{∞}, we can show this in the same way as for G^{∞}.

Definition 41. (Basic T-positive sequent for I) Let $\varDelta_0, \varDelta_1, \cdots \varDelta_m$ be sequents. A *basic* T-*positive* sequent $\varGamma$ is of the following form:

$$\mathrm{T}\ulcorner \varDelta_0 \urcorner, \mathrm{T}\ulcorner \varDelta_1 \urcorner, \cdots, \mathrm{T}\ulcorner \varDelta_m \urcorner.$$

For this $\varGamma$, let the *disquotation* $Disq(\varGamma)$ be the sequent:

$$\varDelta_0, \varDelta_1, \cdots \varDelta_m.$$

Definition 42. 1. $\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}}} \varGamma$ means that $\varGamma$ is derived without (cut) nor ($\mathrm{T}_{\mathsf{cut}}$).

2. $\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma$ means that $\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}}} \varGamma$ is obtained without (Del).

The basic properties and Cut Elimination hold *inside* the truth predicate, the proofs of which are almost the same as for $\mathsf{G}^{\infty}_{(rec)}$.

Lemma 43. ($\mathrm{T}_{\mathsf{cut}}$-Weakening for $\mathsf{I}^{\infty}_{(rec)}$) Let $\varGamma$ be a basic T-positive sequent, and we take any sequents $\varDelta$ and $\varDelta'$ such that $\varDelta \subseteq \varDelta'$. If $\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner \varDelta \urcorner$, then $\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner \varDelta' \urcorner$.

Lemma 44. ($\mathrm{T}_{\mathsf{cut}}$-Elimination for $\mathsf{I}^{\infty}_{(rec)}$) Let $\varGamma$ be basic T-positive. The following is admissible:

$$\frac{\vdash^{\alpha_0}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner \varDelta, A \urcorner \qquad \vdash^{\alpha_1}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner \varSigma, \neg A \urcorner}{\vdash^{\omega_{|A|}(\alpha_0 \# \alpha_1)}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner \varDelta, \varSigma \urcorner}.$$

Lemma 45. (Elimination of $\mathrm{T}_{\mathsf{cut}}$ and Del) Let $\varGamma, \varDelta$ be basic T-positive.

1. If $\mathsf{I}^{\infty}_{(rec)} \vdash^{\alpha}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner \varDelta \urcorner$, then $\mathsf{I}^{\infty}_{(rec)} \vdash^{\alpha}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma, \mathrm{T}\ulcorner Disq(\varDelta) \urcorner$.
2. If $\mathsf{I}^{\infty}_{(rec)} \vdash^{\alpha}_{0} \varGamma$, then $\mathsf{I}^{\infty}_{(rec)} \vdash^{\varphi 1 \alpha}_{\text{-T}_{\mathsf{cut}},\text{-}\mathsf{Del}} \varGamma$.

Proof. Although we consider only I^{∞}, it should be clear from the proof that the result holds also in $\mathsf{I}^{\infty}_{rec}$, similar to Lemma 32.

1. By induction on α. We write $\Delta := \mathrm{T}\ulcorner A_0 \urcorner, \cdots, \mathrm{T}\ulcorner A_m \urcorner$. Assume $\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}}$ $\Gamma, \mathrm{T}\ulcorner \mathrm{T}\ulcorner A_0 \urcorner, \cdots, \mathrm{T}\ulcorner A_m \urcorner \urcorner$. We consider the case where $\mathrm{T}\ulcorner \mathrm{T}\ulcorner A_0 \urcorner, \cdots, \mathrm{T}\ulcorner A_m \urcorner \urcorner$ is principal in the last rule of the derivation. This is possible only when the rule (Rep) is applied:
$$\frac{\vdash^{\alpha_0}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}} \Gamma, \mathrm{T}\ulcorner A_m \urcorner}{\vdash^{\alpha}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}} \Gamma, \mathrm{T}\ulcorner \mathrm{T}\ulcorner A_0 \urcorner, \cdots, \mathrm{T}\ulcorner A_{m-1} \urcorner, \mathrm{T}\ulcorner A_m \urcorner \urcorner}\ (\mathsf{Rep}).$$
Then, using Weakening and $\mathrm{T}_{\mathsf{cut}}$-Weakening, the claim is obvious from the induction hypothesis.
2. By induction on α. The important cases are where $(\mathrm{T}_{\mathsf{cut}})$ or (Del) is the last rule of the derivation.

   (Del) Assume the following derivation:
$$\frac{\vdash^{\alpha_0}_{0} \Gamma', \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner}{\vdash^{\alpha}_{0} \Gamma', \mathrm{T}\ulcorner \Sigma, A \urcorner}\ (\mathsf{Del}),$$
   where $\Gamma = \Gamma' \cup \{\mathrm{T}\ulcorner \Sigma, A \urcorner\}$.

   By the induction hypothesis, we have:
$$\vdash^{\varphi 1 \alpha_0}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}} \Gamma', \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner.$$
   Thus, by item 1 of this lemma, it follows that:
$$\vdash^{\varphi 1 \alpha_0}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}} \Gamma', \mathrm{T}\ulcorner A \urcorner.$$
   Therefore, Weakening and $\mathrm{T}_{\mathsf{cut}}$-Weakening imply the claim:
$$\vdash^{\varphi 1 \alpha}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}} \Gamma', \mathrm{T}\ulcorner \Sigma, A \urcorner.$$

■

Lemma 46. (Admissibility of T-Elim) If $\mathsf{I}^{\infty}_{(rec)} \vdash^{\alpha}_{0} \mathrm{T}\ulcorner \{A\} \urcorner$, then $\mathsf{I}^{\infty}_{(rec)} \vdash^{\varphi 1 (\varphi 1 \alpha)}_{0} A$.

Proof. Assume $\vdash^{\alpha}_{0} \mathrm{T}\ulcorner \{A\} \urcorner$. By Lemma 45, we have $\vdash^{\varphi 1 (\alpha)}_{\text{-T}_{\mathsf{cut}}\text{,-}\mathsf{Del}} \mathrm{T}\ulcorner \{A\} \urcorner$. Since this derivation does not contain (Del), it can be seen as in G^{∞}. Thus, by the same proof as for Lemma 32, we obtain $\vdash^{\varphi 1 (\varphi 1 \alpha)}_{0} A$ in I^{∞}. ■

The truth predicate of $\mathsf{I}^{\infty}_{(rec)}$ represents sequents, not sentences. Thus, to embed I into $\mathsf{I}^{\infty}_{rec}$, we use an $\mathcal{L}$-preserving translation A^{seq} as $(\mathrm{T}(x))^{seq} :\equiv \mathrm{T}(\{x\})$, where the primitive recursive function $\{x\}$ returns (the code of) the sequent $\{A\}$ for the code $x = \ulcorner A \urcorner$ of a sentence.

Lemma 47. (Embedding of I) If $\mathsf{I} \vdash A$, then for some $\alpha < \varphi 20$, we have $\mathsf{I}^{\infty}_{rec} \overset{\alpha}{\underset{0}{\vdash}} A^{seq}$.

Theorem 48. (Upper-bound of I) Let A be an $\mathcal{L}$-sentence.
If $\mathsf{I} \vdash A$, then $\mathsf{PA} + \mathsf{TI}(< \varphi 20) \vdash A$.

3.2.3 An Analysis of F

Definition 49. (System F) An $\mathcal{L}_{\mathrm{T}}$-theory F consists of $\mathsf{Base}_{\mathrm{T}}$ with the following axioms and rules:

(Del) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}. \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner \to \mathrm{T}\ulcorner A \urcorner$;

($\mathrm{T}_{\forall}$) $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}. \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner \to \mathrm{T}\ulcorner \forall x A(x) \urcorner$;

(T-Intro) from A infer $\mathrm{T}\ulcorner A \urcorner$;

(T-Elim) from $\mathrm{T}\ulcorner A \urcorner$ infer A;

($\neg$T-Elim) from $\neg \mathrm{T}\ulcorner A \urcorner$ infer $\neg A$.

To analyse F, Leigh and Rathjen [50] define a system (which we call $\mathsf{F}^{\infty}_{L\&R}$) in which the T-Intro-rank is recorded.

Definition 50. (System $\mathsf{F}^{\infty}_{L\&R}$ [50, Definition 3.3]) The derivation system $\mathsf{F}^{\infty}_{L\&R}$ consists of PA^{∞} with the truth rules in Table 3.4. In each derivation $\overset{\alpha}{\underset{k,m}{\vdash}} \Gamma$, the natural number m denotes the T-Intro-rank, i.e., the number of applications of the rule (T-Intro) in the derivation. This rank increases only when the rule (T-Intro) is applied.

For $\mathsf{F}^{\infty}_{L\&R}$, the basic properties and Cut Elimination are proved in [50]. To show the admissibility of T-Elim, they also give the following soundness result:

Proposition 51. (Soundness of $\mathsf{F}^{\infty}_{L\&R}$ [50, Lemma 3.10]) Assume that Γ is basic T-positive. If $\mathsf{F}^{\infty}_{L\&R} \overset{\alpha}{\underset{0,n+1}{\vdash}} \Gamma$, then it follows that $\mathsf{F}^{\infty}_{L\&R} \overset{\varphi(n+1)\alpha}{\underset{0,n}{\vdash}} A$ for some $\mathrm{T}\ulcorner A \urcorner \in \Gamma$.

Proof. By induction on n and α. To see why T-Intro-rank has to be recorded, we consider the case where Γ is derived by (Del). So, for $\Gamma = \Gamma' \cup \{\mathrm{T}\ulcorner A \urcorner\}$, we assume the following derivation:

$$\frac{\overset{\alpha_0}{\underset{0,n+1}{\vdash}} \Gamma', \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner}{\overset{\alpha}{\underset{0,n+1}{\vdash}} \Gamma', \mathrm{T}\ulcorner A \urcorner} (\mathsf{Del}).$$

Table 3.4: System $\mathsf{F}^{\infty}_{L\&R}$

Truth rules (assume: $m_0 < m$ and $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$$\frac{\left|\frac{\alpha_0}{k,m_0}\right. A}{\left|\frac{\alpha}{k,m}\right. \Gamma, \mathrm{T}\ulcorner A \urcorner} \ (\mathsf{T\text{-}Intro}) \qquad \frac{\left|\frac{\alpha_0}{k,m}\right. \Gamma, \mathrm{T}\ulcorner A \urcorner \quad \left|\frac{\alpha_1}{k,m}\right. \Gamma, \mathrm{T}\ulcorner A \to B \urcorner}{\left|\frac{\alpha}{k,m}\right. \Gamma, \mathrm{T}\ulcorner B \urcorner} \ (\mathsf{T}_{\to})$$

$$\frac{\ldots \left|\frac{\alpha_i}{k,m}\right. \Gamma, \mathrm{T}\ulcorner A(i) \urcorner \ldots \ (i \in \mathbb{N})}{\left|\frac{\alpha}{k,m}\right. \Gamma, \mathrm{T}\ulcorner \forall x A \urcorner} \ (\mathsf{T}_{\forall}) \qquad \frac{\left|\frac{\alpha}{k,m}\right. \Gamma, \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner}{\left|\frac{\alpha_0}{k,m}\right. \Gamma, \mathrm{T}\ulcorner A \urcorner} \ (\mathsf{Del})$$

By the sub-induction hypothesis, we get $\left|\frac{\varphi(n+1)\alpha_0}{0,n}\right. B$ for some $\mathrm{T}\ulcorner B \urcorner \in \Gamma' \cup \{\mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner\}$. If $\mathrm{T}\ulcorner B \urcorner \in \Gamma'$, then the claim is obvious by Weakening. Thus, we may suppose $B \equiv \mathrm{T}\ulcorner A \urcorner$. Note that from the definition of $\mathsf{F}^{\infty}_{L\&R}$, a basic T-positive sequent is derivable only when the $\mathsf{T\text{-}Intro}$-rank is larger than 0. Therefore, we have $n > 0$, and hence, we can use the main induction hypothesis to obtain $\left|\frac{\varphi n(\varphi(n+1)\alpha_0)}{0,n-1}\right. A$. Since $\varphi n(\varphi(n+1)\alpha_0) = \varphi(n+1)\alpha_0 < \varphi(n+1)\alpha$, the claim follows by Weakening. ■

As for the recursive version $(\mathsf{F}^{\infty}_{L\&R})_{rec}$ of $\mathsf{F}^{\infty}_{L\&R}$, we have the following counterexample:

Proposition 52. (Counterexample to Soundness of $(\mathsf{F}^{\infty}_{L\&R})_{rec}$)

1. For any $\mathcal{L}$-sentence A, there exists a proof of the sequent $\mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner \neg A \urcorner$ with length $\alpha < \omega$ and with the $\mathsf{T\text{-}Intro}$-rank $= 1$ in $(\mathsf{F}^{\infty}_{L\&R})_{rec}$.
2. There is an $\mathcal{L}$-sentence B such that neither B nor $\neg B$ is derivable in $(\mathsf{F}^{\infty}_{L\&R})_{rec}$ with length $\varphi 1 \omega$.
3. Thus, there is an $\mathcal{L}$-sentence B such that $(\mathsf{F}^{\infty}_{L\&R})_{rec} \left|\frac{\omega}{0,1}\right. \mathrm{T}\ulcorner B \urcorner, \mathrm{T}\ulcorner \neg B \urcorner$, whereas $(\mathsf{F}^{\infty}_{L\&R})_{rec} \not\left|\frac{\varphi 1 \omega}{0,n}\right. B$ and $(\mathsf{F}^{\infty}_{L\&R})_{rec} \not\left|\frac{\varphi 1 \omega}{0,n}\right. \neg B$ for any n.

Thus, to determine the upper bound of F, we need a recursive procedure that transforms each derivation of $\mathrm{T}\ulcorner A \urcorner$ into that of A. Let us consider the following case:

$$\frac{\vdash \mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner\urcorner}{\vdash \mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner B \urcorner} \ (\mathsf{Del})$$

To show the strong disquotability as for $\mathsf{G}^{\infty}_{rec}$, we need to obtain derivations of three sequents, $A, \mathrm{T}\ulcorner B \urcorner$ and $\mathrm{T}\ulcorner A \urcorner, B$ and $A, B,$. As an example, we want to get

$\mathrm{T}\ulcorner A \urcorner, B$ as follows: by applying IH to the premise, we have $\mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner B \urcorner$, and then one more application of IH should give $\mathrm{T}\ulcorner A \urcorner, B$.

As we have seen in the proof of Proposition 51, we can indeed apply the induction hypothesis twice, by considering the T-Intro-rank. However, T-Intro rank might not be useful when considering recursive derivations. Consider the following derivation:

$$\dfrac{\dfrac{\vdash^{n} A}{\vdash^{n+1} \mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner}\ \textsf{T-Intro}}{\vdash^{n+1} \mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner B \urcorner}\ \textsf{Del}$$

In this case, we cannot achieve $\mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner B \urcorner$ with a T-Intro rank $\leq n$. Although $\mathrm{T}\ulcorner A \urcorner, B$ is luckily obtained directly from A by T-Intro, the problem here is that we need to get the required sequent $\mathrm{T}\ulcorner A \urcorner, B$ in a computable way. If we want to give a recursive procedure, then the operation at each step has to be uniquely determined without trial and error.

To resolve this problem, I will present F^{∞} as a *labelled* sequent system as follows:

$$\dfrac{\dfrac{\vdash n : A}{\vdash n+1 : \mathrm{T}\ulcorner A \urcorner, m : \mathrm{T}\ulcorner \mathrm{T}\ulcorner B \urcorner \urcorner}\ \textsf{T-Intro}}{\vdash n+1 : \mathrm{T}\ulcorner A \urcorner, m : \mathrm{T}\ulcorner B \urcorner}\ \textsf{Del}$$

Informally speaking, this system assigns T-Intro rank to each formula in the sequent.

To give a sketch of the proof of strong disquotability, I here explain the derivation system in more detail. If the rule Del is not used, then the system is a subsystem of G^{∞}, and thus, it is easy to show the strong disquotabiliy of the system. Thus, my strategy is to eliminate applications of Del. Since we consider a labelled system, the rule Del is now indexed:

$$\dfrac{\Gamma, n : \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner}{\Gamma, n : \mathrm{T}\ulcorner A \urcorner}\ \textsf{Del}_n.$$

Therefore, we will eliminate applications of $\textsf{Del}_n$ by induction on n. For this purpose, it is useful to keep the information about what indices for Del are used in the derivation.

For the above reasons, we shall present derivations as the following form:

$$\vdash^{\alpha}_{S,k} \Gamma,$$

where Γ consists of labelled sentences of the form $n : A$; a natural number k is the cut-rank; an ordinal α is the derivation length; and a finite set S of natural numbers indicates that $\textsf{Del}_n$ may occur in the derivation only if $n \in S$ or T-Intro is applied below Del_n. [2] For example, the following is a correct derivation:

$$\dfrac{\dfrac{\dfrac{\dfrac{\vdots}{\vdash^{\alpha}_{\{2\},k} 2 : \mathrm{T}\ulcorner \mathrm{T}\ulcorner A\urcorner\urcorner}}{\vdash^{\alpha+1}_{\{2\},k} 2 : \mathrm{T}\ulcorner A\urcorner}\ \mathsf{Del}_2}{\vdash^{\alpha+2}_{\{3,5\},k} 3 : \mathrm{T}\ulcorner \mathrm{T}\ulcorner A\urcorner\urcorner}\ \text{T-}\mathsf{Intro}}{\vdash^{\alpha+3}_{\{3,5\},k} 3 : \mathrm{T}\ulcorner A\urcorner}\ \mathsf{Del}_3$$

In this derivation, although $2 \notin \{3, 5\}$, the rule Del_2 can occur above T-Intro, because $2 \in \{2\}$.

Let A be an arithmetical sentence. The upper-bound proof of F proceeds as follows:

1. $\mathsf{F} \vdash A$
2. $\mathsf{F}^{\infty}_{rec} \vdash^{\alpha}_{S,k} n : A$, for some $S \subseteq_{\text{fin}} \mathbb{N}$, some $n, k \in \mathbb{N}$, and some $\alpha < \varphi\omega 0$
   ($\because$ Embedding Lemma (Lemma 62))
3. $\mathsf{F}^{\infty}_{rec} \vdash^{\omega_k(\alpha)}_{S,0} n : A$
   ($\because$ Cut elimination (Lemma 57))
4. $\mathsf{PA} + \mathsf{TI}(< \varphi\omega 0) \vdash A$
   ($\because$ formalising the above argument in $\mathsf{PA} + \mathsf{TI}(< \varphi\omega 0)$)

Thus, we have $\mathsf{F} \leq \mathsf{PA} + \mathsf{TI}(< \varphi\omega 0)$, as desired.

To show Embedding Lemma, the admissibility of T-Elim is proved as follows:

1. $\mathsf{F}^{\infty} \vdash^{\alpha}_{S,0} n : \mathrm{T}\ulcorner A\urcorner$
2. $\mathsf{F}^{\infty} \vdash^{\varphi(n+1)\alpha}_{S\setminus\{0,1,\cdots,n\},0} n : \mathrm{T}\ulcorner A\urcorner$
   ($\because$ Del-Elimination (Lemma 60), which states that $n : \mathrm{T}\ulcorner A\urcorner$ is derivable without using $\mathsf{Del}_0, \mathsf{Del}_1, \cdots, \mathsf{Del}_n$)
3. $\mathsf{F}^{\infty} \vdash^{\varphi 1(\varphi(n+1)\alpha)}_{S\setminus\{n\},0} n - 1 : A$
   ($\because$ Disquotation (Lemma 59), which states that we can disquote $n : \mathrm{T}\ulcorner A\urcorner$ without using Del_0, Del_1, $\cdots$, Del_{n-1}.)

Definition 53. (Labelled sequents) A *labelled sequent* Γ is a finite set of labelled $\mathcal{L}_{\mathrm{T}}$-sentences of the form $n : A$, where $n \in \mathbb{N}$. For a normal sequent Δ, we define a labelled sequent $n : \Delta$ as $\{n : A \mid A \in \Delta\}$.

[2] Here, we use the notion of *finite sets* of natural numbers. Nevertheless, concepts and operations on finite sets can be formalised in weak arithmetic, such as PA. Thus, they are not problematic for the proof-theoretic purpose. For more details, see [27].

Definition 54. (System F^∞) The labelled derivation system F^∞ comprises the labelled version of PA^∞ in Table 3.5 and the truth rules in Table 3.6. The system F^∞_{rec} is a subsystem of F^∞ in which only recursive derivations are allowed.

Table 3.5: Labelled PA^∞

Basic axioms

($\mathsf{Ax.1}$) $\vdash^{\alpha}_{S,k} \Gamma, n : A$, if A is a true arithmetical literal.
($\mathsf{Ax.2}$) $\vdash^{\alpha}_{S,k} \Gamma, n : \neg \mathrm{T}s, n : \mathrm{T}t$, if $s^{\mathbb{N}} = t^{\mathbb{N}}$.

Basic rules (assume: $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : A_0, n : A_1}{\vdash^{\alpha}_{S,k} \Gamma, n : A_0 \vee A_1}\ (\vee) \qquad \frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : A_0 \quad \vdash^{\alpha_1}_{S,k} \Gamma, n : A_1}{\vdash^{\alpha}_{S,k} \Gamma, n : A_0 \wedge A_1}\ (\wedge)$$

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : A(i)}{\vdash^{\alpha}_{S,k} \Gamma, n : \exists x A}\ (\exists) \qquad \frac{\ldots \vdash^{\alpha_i}_{S,k} \Gamma, n : A(i) \ldots \ (i \in \mathbb{N})}{\vdash^{\alpha}_{S,k} \Gamma, n : \forall x A}\ (\forall)$$

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : A \quad \vdash^{\alpha_1}_{S,k} \Gamma, n : \neg A}{\vdash^{\alpha}_{S,k} \Gamma}\ (\mathsf{cut}), \text{ where } |A| < k.$$

First, we remark that the basic properties hold even in the labelled system. In particular, we have three kinds of Weakening for $\mathsf{F}^\infty_{(rec)}$.

Lemma 55. (Weakening) The following three types of Weakening hold in $\mathsf{F}^\infty_{(rec)}$.

1. Assume $\vdash^{\alpha}_{S,k} \Gamma$ and take arbitrary $S' \supseteq S, k' \geq k$, and $\alpha' \geq \alpha$. Then, it holds that $\vdash^{\alpha'}_{S',k'} \Gamma, n : A$.
2. Assume $\vdash^{\alpha}_{S,k} n : \Gamma$ and take arbitrary $n' \geq n$. If $n' \in S$, then it holds that $\vdash^{\alpha}_{S,k} n' : \Gamma$.
3. Assume $\vdash^{\alpha}_{S,k} \Gamma, n : A$. If $n \notin S$ and $n' \geq n$, then it holds that $\vdash^{\alpha}_{S,k} \Gamma, n' : A$.

Lemma 56. The following basic properties hold in $\mathrm{F}^\infty_{(rec)}$.

Table 3.6: System F^∞

Truth rules (assume: $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$$\frac{\vdash^{\alpha_0}_{S_0,k} n_0 : A}{\vdash^{\alpha}_{S,k} \Gamma, n : \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{T\text{-}Intro}), \text{ for any } n > n_0 \text{ and any } S \subseteq_{\text{fin}} \mathbb{N}.$$

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : \mathrm{T}\ulcorner A \urcorner \quad \vdash^{\alpha_1}_{S,k} \Gamma, n : \mathrm{T}\ulcorner A \to B \urcorner}{\vdash^{\alpha}_{S,k} \Gamma, n : \mathrm{T}\ulcorner B \urcorner}\ (\mathsf{T}_{\to})$$

$$\frac{\ldots \vdash^{\alpha_i}_{S,k} \Gamma, n : \mathrm{T}\ulcorner A(i) \urcorner \ldots \ (i \in \mathbb{N})}{\vdash^{\alpha}_{S,k} \Gamma, n : \mathrm{T}\ulcorner \forall x A \urcorner}\ (\mathsf{T}_{\forall})$$

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner}{\vdash^{\alpha}_{S,k} \Gamma, n : \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{Del}_n), \text{ where } n \in S.$$

1. ($\wedge$-Inversion) If $\vdash^{\alpha}_{S,k} \Gamma, n : A_0 \wedge A_1$, then $\vdash^{\alpha}_{S,k} \Gamma, n : A_i$ for every $i \leq 1$.
2. ($\forall$-Inversion) If $\vdash^{\alpha}_{S,k} \Gamma, n : \forall x A(x)$, then $\vdash^{\alpha}_{S,k} \Gamma, n : A(i)$ for every $i \in \mathbb{N}$.
3. ($\vee$-Exportation) If $\vdash^{\alpha}_{S,k} \Gamma, n : A \vee B$, then $\vdash^{\alpha}_{S,k} \Gamma, n : A, n : B$.
4. (False Literal) If $\vdash^{\alpha}_{S,k} \Gamma, n : A$ for a false $\mathcal{L}$-literal A, then $\vdash^{\alpha}_{S,k} \Gamma$.
5. (Substitution) If $\vdash^{\alpha}_{S,k} \Gamma, n : A(s)$ and $s^{\mathbb{N}} = t^{\mathbb{N}}$, then $\vdash^{\alpha}_{S,k} \Gamma, n : A(t)$.

Lemma 57. (Cut Elimination) For any $n \in \mathbb{N}$, the following is admissible:

$$\frac{\vdash^{\alpha}_{S,0} \Gamma, n : A \quad \vdash^{\beta}_{S,0} \Delta, n : \neg A}{\vdash^{\omega_{|A|}(\alpha \# \beta)}_{S,0} \Gamma, \Delta}.$$

Definition 58. (Labelled basic T-positive sequent) A *labelled basic* T-*positive* sequent is of the following form:

$$n_0 : \mathrm{T}\ulcorner A_0 \urcorner, n_1 : \mathrm{T}\ulcorner A_1 \urcorner, \cdots, n_m : \mathrm{T}\ulcorner A_m \urcorner.$$

For a sequent $\Gamma := \{\mathrm{T}\ulcorner A_0 \urcorner, \mathrm{T}\ulcorner A_1 \urcorner, \cdots, \mathrm{T}\ulcorner A_m \urcorner\}$, recall that the *disquotation* $Disq(\Gamma)$ of Γ is the sequent $\{A_0, A_1, \cdots, A_m\}$. We also define a labelled sequent $n : Disq(\Gamma)$ as $\{n : A \mid A \in Disq(\Gamma)\}$.

Lemma 59. (Disquotation for $\mathrm{F}^{\infty}_{(rec)}$) Assume $0 < n \notin S$; let $\Gamma, n : \Delta$ be a labelled basic T-positive sequent; define a set $S' := S \cup \{0, 1, \ldots, n-1\}$. If $\vdash^{\alpha}_{S,0} \Gamma, n : \Delta$, then we have $\vdash^{\varphi 1 \alpha}_{S',0} \Gamma, n-1 : Disq(\Delta)$.

Proof. By induction on α. We remark that because $0 < n \notin S$, the labelled sequent $\Gamma, n : \Delta$ is obtained by a rule other than Del_n.

As an example, we consider the following case:

$$\dfrac{\vdash^{\alpha_0}_{S_0,0} m : A}{\vdash^{\alpha}_{S,0} \Gamma, n : \Delta', n : \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{T\text{-}Intro}),$$

where $\Delta = \Delta' \cup \{\mathrm{T}\ulcorner A \urcorner\}$ and $m < n$.

Here, S_0 may contain natural numbers larger than m, but it is easy to show that $\vdash^{\alpha_0}_{\{0,1,\cdots,m\},0} m : A$. Therefore, by item 3 of Weakening, it follows that $\vdash^{\varphi 1 \alpha}_{S',0} \Gamma, n-1 : Disq(\Delta)$. ■

Lemma 60. (Del-Elimination for $\mathrm{F}^{\infty}_{(rec)}$) Assume that Γ is a labelled basic T-positive sequent.

1. If $\vdash^{\alpha}_{S,0} \Gamma, 0 : \mathrm{T}\ulcorner A \urcorner$, then $\vdash^{\alpha}_{S,0} \Gamma$.
2. If $\vdash^{\alpha}_{S,0} \Gamma$, then $\vdash^{\varphi(n+1)\alpha}_{S \setminus \{0,1,\cdots,n\},0} \Gamma$.

Proof.

1. For a labelled basic T-positive sequent Γ, let Γ' be the subset of Γ that consists of all and only sentences with a non-zero label. By induction on α, we can show the claim: for every labelled T-positive sequent Γ, if $\vdash^{\alpha}_{S,0} \Gamma$, then $\vdash^{\alpha}_{S,0} \Gamma'$. Therefore, the required claim follows.
2. By induction on n and α. If $n = 0$, from item 1, we obtain $\vdash^{\alpha}_{S,0} \Gamma'$, where Γ' is the result of eliminating from Γ all sentences with the label 0. As long as we only see below the applications of $\mathsf{T\text{-}Intro}$, the rule Del_0 is not used in the derivation $\vdash^{\alpha}_{S,0} \Gamma'$. Therefore, we get $\vdash^{\alpha}_{S \setminus \{0\},0} \Gamma'$. Thus, by item 1 of Weakening Lemma (Lemma 55), we have $\vdash^{\varphi 1 \alpha}_{S \setminus \{0\},0} \Gamma$.
   Now we assume $n > 0$. As in the important case, we consider the situation where $\vdash^{\alpha}_{S,0} \Gamma$ is obtained by (Del_n):

$$\dfrac{\vdash^{\alpha_0}_{S,0} \Gamma', n : \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner \urcorner}{\vdash^{\alpha}_{S,0} \Gamma', n : \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{Del}_n),$$

where $\Gamma = \Gamma' \cup \{n : \mathrm{T}\ulcorner A \urcorner\}$.
We can assume $\{0, 1, \cdots, n-1\} \subseteq S$. Then, by the sub IH, we have

$$\vdash^{\varphi(n+1)\alpha_0}_{S\setminus\{0,1,\cdots,n\},0} \Gamma', n : \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner.$$

Thus, Disquotation Lemma (Lemma 59) yields that

$$\vdash^{\varphi 1(\varphi(n+1)\alpha_0)}_{S\setminus\{n\},0} \Gamma', n-1 : \mathrm{T}\ulcorner A \urcorner.$$

Therefore, from the main IH, we obtain:

$$\vdash^{\varphi n(\varphi 1(\varphi(n+1)\alpha_0))}_{S\setminus\{0,1,\cdots,n\},0} \Gamma', n-1 : \mathrm{T}\ulcorner A \urcorner.$$

As $\varphi n(\varphi 1(\varphi(n+1)\alpha_0)) = \varphi(n+1)\alpha_0 < \varphi(n+1)\alpha$, it follows by item 3 of Weakening (Lemma 55) that $\vdash^{\varphi(n+1)\alpha}_{S\setminus\{0,1,\cdots,n\},0} \Gamma$, as required.

■

As I explained above, we now immediately have the upper bound of F.

Lemma 61. (Admissibility of T-Elim)
If $\mathsf{F}^{\infty}_{(rec)} \vdash^{\alpha}_{S,0} n : \mathrm{T}\ulcorner A \urcorner$, then $\mathsf{F}^{\infty}_{(rec)} \vdash^{\varphi(n+1)\alpha}_{S\cup\{0,1,\cdots,n-1\},0} n-1 : A$.

Proof. If $\vdash^{\alpha}_{S,0} n : \mathrm{T}\ulcorner A \urcorner$, then n cannot be 0. By Del-Elimination (Lemma 60), we have $\vdash^{\varphi(n+1)\alpha}_{S\setminus\{0,1,\cdots,n\},0} n : \mathrm{T}\ulcorner A \urcorner$. Thus, Disquotation (Lemma 59) implies

$$\vdash^{\varphi 1(\varphi(n+1)\alpha)}_{S\cup\{0,1,\cdots,n-1\},0} n-1 : A.$$

Since $n+1 > 1$, we have $\varphi 1(\varphi(n+1)\alpha) = \varphi(n+1)\alpha$. ■

Lemma 62. (Embedding of F) If $\mathsf{F} \vdash A$, then we have $\mathsf{F}^{\infty}_{rec} \vdash^{\alpha}_{\{0,1,\cdots,n\},0} n : A$ for some n and α.

Theorem 63. (Upper-bound of F) Let A be an $\mathcal{L}$-sentence.
If $\mathsf{F} \vdash A$, then $\mathsf{PA} + \mathsf{TI}(< \varphi\omega 0) \vdash A$.

Proof. The proof is by formalising the above lemmata. In particular, given a derivation of $\mathrm{T}\ulcorner A \urcorner$ in $\mathsf{F}^{\infty}_{rec}$, the above proof gives a particular derivation of A in $\mathsf{F}^{\infty}_{rec}$. ■

3.2.4 An Analysis of E

Finally, we consider the system E. Note that Leigh and Rathjen's proof faces the same problem as F (Proposition 52), so we need to modify the proof. Since E has (T-Intro) and (Del), we again need a labelled system for E, similar to $\mathsf{F}^{\infty}_{(rec)}$. Moreover, E has the axiom (T-Cons), which makes the analysis of E more complicated.

Definition 64. (System E [50]) An $\mathcal{L}_{\mathrm{T}}$-theory E consists of $\mathsf{Base}_{\mathrm{T}}$ with the following axioms and rules:

(Del) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner \mathrm{T}\ulcorner A \urcorner\urcorner \to \mathrm{T}\ulcorner A \urcorner$;

($\mathrm{T}_{\forall}$) $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner \to \mathrm{T}\ulcorner \forall x A(x) \urcorner$;

(T-Cons) $\neg \mathrm{T}\ulcorner \bot \urcorner$;

(T-Intro) from A infer $\mathrm{T}\ulcorner A \urcorner$;

(T-Elim) from $\mathrm{T}\ulcorner A \urcorner$ infer A.

Similar to $\mathsf{F}^{\infty}_{(rec)}$, we formulate $\mathsf{E}^{\infty}_{(rec)}$ as a labelled system. However, due to the axiom (T-Cons), we impose a condition on labels in the rules (cut) and (T-Cons).

Let n be a natural number and Γ be a labelled sequent. We write $n \in \Gamma$, when there exists a sentence A such that $n : A \in \Gamma$.

Definition 65. (System $\mathsf{E}^{\infty}_{(rec)}$) The labelled system E^{∞} is almost the same as F^{∞}, though the cut rule is now restricted to the following form:

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : A \qquad \vdash^{\alpha_1}_{S,k} \Gamma, n : \neg A}{\vdash^{\alpha}_{S,k} \Gamma}\ (\mathsf{cut}),\ \text{where } \alpha_0, \alpha_1 < \alpha,\ |A| < k \text{ and } n \in \Gamma.$$

In addition, E^{∞} has the rule (T-Cons) :

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : \mathrm{T}\ulcorner \bot \urcorner}{\vdash^{\alpha}_{S,k} \Gamma}\ (\text{T-}\mathsf{Cons}),\ \text{where } \alpha_0 < \alpha \text{ and } n \in \Gamma.$$

Similar to $\mathsf{F}^{\infty}_{(rec)}$, we observe that all of the basic properties (Lemma 56), Weakening (Lemma 55), Cut Elimination, Disquotation, Admissibility of T-Elim, and Embedding Lemma hold in $\mathsf{E}^{\infty}_{(rec)}$. We restate the claims:

Lemma 66. (Weakening for $\mathsf{E}^{\infty}_{(rec)}$) 1. Assume $\vdash^{\alpha}_{S,k} \Gamma$ and take arbitrary $S' \supseteq S$, $k' \geq k$, and $\alpha' \geq \alpha$. Then, it holds that $\vdash^{\alpha'}_{S',k'} \Gamma, n : A$.

2. Assume $\vdash^{\alpha}_{S,k} n : \Gamma$ and take arbitrary $n' \geq n$. If $n' \in S$, then it holds that $\vdash^{\alpha}_{S,k} n' : \Gamma$.
3. Assume $\vdash^{\alpha}_{S,k} \Gamma, n : A$. If $n \notin S$ and $n' \geq n$, then it holds that $\vdash^{\alpha}_{S,k} \Gamma, n' : A$.

Proof. In the same way as in the proof of Lemma 55, we show each claim by induction on α. Thus, it is enough to check the cases where (T-Cons) is related.

1. Taking arbitrary $S' \supseteq S, k' \geq k$, and $\alpha' \geq \alpha$, we assume that the derivation of Γ is of the following form:

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, m : \mathrm{T}\ulcorner\bot\urcorner}{\vdash^{\alpha}_{S,k} \Gamma}\ (\text{T-}\mathsf{Cons}),$$

   where $m \in \Gamma$.
   By the induction hypothesis, we have $\vdash^{\alpha_0}_{S',k'} \Gamma, m : \mathrm{T}\ulcorner\bot\urcorner, n : A$. Since $\alpha_0 < \alpha'$ and $m \in \Gamma \subseteq \Gamma \cup \{n : A\}$, we obtain the claim $\vdash^{\alpha'}_{S',k'} \Gamma, n : A$ by (T-Cons).
2. Similar to item 1.
3. We assume that the derivation of $\Gamma, n : A$ is of the following form:

$$\frac{\vdash^{\alpha_0}_{S,k} \Gamma, n : A, m : \mathrm{T}\ulcorner\bot\urcorner}{\vdash^{\alpha}_{S,k} \Gamma, n : A}\ (\text{T-}\mathsf{Cons}),$$

   where $m \in \Gamma$ or $m = n$.
   We also suppose $n \notin S$ and $n' \geq n$; then, we need to show $\vdash^{\alpha}_{S,k} \Gamma, n' : A$.
   If $m = n$, then by applying the induction hypothesis twice, we have $\vdash^{\alpha_0}_{S,k} \Gamma, n' : A, n' : \mathrm{T}\ulcorner\bot\urcorner$. Since $n' \in \Gamma \cup \{n' : A\}$, we obtain by (T-$\mathsf{Cons}$) that $\vdash^{\alpha}_{S,k} \Gamma, n' : A$, as desired.
   If $m \in \Gamma$, then we apply the induction hypothesis to get $\vdash^{\alpha_0}_{S,k} \Gamma, n' : A, m : \mathrm{T}\ulcorner\bot\urcorner$. Since $m \in \Gamma$, we have $\vdash^{\alpha_0}_{S,k} \Gamma, n' : A$ by (T-Cons).

■

Lemma 67. (Cut Elimination for $\mathsf{E}^{\infty}_{(rec)}$) For any $n \in \mathbb{N}$, the following is admissible:

$$\frac{\vdash^{\alpha}_{S,0} \Gamma, n : A \quad \vdash^{\beta}_{S,0} \Delta, n : \neg A}{\vdash^{\omega_{|A|}(\alpha\#\beta)}_{S,0} \Gamma, \Delta},\ \text{where } n \in \Gamma, \Delta.$$

Proof. Except for the cases where new rules are used, the proof is almost the same as for Lemma 57. Therefore, we consider only the case where (T-Cons) is concerned. The case of (cut) is similarly treated.

We assume that the left premise is derived by (T-Cons):

$$\frac{\vdash^{\alpha_0}_{S,0} \Gamma, n : A, m : \mathrm{T}\ulcorner\bot\urcorner}{\vdash^{\alpha}_{S,0} \Gamma, n : A} \text{(T-Cons)}$$

where $n \in \Gamma, \Delta$ and $m \in \Gamma \cup \{n : A\}$.

By the induction hypothesis, we have $\vdash^{\omega_{|A|}(\alpha_0 \# \beta)}_{S,0} \Gamma, m : \mathrm{T}\ulcorner\bot\urcorner, \Delta$. Since $m \in \Gamma, \Delta$, it follows by (T-Cons) that $\vdash^{\omega_{|A|}(\alpha \# \beta)}_{S,0} \Gamma, \Delta$. ■

To analyse $\mathsf{E}^{\infty}_{(rec)}$, we slightly modify the definition of labelled basic T-positivity.

Definition 68. (Labelled basic T-positive sequent for E) A labelled sequent is *basic* T-*positive* if it consists of either one of the following forms:

- $n : \mathrm{T}\ulcorner A \urcorner$,
- $n : B$ for an $\mathcal{L}$-sentence B.

We take a sequent $\Gamma := \{\mathrm{T}\ulcorner A_0\urcorner, \mathrm{T}\ulcorner A_1\urcorner, \cdots, \mathrm{T}\ulcorner A_m\urcorner, B_0, \cdots B_l\}$ for $\mathcal{L}$-sentences $B_0, \cdots, B_l$. Let the *disquotation* $Disq(\Gamma)$ of Γ be the set $\{A_0, A_1, \cdots, A_m, B_0, \cdots B_l\}$. Also, we define a labelled sequent $n : Disq(\Gamma) := \{n : A \mid A \in Disq(\Gamma)\}$.

Lemma 69. (Disquotation for $\mathsf{E}^{\infty}_{(rec)}$) Assume $0 < n \notin S$; let a labelled sequent $\Gamma, n : \Delta$ be basic T-positive; define a set $S' := S \cup \{0, 1, \ldots, n-1\}$. If $\vdash^{\alpha}_{S,0} \Gamma, n : \Delta$, then we have $\vdash^{\varphi 1 \alpha}_{S',0} \Gamma, n-1 : Disq(\Delta)$.

Proof. The proof is almost the same as for Lemma 59.

(T-Intro) Recall the case of (T-Intro):

$$\frac{\vdash^{\alpha_0}_{S_0,0} m : A}{\vdash^{\alpha}_{S,0} \Gamma, n : \Delta', n : \mathrm{T}\ulcorner A\urcorner} \text{(T-Intro)},$$

where $\Delta = \Delta' \cup \{\mathrm{T}\ulcorner A\urcorner\}$ and $m < n$.

Owing to the condition on (T-Cons), we can show $\vdash^{\alpha_0}_{\{0,1,\cdots,m\},0} m : A$, similar to the case of F^{∞}.[3] Thus, by Weakening, we have $\vdash^{\varphi 1 \alpha}_{S',0} \Gamma, n-1 : Disq(\Delta)$.

[3]Conversely, if it were not for the condition, the indices of (T-Cons) used in the derivation might be arbitrarily large, and thus, it could not be assured that $\vdash^{\alpha_0}_{\{0,1,\cdots,m\},0} m : A$.

(T-Cons) we next consider the case where $\vdash^{\alpha}_{S,0} \Gamma, n : \Delta$ is derived by (T-Cons):

$$\dfrac{\vdash^{\alpha_0}_{S,0} \Gamma, n : \Delta, m : \mathrm{T}\ulcorner\bot\urcorner}{\vdash^{\alpha}_{S,0} \Gamma, n : \Delta}\ (\text{T-}\mathsf{Cons}),$$

where $m \in \Gamma$ or $m = n$.

If $m \in \Gamma$, the claim immediately follows by the induction hypothesis. Thus, we can assume $m = n$. By the induction hypothesis, we obtain:

$$\vdash^{\varphi 1 \alpha_0}_{S',0} \Gamma, n-1 : Disq(\Delta), n-1 : \bot.$$

By False Literal and item 1 of Weakening Lemma, we also have the required claim:

$$\vdash^{\varphi 1 \alpha}_{S',0} \Gamma, n-1 : Disq(\Delta).$$

■

Lemma 70. (Del-Elimination for $\mathsf{E}^{\infty}_{(rec)}$) Assume that a labelled sequent Γ is basic T-positive.

1. If $\vdash^{\alpha}_{S,0} \Gamma, 0 : \mathrm{T}\ulcorner A\urcorner$, then $\vdash^{\alpha}_{S,0} \Gamma$.
2. If $\vdash^{\alpha}_{S,0} \Gamma$, then $\vdash^{\varphi(n+1)\alpha}_{S\setminus\{0,1,\cdots,n\},0} \Gamma$.

Proof. The proof is exactly the same as for Lemma 60. ■

Lemma 71. (Admissibility of T-Elim for $\mathsf{E}^{\infty}_{(rec)}$) If $\vdash^{\alpha}_{S,0} n : \mathrm{T}\ulcorner A\urcorner$, then $\vdash^{\varphi(n+1)\alpha}_{S\cup\{0,1,\cdots,n-1\},0}$ $n-1 : A$.

Lemma 72. (Embedding for $\mathsf{E}^{\infty}_{(rec)}$) If $\mathsf{E} \vdash A$, then for some n and α, we have $\mathsf{E}^{\infty}_{(rec)} \vdash^{\alpha}_{\{0,1,\cdots,n\},0} n : A$.

Proof. The proof is almost the same as for Lemma 62. We only have to care about the additional conditions on (cut) and (T-Cons). For example, we treat the case of Modus Ponens.

Assume that $\mathsf{E} \vdash C$ is derived from the premises $\mathsf{E} \vdash B$ and $\mathsf{E} \vdash B \to C$. By the induction hypothesis, we have

$$\mathsf{E}^{\infty}_{(rec)} \vdash^{\alpha}_{\{0,1,\cdots,n\},0} n : B \quad \text{and} \quad \mathsf{E}^{\infty}_{(rec)} \vdash^{\alpha'}_{\{0,1,\cdots,n'\},0} n' : B \to C.$$

Let $m := \max(n, n')$ and $\beta := \max(\alpha, \alpha')$. Then, by Weakening, we obtain

$$\mathsf{E}^{\infty}_{(rec)} \left|\frac{\beta}{\{0,1,\cdots,m\},0}\right. m : B \quad \text{and} \quad \mathsf{E}^{\infty}_{(rec)} \left|\frac{\beta}{\{0,1,\cdots,m\},0}\right. m : \neg B, m : C.$$

Therefore, Cut Elimination yields

$$\mathsf{E}^{\infty}_{(rec)} \left|\frac{\omega_{|B|}(\beta)}{\{0,1,\cdots,m\},0}\right. m : C.$$

■

Due to the rule (T-Cons), the proof of Theorem 63 does not work well for E. Nevertheless, thanks to the additional conditions on (cut) and (T-Cons), we can eliminate applications of T-Intro, which makes the analysis of E much easier.

Theorem 73. (Upper Bound of E) Let A be an $\mathcal{L}$-sentence.
If $\mathsf{E} \vdash A$, then $\mathsf{PA} + \mathsf{TI}(< \varphi\omega 0) \vdash A$.

Proof. Assume that an $\mathcal{L}$-sentence A is derived in E, then by Embedding Lemma (Lemma 72), we have $\mathsf{E}^{\infty}_{rec} \left|\frac{\alpha}{\{0,1,\cdots,n\},0}\right. n : A$, for some n and α. Since A, as a sequent, is basic T-positive, it follows by Del-elimination (Lemma 70) that $\mathsf{E}^{\infty}_{rec} \left|\frac{\varphi(n+1)\alpha}{\emptyset,0}\right. n : A$. Similarly, by Disquotation Lemma, we obtain $\mathsf{E}^{\infty}_{rec} \left|\frac{\varphi 1(\varphi(n+1)\alpha)}{\{0,1,\cdots,n-1\},0}\right. n-1 : A$. Therefore, iterating this process yields $\mathsf{E}^{\infty}_{rec} \left|\frac{\beta}{\emptyset,0}\right. 0 : A$ for some $\beta < \varphi\omega 0$. However, by the condition on (T-Cons), this means that A is derived without (Del) nor (T-Intro), and thus, this derivation can be seen as one in, e.g., the system $\mathsf{G}^{\infty}_{rec}$. Therefore, similar to Theorem 34, this is formalisable in $\mathsf{PA} + \mathsf{TI}(< \varphi\omega 0)$. In conclusion, we obtain $\mathsf{PA} + \mathsf{TI}(< \varphi\omega 0) \vdash A$. ■

3.3 Upper Bound of FS

In Section 3.2, we observed that the truth disjunction property does not hold in any of $\mathsf{G}^{\infty}_{rec}$, $\mathsf{I}^{\infty}_{rec}$, $\mathsf{F}^{\infty}_{rec}$, and $\mathsf{E}^{\infty}_{rec}$, even when considering only $\mathcal{L}$-sentences:

Truth disjunction property. If $\vdash \mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner B \urcorner$, then $\vdash \mathrm{T}\ulcorner A \urcorner$ or $\vdash \mathrm{T}\ulcorner B \urcorner$.

In this section, we show that the situation is even more restrictive in the presence of the axiom $\mathsf{T\text{-}Comp}$: $\mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner \neg A \urcorner$. More precisely, taking FS as an example, we prove that the infinitary system FS^{∞} for FS does not satisfy the truth disjunction property, whether it is recursive or not (Remark 84). On the other hand, FS is shown to satisfy *disquotability*, a property weaker than strong disquotability. Using this property, we give a direct upper-bound proof of FS.

In this section, we adopt Definition 5 as the definition of FS.

Definition 74. (System FS [50]) An $\mathcal{L}_{\mathrm{T}}$-theory FS consists of $\mathsf{Base}_{\mathrm{T}}$ with the following axioms and rules:

($\mathrm{T}_\forall$) $\forall \ulcorner A_x \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \forall v \mathrm{T}\ulcorner A(\dot{v})\urcorner \to \mathrm{T}\ulcorner \forall x A(x)\urcorner$;

($\mathsf{T\text{-}Cons}$) $\neg \mathrm{T}\ulcorner \bot \urcorner$;

($\mathsf{T\text{-}Comp}$) $\forall \ulcorner A \urcorner \in \mathrm{Sent}_{\mathcal{L}_{\mathrm{T}}}.\ \mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner \neg A \urcorner$;

($\mathsf{T\text{-}Intro}$) from A infer $\mathrm{T}\ulcorner A \urcorner$;

($\mathsf{T\text{-}Elim}$) from $\mathrm{T}\ulcorner A \urcorner$ infer A.

Definition 75. (System FS^∞) The (non-labelled) derivation system FS^∞ consists of PA^∞ with Truth axiom and Truth rules in Table 3.7. The system FS^∞_{rec} is a subsystem of FS^∞ in which only recursive derivations are allowed.

Table 3.7: System FS^∞

Truth axiom

$$(\mathsf{T\text{-}Comp})\ \vdash^{\alpha}_{k,m} \Gamma, \mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner \neg A \urcorner$$

Truth rules (assume: $\alpha_i < \alpha$ for $i \in \mathbb{N}$)

$$\frac{\vdash^{\alpha_0}_{k,m_0} A}{\vdash^{\alpha}_{k,m} \Gamma, \mathrm{T}\ulcorner A \urcorner}\ (\mathsf{T\text{-}Intro}) \qquad \frac{\vdash^{\alpha_0}_{k,m} \Gamma, \mathrm{T}\ulcorner A \urcorner \quad \vdash^{\alpha_1}_{k,m} \Gamma, \mathrm{T}\ulcorner A \to B \urcorner}{\vdash^{\alpha}_{k,m} \Gamma, \mathrm{T}\ulcorner B \urcorner}\ (\mathsf{T}_{\to})$$

$$\frac{\ldots \vdash^{\alpha_i}_{k,m} \Gamma, \mathrm{T}\ulcorner A(i) \urcorner \ldots \ (i \in \mathbb{N})}{\vdash^{\alpha}_{k,m} \Gamma, \mathrm{T}\ulcorner \forall x A \urcorner}\ (\mathsf{T}_\forall) \qquad \frac{\vdash^{\alpha_0}_{k,m} \Gamma, \mathrm{T}\ulcorner \bot \urcorner}{\vdash^{\alpha}_{k,m} \Gamma}\ (\mathsf{T\text{-}Cons})$$

In $\mathrm{FS}^\infty_{(rec)}$, $\mathsf{Base}_{\mathsf{T}}$, ($\mathsf{T\text{-}Cons}$), ($\mathsf{T\text{-}Comp}$), ($\forall$-$\mathsf{Inf}$), and ($\mathsf{T\text{-}Intro}$) are explicitly covered. On the other hand, ($\exists$-Inf), ($\neg\mathsf{T\text{-}Intro}$), and ($\neg\mathsf{T\text{-}Elim}$) are known to be derivable from the other axioms and rules of D (see [28, p. 316]). Thus, it suffices to show the admissibility of ($\mathsf{T\text{-}Elim}$).

The basic properties in Lemma ref and Cut-Elimination for $\mathrm{FS}^\infty_{(rec)}$ (Lemma 78) hold in the same way as for $\mathrm{G}^\infty_{(rec)}$.

Lemma 76. The following basic properties hold in $\mathsf{FS}^{\infty}_{(rec)}$:

1. (Weakening) If $\vdash^{\alpha}_{k,m} \Gamma$, $\alpha \leq \beta$, $k \leq l$ and $m \leq n$, then $\vdash^{\beta}_{l,n} \Delta$ for $\Gamma \subseteq \Delta$.
2. ($\wedge$-Inversion) If $\vdash^{\alpha}_{k,m} \Gamma, A_0 \wedge A_1$, then $\vdash^{\alpha}_{k,m} \Gamma, A_i$ for every $i \leq 1$.
3. ($\forall$-Inversion) If $\vdash^{\alpha}_{k,m} \Gamma, \forall x A(x)$, then $\vdash^{\alpha}_{k,m} \Gamma, A(\bar{n})$ for every $n \in \mathbb{N}$.
4. ($\vee$-exportation) If $\vdash^{\alpha}_{k,m} \Gamma, A \vee B$, then $\vdash^{\alpha}_{k,m} \Gamma, A, B$.
5. (False Literal) If $\vdash^{\alpha}_{k,m} \Gamma, A$ for a false $\mathcal{L}$-literal A, then $\vdash^{\alpha}_{k,m} \Gamma$.
6. (Substitution) If $\vdash^{\alpha}_{k,m} \Gamma, A(s)$ and $s^{\mathbb{N}} = t^{\mathbb{N}}$, then $\vdash^{\alpha}_{k,m} \Gamma, A(t)$.

Lemma 77. (Reduction Lemma for $\mathsf{FS}^{\infty}_{(rec)}$) If $\vdash^{\alpha}_{k,m} \Gamma, A$, $\vdash^{\beta}_{k,m} \Delta, \neg A$ and $|A| = k$, then $\vdash^{\alpha \# \beta}_{k,m} \Gamma, \Delta$.

Proof. The proof is again by induction on $\alpha \# \beta$ and is almost the same as for the systems $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I in Section 3.2. We consider only the case where both A and $\neg A$ are principal and $|A| = k = 0$. The proof is almost the same as that for $\mathsf{G}^{\infty}_{(rec)}$ in Lemma 27. As a new case, we suppose that Γ, A is an instance of (T-Comp). Then, A must be of the form $\mathrm{T}(t)$ and $\Delta, \neg A$ is an instance of (Ax.2), as $\neg\mathrm{T}(t)$ can be principal only in (Ax.2). Therefore, A is contained in Δ, and thus the conclusion follows from $\vdash^{\alpha}_{k,m} \Gamma, A$ by Weakening (Lemma 76). ■

Lemma 78. (Cut Elimination for $\mathsf{FS}^{\infty}_{(rec)}$) If $\mathsf{FS}^{\infty}_{(rec)} \vdash^{\alpha}_{k+1,m} \Gamma$, then $\mathsf{FS}^{\infty}_{(rec)} \vdash^{\omega^{\alpha}}_{k,m} \Gamma$.

To prove Embedding Lemma for FS, the next step is to show the admissibility of ($\mathsf{T\text{-}Elim}$). This task is achieved by proving the disquotability for $\mathrm{FS}^{\infty}_{(rec)}$.

Definition 79. Similar to Definition 68, we define a *basic* T-*positive* sequent for FS to be a sequent Γ of the form $\mathrm{T}\ulcorner A_1 \urcorner$, ..., $\mathrm{T}\ulcorner A_a \urcorner$, B_1, ..., B_b, where B_1, ..., B_b are $\mathcal{L}$-sentences.

For such a Γ, we define its *disquotation* $Disq(\Gamma)$ as the following sequent:

$$A_1, ..., A_a, B_1, ..., B_b.$$

Lemma 80. (Disquotation) Let Γ be a basic T-positive sequent, and $Disq(\Gamma)$ be the disquotation of Γ. Then, for each natural number m and $\alpha \in \mathrm{OT}$, the following hold in $\mathrm{FS}^{\infty}_{(rec)}$:

1. If $\vdash^{\alpha}_{0,m+1} \Gamma$, then $\vdash^{\varepsilon_\alpha}_{0,m} Disq(\Gamma)$.

2. ($\mathsf{T\text{-}Elim}$) If $\vdash^{\alpha}_{0,m+1} \mathrm{T}\ulcorner A \urcorner$, then $\vdash^{\varepsilon_\alpha}_{0,m} A$.

Proof. Item 2 is an immediate consequence of Item 1 because the sequent $\{\mathrm{T}\ulcorner A \urcorner\}$ is itself basic T-positive. Thus, it suffices to prove Item 1. The proof is by induction on α, as with the proof of Strong disquotation for $\mathsf{G}^{\infty}_{(rec)}$ (Lemma 32). We only adress (T-Comp) and (T-Intro).

(T-Comp) Assume that $\vdash^{\alpha}_{0,m+1} \Gamma, \mathrm{T}\ulcorner A \urcorner, \mathrm{T}\ulcorner \neg A \urcorner$ is derived as an instance of (T-Comp). Letting $Disq(\Gamma)$ be the disquotation of Γ, we have $\vdash^{\omega}_{0,0} Disq(\Gamma), A, \neg A$ (cf. Lemma 17). Because $\omega < \varepsilon_\alpha$, the conclusion follows by Weakening (Lemma 76).

(T-Intro) For $\alpha' < \alpha$and $n < m+1$, assume that $\vdash^{\alpha}_{0,m+1} \Gamma, \mathrm{T}\ulcorner A \urcorner$ is derived from $\vdash^{\alpha'}_{0,n} A$. Letting the disquotation of Γ be $Disq(\Gamma)$, then the disquotation of $\Gamma, \mathrm{T}\ulcorner A \urcorner$ is $Disq(\Gamma), A$. Hence, by Weakening (Lemma 76), we directly have $\vdash^{\varepsilon_\alpha}_{0,m} Disq(\Gamma), A$ from the assumption.

∎

Similar to for $\mathsf{G}^{\infty}_{(rec)}$ (Lemma 33), we obtain Embedding Lemma for $\mathsf{FS}^{\infty}_{(rec)}$.

Lemma 81. (Embedding Lemma) For each $\mathcal{L}_{\mathrm{T}}$-formula A and the universal closure $A^{\forall}$, if $\mathsf{FS} \vdash A$, then $\mathsf{FS}^{\infty}_{(rec)} \vdash^{\alpha}_{0,m} A^{\forall}$ for some $\alpha < \varphi 20$ and $m \in \mathbb{N}$.

Since $\mathsf{FS}^{\infty}_{(rec)}$ has ($\mathsf{T\text{-}Cons}$), Cut elimination is not enough to obtain the upper bound of FS. Thus, we next want to eliminate applications of ($\mathsf{T\text{-}Cons}$).

Lemma 82. (T-Cons Elimination) For any natural number m, basic T-positive sequent Δ, and $\mathcal{L}$-sequent Γ, the following holds:

1. If $\mathsf{FS}^{\infty}_{(rec)} \vdash^{\alpha}_{0,0} \Delta$, then $\mathsf{PA}^{\infty}_{(rec)} \vdash^{\varepsilon_\alpha}_{0} Disq(\Delta)$.
2. If $\mathsf{FS}^{\infty}_{(rec)} \vdash^{\alpha}_{0,m} \Gamma$, then $\mathsf{PA}^{\infty}_{(rec)} \vdash^{\varepsilon_{m+1}(\alpha)}_{0} \Gamma$.

Proof.

1. The proof is almost the same as for the item 1 of Lemma 80. In particular, since $\mathsf{T\text{-}Intro}$ is not used in the derivation of Δ, its disquotation $Disq(\Delta)$ is derivable in $\mathsf{PA}^{\infty}_{(rec)}$.
2. We prove the claim by induction of m and α. The case $m = 0$ is a consequence of the item 1. Thus, we can assume $m > 0$. If Γ is an axiom, then that must be an instance of (Ax.1), and thus Γ is derivable in $\mathrm{PA}^{\infty}_{(rec)}$, too. If Γ is derived by one of the rules of $\mathrm{PA}^{\infty}_{(rec)}$, then its premises are also $\mathcal{L}$-sequents.

Therefore, by the sub-induction hypothesis and the same rule, we obtain the required conclusion. Finally, we assume that $\mathrm{FS}^{\infty}_{(rec)} \mid\frac{\alpha}{0,m} \Gamma$ is obtained by (T-Cons) from the premise $\mathsf{FS}^{\infty}_{(rec)} \mid\frac{\alpha_0}{0,m} \Gamma, \mathrm{T}\ulcorner\bot\urcorner$. As $\Gamma, \mathrm{T}\ulcorner\bot\urcorner$ is basic T-positive, Lemma 80 yields that $\mathsf{FS}^{\infty}_{(rec)} \mid\frac{\varepsilon_{\alpha_0}}{0,m-1} \Gamma, \bot$. Here, $\bot$ is eliminable by False literal (Lemma 76), and hence we also have $\mathsf{FS}^{\infty}_{(rec)} \mid\frac{\varepsilon_{\alpha_0}}{0,m-1} \Gamma$. Thus, by the main-induction hypothesis, it follows that $\mathsf{PA}^{\infty}_{(rec)} \mid\frac{\varepsilon_{m+1}(\alpha_0)}{0} \Gamma$, from which we conclude that $\mathsf{PA}^{\infty}_{(rec)} \mid\frac{\varepsilon_{m+1}(\alpha)}{0} \Gamma$.
The other cases are impossible, thus the proof completes.

■

To conclude, we obtain another proof of Halbach's ordinal analysis for FS [28]:

Theorem 83. Any $\mathcal{L}$-sentence provable in FS is a theorem of $\mathsf{PA} + \mathsf{TI}(< \varphi 20)$.

Proof. Assume that an $\mathcal{L}$-sentence A is derived in FS. Then, by Embedding Lemma, we have $\mathrm{FS}^{\infty}_{rec} \mid\frac{\alpha}{0,m} A$ for some $\alpha < \varphi 20$ and $m \in \mathbb{N}$. Thus, by T-Cons Elimination, we have $\mathrm{PA}^{\infty}_{rec} \mid\frac{\varepsilon_{m+1}(\alpha)}{0} A$. Because $\varepsilon_{m+1}(\alpha) < \varphi 20$, this can be formalised in $\mathsf{PA} + \mathsf{TI}(< \varphi 20)$, whence it follows that $\mathsf{PA} + \mathsf{TI}(< \varphi 20) \vdash A$.

■

Remark 84. Finally, we observe that both the truth disjunction property and strong disquotability fail in FS^{∞}. Since FS^{∞} has the rule T-Intro, the truth disjunction property implies strong disquotability. Thus, it suffices to prove the failure of strong disquotability. For the liar sentence λ, we have $\mathsf{FS}^{\infty} \vdash \mathrm{T}\ulcorner\lambda\urcorner \vee \mathrm{T}\ulcorner\neg\lambda\urcorner$ by T-Comp and ($\vee$). Thus, strong disquotability requires the derivability of $\mathrm{T}\ulcorner\lambda\urcorner, \neg\lambda$; and thus, we have $\mathsf{FS}^{\infty} \vdash \mathrm{T}\ulcorner\lambda\urcorner$. As FS^{∞} is closed under T-Elim (Lemma 80), it follows that $\mathsf{FS}^{\infty} \vdash \lambda$. Therefore, FS^{∞} is inconsistent, which contradicts Lemma 82. In conclusion, strong disquotability must fail in FS^{∞}.

Chapter 4
Truth as Classical Realisability

In Section 3.3 of Chapter 3, we observed that the axiom $\mathrm{T}\ulcorner A \urcorner \vee \mathrm{T}\ulcorner \neg A \urcorner$ is in general incompatible with truth-as-provability interpretation. To put it another way, this implies that it is challenging to understand the truth predicate of Tarskian fully compositional theory like FS as provability. Nevertheless, authors (cf. [2, 3, 4, 7, 8, 9, 47, 48, 49]) provide realisability (computational) interpretations for classical principles that ensure the soundness of classical theories such as PA. Therefore, by adopting such a semantic framework as an alternative to Tarskian truth definition, we can expect to obtain fully-compositional axiomatic theories based on realisability, a weak notion of provability.

As a famous semantic framework for classical theories, Krivine formulated *classical realisability* [47, 48, 49], which is a classical version of Kleene's intuitionistic realisability [45]. [1] In the following, we briefly explain Krivine's classical realisability. There are two kinds of syntactic expressions, *terms* t, which represent programs, and *stacks* π, which represent evaluation contexts. A *process* $t \star \pi$ is a pair of a term t and a stack π. In addition, we fix a set $\perp\!\!\!\perp$ (*pole*) of processes that is closed under several evaluation rules for processes. Then, for each sentence A, the set $|A|_{\perp\!\!\!\perp}$ (*realisers*) of terms and the set $\|A\|_{\perp\!\!\!\perp}$ (*refuters*) of stacks are defined inductively such that $t \star \pi \in \perp\!\!\!\perp$ for any $t \in |A|_{\perp\!\!\!\perp}$ and $\pi \in \|A\|_{\perp\!\!\!\perp}$. Here, a term t *universally realises* A if $t \in |A|_{\perp\!\!\!\perp}$ for every pole $\perp\!\!\!\perp$. Then, Krivine proved that each theorem of second-order arithmetic is universally realisable.[2] In particular, we can take the empty set $\emptyset$ as a pole, and it is easy to show that if $|A|_{\emptyset}$ is not empty, then A is true in the standard model $\mathbb{N}$ of arithmetic. In summary, universal realisability implies truth in $\mathbb{N}$; thus, Tarskian truth can be positioned in

[1] Kleene's realisability is regerded as one formalisation of the so-called Brouwer–Heyting–Kolmogorov-interpretation, an intuitionist (constructivist) reading of the logical symbols. For more details, see, e.g., [62]

[2] Moreover, Krivine's classical realisability can be given to other strong theories, such as Zermelo–Fraenkel set theory [47].

Krivine's general framework. In summary, the inclusion relations are as follows:

$$\text{provability in second-order arithmetic} \subset \text{universal realisability} \subset \text{truth in } \mathbb{N}$$

The purpose of this chapter is to axiomatise fully compositional theories for Krivine's classical realisability in a similar manner to CT and FS (see Chapter 2) for Tarskian truth definition. To clarify the relationship with the corresponding truth theories, we formulate realisability theories over PA. As clarified in the above explanation of Krivine's realisability, we require additional vocabularies for a pole $\perp\!\!\!\perp$ and the relations $t \in |A|_{\perp\!\!\!\perp}$ and $\pi \in \|A\|_{\perp\!\!\!\perp}$. With the help of Gödel-numbering, they can be expressed by a unary predicate $x \in \perp\!\!\!\perp$ and binary predicates $x\mathrm{T}y$ and $x\mathrm{F}y$, respectively. Although Krivine's realisability uses λ-terms to express terms, stacks, and processes, we define realisers and refuters as natural numbers, similar to Kleene's number realisability. Since our base theory is PA, this modification can simplify the formulation substantially (Section 4.2.1).

The remainder of this chapter is organised as follows. Section 4.1 is devoted to technical preliminaries required in this chapter. In Section 4.2 we define classical number realisability as a combination of Krivine's classical realisability and Kleene's intuitionistic number realisability. By formalising our realisability, we obtain a first-order theory CR of compositional realisability (Definition 90). In Section 4.3, we observe that our classical number realisability can realise every theorem of PA, which is formalisable in CR (Theorem 96). Then, in Section 4.4, we study the proof-theoretic strength of CR. First, CR is shown to be conservative over PA (Proposition 97). Then, we formulate a kind of reflection principle under which CR essentially amounts to CT (Proposition 100). We also consider a weaker reflection principle that is sound with respect to any pole, and we prove that the principle can make CR as strong as CT (Theorem 103). In Section 4.5, we generalise CR to a Friedman–Sheard-style system FSR for realisability.

The content of this chapter is based on the co-authored paper [33]. The author's contribution is as follows. The author first defined a formal theory of classical realisability, which was more like Krivine's original formulation than the current one. Subsequently, through discussions with the second author of [33], the author of this thesis reformulated it into a form of number realisability, as given in this chapter. In addition, a further improvement of the formulation was given by the second author of [33] (see Remark 91). The proof-theoretic results are given by the author of this thesis through discussions with the second author.

4.1 Preliminaries

This section explains necessary notation and conventions, most of which are identical to those in Chapter 2. The main difference is that only $=$, $\rightarrow$, and $\forall$ are now allowed as logical symbols.

4.1.1 Conventions and Notations

We introduce abbreviations for common formal concepts concerning coding and recursive functions. We denote by $\mathcal{L}$ the first-order language of PA. The logical symbols of $\mathcal{L}$ are $\to$, $\forall$ and $=$. The non-logical symbols are a constant symbol $\underline{0}$ and the function symbols for all primitive recursive functions. In particular, $\mathcal{L}$ has the successor function $x+1$, with which we can define *numerals* $\underline{0}, \underline{1}, \underline{2}, \ldots$. Thus, we identify natural numbers with the corresponding numeral. We also employ the false equation $0 = 1$ as the propositional constant $\bot$ for contradiction, and then the other logical symbols are defined in a standard manner, e.g., $\neg A = A \to \bot$.

The primitive recursive pairing function is denoted $\langle \cdot, \cdot \rangle$ with projection functions $(\cdot)_0$ and $(\cdot)_1$ satisfying $(\langle x, y \rangle)_0 = x$ and $(\langle x, y \rangle)_1 = y$. Sequences are treated as iterated pairing: $\langle x_0, x_1, \ldots, x_k \rangle := \langle x_0, \langle x_1, \ldots, x_k \rangle \rangle$. Based on these constructors, each finite extension of $\mathcal{L}$ is associated a fixed Gödel coding (denoted $\ulcorner e \urcorner$ where e is a finite string of symbols of the extended language) for which the basic syntactic constructions are primitive recursive. In particular, $\mathcal{L}$ contains a binary function symbol *sub* representing the mapping $\ulcorner A(x) \urcorner, n \mapsto \ulcorner A(\underline{n}) \urcorner$ in the case that x is the only free variable of $A(x)$, and binary function symbols $\dot{=}$, $\dot{\to}$ and $\dot{\forall}$ representing, respectively, the operations $\ulcorner s \urcorner, \ulcorner t \urcorner \mapsto \ulcorner s = t \urcorner$, $\ulcorner A \urcorner, \ulcorner B \urcorner \mapsto \ulcorner A \to B \urcorner$ and $\ulcorner x \urcorner, \ulcorner A \urcorner \mapsto \ulcorner \forall x A \urcorner$. These operations will sometimes be omitted and we write $\ulcorner s = t \urcorner$ and $\ulcorner A \to B \urcorner$ for $\dot{=}(\ulcorner s \urcorner, \ulcorner t \urcorner)$ and $\dot{\to}(\ulcorner A \urcorner, \ulcorner B \urcorner)$, etc.

We introduce a number of abbreviations for $\mathcal{L}$-expressions corresponding to common properties or operations on Gödel codes. The property of being the code of a variable is expressed by the formula $\mathit{Var}(x)$, $\mathit{ClTerm}(x)$ denotes the formula expressing that x is a code of a *closed* $\mathcal{L}$-term, and for a fixed extension $\mathcal{L}'$ of $\mathcal{L}$, $\mathit{Sent}_{\mathcal{L}'}(x)$ expresses that x is the code of a sentence. The concepts above are primitive recursively definable meaning that the representing $\mathcal{L}$-formula is simply an equation between $\mathcal{L}$-terms. Given a formula $A(x)$ with at most x is free, $\ulcorner A(\dot{x}) \urcorner$ abbreviates the term $\mathit{sub}(\ulcorner A(x) \urcorner, x)$ expressing the code of the formula $A(\underline{n})$ where n is the value of x. For sequences $\vec{x}$ of variables, $\ulcorner A(\dot{\vec{x}}) \urcorner$ is defined similarly.

Quantification over codes is associated similar abbreviations:

- $\forall \ulcorner A \urcorner \in \mathit{Sent}_{\mathcal{L}'}.\ B(\ulcorner A \urcorner)$ abbreviates $\forall x (\mathit{Sent}_{\mathcal{L}'}(x) \to B(x))$.
- $\forall \ulcorner s \urcorner.\ B(\ulcorner s \urcorner)$ abbreviates $\forall x (\mathit{ClTerm}(x) \to B(x))$.
- $\forall \ulcorner A_v \urcorner \in \mathit{Sent}_{\mathcal{L}'}.\ B(v, \ulcorner A \urcorner)$ abbr. $\forall x \forall v (\mathit{Var}(v) \land \mathit{Sent}_{\mathcal{L}'}(\dot{\forall} v x) \to B(v, x))$, namely quantification relative to codes of formulas with at most one distinguished variable free.

Partial recursive functions can be expressed in $\mathcal{L}$ via the Kleene 'T predicate' method. The ternary relation $x \cdot y \simeq z$ expresses that the result of evaluating the x-th partial recursive function on input y terminates with output z. Note that this relation has a Σ^0_1 definition in PA as a formula $\exists w(T(x,y,w) \wedge (w)_0 = z)$ where T is primitive recursive. It will be notationally convenient to use $x \cdot y$ in place of a *term* (with the obvious interpretation) though use of this abbreviation will be constrained to contexts in which potential for confusion is minimal.

With the above ternary relation we can express the property of two closed terms having equal value, via a Σ^0_1-formula $Eq(y,z)$. That is, $Eq(y,z)$ expresses that y and z are codes of closed $\mathcal{L}$-terms s and t respectively such that $s = t$ is a true equation.

4.1.2 Classical Compositional Truth

Tarskian truth for $\mathcal{L}$ is characterised inductively in the standard manner:

- A closed equation $s = t$ is true iff s and t denote the same value in $\mathbb{N}$;
- $A \to B$ is true iff if A is true, then B is true;
- $\forall x A(x)$ is true iff $A(s)$ is true for all closed terms s.

Quantification over $\mathcal{L}$-sentences and the operations on syntax implicit in the above clauses can be expressed via a Gödel-numbering. Thus, employing a unary (truth) predicate T, a formal system CT can be defined corresponding, in a straightforward manner, to the Tarskian truth clauses.

Definition 85. (CT) For a unary predicate T, let $\mathcal{L}_{\mathrm{T}} = \mathcal{L} \cup \{\mathrm{T}\}$. The $\mathcal{L}_{\mathrm{T}}$-theory CT (compositional truth) consists of PA formulated for the language $\mathcal{L}_{\mathrm{T}}$ plus the following three axioms.

($\mathsf{CT}_=$) $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(s,t)$.

($\mathsf{CT}_\to$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner A \to B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \to \mathrm{T}\ulcorner B \urcorner)$.

($\mathsf{CT}_\forall$) $\forall \ulcorner A_v \urcorner \in Sent_{\mathcal{L}}.\ \mathrm{T}\ulcorner \forall x A(x) \urcorner \leftrightarrow \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$.

Two consequences of the above axioms are of particular relevance. The first is the observation that CT staisfies Tarski's Convention T [61, pp. 187–188] for formulas in $\mathcal{L}$:

Lemma 86. The Tarski-biconditional is derivable in CT for every formula of $\mathcal{L}$. That is, for each formula $A(x_1,\dots,x_k)$ of $\mathcal{L}$ in which only the distinguished variables occur free, we have

$$\mathsf{CT} \vdash \forall x_1,\dots,x_k(\mathrm{T}\ulcorner A(\dot{x}_1,\dots,\dot{x}_k)\urcorner \leftrightarrow A(x_1,\dots,x_k)).$$

Second is the 'term regularity principle' stating that the truth value of each formula depends only on the value of terms and not their 'structure'. Let *subt* be a primitive recursive function function such that $subt\colon \ulcorner A(x)\urcorner, \ulcorner x\urcorner, \ulcorner s\urcorner \mapsto \ulcorner A(s)\urcorner$ for each formula A, variable x and term s.

Lemma 87. Provable in CT is the term regularity principle:

$$\forall\ulcorner s\urcorner\forall\ulcorner t\urcorner\forall\ulcorner A_v\urcorner \in Sent_{\mathcal{L}}.\ Eq(\ulcorner s\urcorner, \ulcorner t\urcorner) \to (\mathrm{T}\ulcorner A(s)\urcorner \to \mathrm{T}\ulcorner A(t)\urcorner)$$

where the term $\ulcorner A(s)\urcorner$ is shorthand for $subt(\ulcorner A\urcorner, v, \ulcorner s\urcorner)$, and $\ulcorner A(t)\urcorner$ likewise.

4.2 Classical Realisability

We present a classical number realisability interpretation for PA and the corresponding axiomatic theory CR.

4.2.1 Classical Number Realisability

We introduce a realisability interpretation for PA based on Krivine's classical realisability. For our setting we require several modifications from Krivine's original definition. The first modification is about realisers. In Krivine's formulation, realisers are essentially lambda terms (cf. [52]). As we seek to formalise realisability over PA, it is natural to assume that realisers are natural numbers, similar to Kleene's number realisability for the intuitionistic arithmetic [45].

Second, we must define the realisability and refutability conditions explicitly for equality. In the language of second-order arithmetic, the equality $a = b$ between a and b is definable by Leibniz equality $\forall X(a \in X \to b \in X)$, which in Krivine's definition, determines the realisability and refutability conditions of the equation uniquely. For a first-order language, equality is a primitive logical symbol and it is a matter of choice what constitutes a refutation of a closed equation. In some sense, we take the most naive approach motivated by Kleene and Krivine's choices: a true equation is refuted by every element of the pole $\perp\!\!\!\perp \subseteq \mathbb{N}$ and a false equation by every natural number. Although a natural question, we do not delve into the possibility of other definitions.

The third modification involves the interpretation of the first-order universal quantifier $\forall x$. In Krivine's definition, it is interpreted *uniformly*, i.e., a term t realises a universal sentence $\forall x A$ when t realises every instance $A(n)$. This definition is sufficient for the interpretation of second-order arithmetic because the set of natural numbers $\mathbb{N}$ is definable so the axiom of induction does not need to be realised explicitly. In contrast, in case of the first-order arithmetic, realisers of induction must be presented and the uniform interpretation is not ideal for this purpose. As an alternative, we use Kleene's interpretation, where a realiser of a universal sentence $\forall x A$ is, in essence, the code of a recursive function that maps each natural number n to a realiser of $A(n)$. The modifier 'in essence' above is merely because it is the notion of 'refuter' (not 'realiser') which is primitive. The relation between the two in the case of quantifiers is qualified in Lemma 94.

In light of the above remarks, we introduce the following definitions.

Definition 88. A **pole** $\perp\!\!\!\perp$ is a subset of $\mathbb{N}$ such that it is conversely closed under computation: for all $e, m, n \in \mathbb{N}$, if $e \cdot m \simeq n$ and $n \in \perp\!\!\!\perp$, then $\langle e, m\rangle \in \perp\!\!\!\perp$.

Note that the empty set $\emptyset$ and natural numbers $\mathbb{N}$ trivially satisfy the above condition; thus, they are poles.

Given a pole $\perp\!\!\!\perp$ and $\mathcal{L}$-sentence A, we define sets $\|A\|_{\perp\!\!\!\perp}, |A|_{\perp\!\!\!\perp} \subseteq \mathbb{N}$ of, respectively, *refutations* (or a counter-proofs) and *realisations* (or proofs) of A. The sets are defined such that every pair $\langle n, m\rangle \in |A|_{\perp\!\!\!\perp} \times \|A\|_{\perp\!\!\!\perp}$ of a realisation and a refutation is an element of the pole $\perp\!\!\!\perp$. Thus, $\perp\!\!\!\perp$ can be seen as the set of contradictions.

Definition 89. Fix a pole $\perp\!\!\!\perp$. For each $\mathcal{L}$-sentence A, the sets $|A|_{\perp\!\!\!\perp}, \|A\|_{\perp\!\!\!\perp} \subseteq \mathbb{N}$ are defined as follows. The set $|A|_{\perp\!\!\!\perp}$ is defined directly from $\|A\|_{\perp\!\!\!\perp}$:

$$|A|_{\perp\!\!\!\perp} = \{n \in \mathbb{N} \mid \forall m \in \|A\|_{\perp\!\!\!\perp}.\ \langle n, m\rangle \in \perp\!\!\!\perp\}.$$

The set $\|A\|_{\perp\!\!\!\perp}$ is defined inductively:

- $\|s = t\|_{\perp\!\!\!\perp} = \begin{cases} \mathbb{N}, & \text{if } \mathbb{N} \not\models s = t, \\ \perp\!\!\!\perp, & \text{otherwise.} \end{cases}$
- $\|A \to B\|_{\perp\!\!\!\perp} = \{n \mid (n)_0 \in |A|_{\perp\!\!\!\perp} \text{ and } (n)_1 \in \|B\|_{\perp\!\!\!\perp}\}$.
- $\|\forall x A\|_{\perp\!\!\!\perp} = \{n \mid (n)_1 \in \|A((n)_0)\|_{\perp\!\!\!\perp}\}$.

The motivation for the definitions of $\|A\|_{\perp\!\!\!\perp}$ and $|A|_{\perp\!\!\!\perp}$ should be clear. A false equation is refuted by every number whereas a true equation is refuted only by 'contradictions', i.e., elements of the pole. A refutation of $A \to B$ is a pair $\langle m, n\rangle$ for which m realises A and n refutes B. A refutation of $\forall x A$ is a pair $\langle m, n\rangle$ such that n refutes $A(m)$. Finally, a realiser of A is a number n that contradicts all refutations of A, i.e., $\langle n, m\rangle \in \perp\!\!\!\perp$ for every $m \in \|A\|_{\perp\!\!\!\perp}$. In particular, the closure condition on poles implies that every partial recursive function $e : \|A\|_{\perp\!\!\!\perp} \to \perp\!\!\!\perp$ is a realiser of A: if for every $n \in \|A\|_{\perp\!\!\!\perp}$, $e \cdot n$ is defined and an element of $\perp\!\!\!\perp$ then, by definition, $\langle e, n\rangle \in \perp\!\!\!\perp$ for every $n \in \|A\|_{\perp\!\!\!\perp}$. It is also clear that every realiser of A induces a canonical partial recursive function mapping $\|A\|_{\perp\!\!\!\perp}$ to $\perp\!\!\!\perp$.

4.2.2 Compositional Theory for Realisability

A corollary of Tarski's undefinability of truth, the language of PA is insufficient to express classical realisability fully without expanding the non-logical vocabulary. Let $\mathcal{L}_R$ extend $\mathcal{L}$ by three new predicate symbols:

- a unary predicate $\perp\!\!\!\perp$ for a pole (written $x \in \perp\!\!\!\perp$);
- a binary predicate F for refutation (written $x\mathrm{F}y$);
- a binary predicate T for realisation (written $x\mathrm{T}y$).[3]

Definition 90. (Compositional Realisability) The $\mathcal{L}_R$-theory CR extends PA formulated over $\mathcal{L}_R$ by the universal closures of the following axioms:

($\mathsf{Ax}_{\perp\!\!\!\perp}$) $\forall x, y, z(x \cdot y \simeq z \to (z \in \perp\!\!\!\perp \to \langle x, y\rangle \in \perp\!\!\!\perp))$

(Ax_{T}) $\forall \ulcorner A \urcorner \in \mathit{Sent}_{\mathcal{L}}\, \forall a(a\mathrm{T}\ulcorner A\urcorner \leftrightarrow \forall b(b\mathrm{F}\ulcorner A\urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp))$

($\mathsf{CR}_{=}$) $\forall \ulcorner s\urcorner, \ulcorner t\urcorner \forall a(a\mathrm{F}\ulcorner s = t\urcorner \leftrightarrow (\mathit{Eq}(\ulcorner s\urcorner, \ulcorner t\urcorner) \to a \in \perp\!\!\!\perp))$

($\mathsf{CR}_{\to}$) $\forall \ulcorner A\urcorner, \ulcorner B\urcorner \in \mathit{Sent}_{\mathcal{L}}\, \forall a(a\mathrm{F}\ulcorner A \to B\urcorner \leftrightarrow ((a)_0\mathrm{T}\ulcorner A\urcorner \wedge (a)_1\mathrm{F}\ulcorner B\urcorner))$

($\mathsf{CR}_{\forall}$) $\forall \ulcorner A_v\urcorner \in \mathit{Sent}_{\mathcal{L}}\, \forall a(a\mathrm{F}\ulcorner \dot\forall v A\urcorner \leftrightarrow (a)_1\mathrm{F}(\mathit{sub}(\ulcorner A\urcorner, (a)_0)))$

Remark 91. The universal closure of the axiom ($\mathsf{CR}_{=}$) is equivalent to the conjunction of the following:

($\mathsf{CR}_{=1}$) $\forall \ulcorner s\urcorner \forall \ulcorner t\urcorner(\neg \mathit{Eq}(\ulcorner s\urcorner, \ulcorner t\urcorner) \to \forall a.\ a\mathrm{F}\ulcorner s = t\urcorner)$;

($\mathsf{CR}_{=2}$) $\forall \ulcorner s\urcorner \forall \ulcorner t\urcorner(\mathit{Eq}(\ulcorner s\urcorner, \ulcorner t\urcorner) \to \forall a(a\mathrm{F}\ulcorner s = t\urcorner \leftrightarrow a \in \perp\!\!\!\perp))$.

A straightforward formal induction in CR verifies the term regularity principle for refutations (cf. Lemma 87).

Proposition 92. Refutations are provably invariant under term values:

$$\mathsf{CR} \vdash \forall \ulcorner s\urcorner \forall \ulcorner t\urcorner \forall \ulcorner A_v\urcorner \in \mathit{Sent}_{\mathcal{L}}.\ \mathit{Eq}(\ulcorner s\urcorner, \ulcorner t\urcorner) \to \forall x(x\mathrm{F}\ulcorner A(s)\urcorner \to x\mathrm{F}\ulcorner A(t)\urcorner).$$

We provide a model of CR based on classical number realisability. First, the interpretation of the vocabularies of $\mathcal{L}$ is naturally given by the standard model $\mathbb{N}$ of arithmetic. Second, we fix any pole $\perp\!\!\!\perp \subseteq \mathbb{N}$ for the interpretation of the predicate $x \in \perp\!\!\!\perp$. The sets $\mathbb{T}_{\perp\!\!\!\perp}, \mathbb{F}_{\perp\!\!\!\perp} \subseteq \mathbb{N} \times \mathbb{N}$ (for the interpretations of $x\mathrm{T}y$ and $x\mathrm{F}y$, respectively) are defined by the sets $|A|_{\perp\!\!\!\perp}$ and $\|A\|_{\perp\!\!\!\perp}$ in Definition 89:

[3] Although T is definable by F and $\perp\!\!\!\perp$, we introduce T as a primitive to simplify the notation.

$$\mathbb{T}_{\perp\!\!\!\perp} := \{(n,m) \in \mathbb{N}^2 \mid m \text{ is a code of an } \mathcal{L}\text{-sentence } A \text{ and } n \in |A|_{\perp\!\!\!\perp}\},$$
$$\mathbb{F}_{\perp\!\!\!\perp} := \{(n,m) \in \mathbb{N}^2 \mid m \text{ is a code of an } \mathcal{L}\text{-sentence } A \text{ and } n \in \|A\|_{\perp\!\!\!\perp}\}.$$

Then, the following is clear.

Proposition 93. Fix any pole $\perp\!\!\!\perp$ and take any $\mathcal{L}_R$-sentence A. If $\mathsf{CR} \vdash A$, then the $\mathcal{L}_R$-model $\langle \mathbb{N}, \perp\!\!\!\perp, \mathbb{T}_{\perp\!\!\!\perp}, \mathbb{F}_{\perp\!\!\!\perp} \rangle$ satisfies A.

4.3 Formalised Realisation of Peano Arithmetic

We demonstrate that the theory of classical number realisability realises every theorem of PA. In particular, we observe that this is formalisable in CR. For that purpose, the following lemmas are useful.

Lemma 94. There exist numbers $\mathsf{k}_{\to}$, $\mathsf{k}_{\forall}$ and k_{sub} such that

1. $\mathsf{CR} \vdash \forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}} \forall a, b(a \mathrm{T} \ulcorner A \to B \urcorner \wedge b \mathrm{T} \ulcorner A \urcorner \to (\mathsf{k}_{\to} \cdot \langle a, b \rangle) \mathrm{T} \ulcorner B \urcorner)$.
2. $\mathsf{CR} \vdash \forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}}.\ (\forall x.\, (a \cdot x) \mathrm{T} \ulcorner A(\dot{x}) \urcorner) \to (\mathsf{k}_{\forall} \cdot a) \mathrm{T} \ulcorner \forall x A \urcorner$.
3. $\mathsf{CR} \vdash \forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}} \forall a(a \mathrm{T} \ulcorner \dot{\forall} x A \urcorner \to \forall y (\mathsf{k}_{sub} \cdot \langle a, y \rangle) \mathrm{T} \ulcorner A(\dot{y}) \urcorner)$.

The meaning of these functions should be clear, i.e., $\mathsf{k}_{\to}$ computes a realiser of B from those for $A \to B$ and A, and $\mathsf{k}_{\forall}$ expresses that if there exists a procedure that computes every instance $A(n)$, then $\forall x A$ is realised. Conversely, k_{sub} computes a realiser of $A(n)$ for each n.

Proof.

1. Let $\mathsf{k}_{\to}$ be such that $\mathsf{k}_{\to} \cdot \langle a, b \rangle \simeq \lambda x. \langle a, b, x \rangle$. To show that $\mathsf{k}_{\to}$ is the required one, we take any $\mathcal{L}$-sentence $A \to B$ and assume $a \mathrm{T} \ulcorner A \to B \urcorner$ and $b \mathrm{T} \ulcorner A \urcorner$. Then, we must show $(\mathsf{k}_{\to} \cdot \langle a, b \rangle) \mathrm{T} \ulcorner B \urcorner$. By the axiom $(\mathsf{Ax}_{\mathrm{T}})$, taking any c such that $c \mathrm{F} \ulcorner B \urcorner$, we prove $\langle \mathsf{k}_{\to} \cdot \langle a, b \rangle, c \rangle \in \perp\!\!\!\perp$. As $(\mathsf{k}_{\to} \cdot \langle a, b \rangle) \cdot c \simeq \langle a, b, c \rangle$, it is sufficient by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$ to show that $\langle a, b, c \rangle$ is in $\perp\!\!\!\perp$. From the assumptions $b \mathrm{T} \ulcorner A \urcorner$ and $c \mathrm{F} \ulcorner B \urcorner$, as well as the axiom $(\mathsf{Ax}_{\mathrm{T}})$, we obtain $\langle b, c \rangle \mathrm{F} \ulcorner A \to B \urcorner$. Thus, $(\mathsf{Ax}_{\mathrm{T}})$ yields that $\langle a, b, c \rangle = \langle a, \langle b, c \rangle \rangle \in \perp\!\!\!\perp$.
2. Let $\mathsf{k}_{\forall}$ be such that $\mathsf{k}_{\forall} \cdot a \simeq \lambda x. \langle a \cdot (x)_0, (x)_1 \rangle$. To prove that this $\mathsf{k}_{\forall}$ is the required one, we take any a and any $\mathcal{L}$-sentence $\forall x A$. Then, under the assumption that a is total and $\forall x.\, (a \cdot x) \mathrm{T} \ulcorner A(\dot{x}) \urcorner$ holds, we must show $(\mathsf{k}_{\forall} \cdot a) \mathrm{T} \ulcorner \forall x A \urcorner$. By the axiom $(\mathsf{Ax}_{\mathrm{T}})$, taking any b such that $b \mathrm{F} \ulcorner \forall x A \urcorner$, we prove $\langle \mathsf{k}_{\forall} \cdot a, b \rangle \in \perp\!\!\!\perp$. As $(\mathsf{k}_{\forall} \cdot a) \cdot b \simeq \langle a \cdot (b)_0, (b)_1 \rangle$, it is sufficient by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$ to show the latter is in $\perp\!\!\!\perp$. From the assumption, we obtain the formula $(a \cdot (b)_0) \mathrm{T} \ulcorner A((\dot{b})_0) \urcorner$. In addition, by the axiom $(\mathsf{CR}_{\forall})$, we have $(b)_1 \mathrm{F} \ulcorner A((\dot{b})_0) \urcorner$. Thus, the axiom $(\mathsf{Ax}_{\mathrm{T}})$ implies $\langle a \cdot (b)_0, (b)_1 \rangle \in \perp\!\!\!\perp$.

3. Let k_{sub} be such that $\mathsf{k}_{sub} \cdot \langle a, b\rangle \simeq \lambda c.\langle a, b, c\rangle$, and we show that this function is a required one. Thus, taking any $\ulcorner \forall x A(x) \urcorner \in Sent_{\mathcal{L}}$ and any a, b, c, we assume $a\mathrm{T}\ulcorner \forall x A\urcorner$ and $c\mathrm{F}\ulcorner A(\dot{b})\urcorner$. Then, by the axiom $(\mathsf{CR}_{\forall})$, we obtain $\langle b, c\rangle\mathrm{F}\ulcorner \forall x A\urcorner$. Thus, it follows that $\langle a, b, c\rangle \in \perp\!\!\!\perp$ by the axiom $(\mathsf{Ax}_{\mathrm{T}})$. Therefore, the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$ implies that $\langle \mathsf{k}_{sub} \cdot \langle a, b\rangle, c\rangle \in \perp\!\!\!\perp$. Here, the c is arbitrary; thus, we obtain $(\mathsf{k}_{sub} \cdot \langle a, b\rangle)\mathrm{T}\ulcorner A(\dot{b})\urcorner$ again by $(\mathsf{Ax}_{\mathrm{T}})$.

■

Lemma 95. There are numbers k_{π} and $\mathsf{k}_{\perp\!\!\!\perp}$ such that

1. $\mathsf{CR} \vdash \forall \ulcorner A\urcorner, \ulcorner B\urcorner \in Sent_{\mathcal{L}}.\ \forall a(a\mathrm{F}\ulcorner A\urcorner \to (\mathsf{k}_{\pi} \cdot a)\mathrm{T}\ulcorner A \to B\urcorner)$.
2. $\mathsf{CR} \vdash \forall a(a \in \perp\!\!\!\perp \to \forall \ulcorner A\urcorner \in Sent_{\mathcal{L}}.\ (\mathsf{k}_{\perp\!\!\!\perp} \cdot a)\mathrm{T}\ulcorner A\urcorner)$.

Proof.

1. Let $\mathsf{k}_{\pi} := \lambda a.\lambda b.\langle (b)_0, a\rangle$. We prove that this k_{π} is the required number. By taking any a and any $\mathcal{L}$-sentence $A \to B$, we assume $a\mathrm{F}\ulcorner A\urcorner$. To demonstrate that $(\mathsf{k}_{\pi} \cdot a)\mathrm{T}\ulcorner A \to B\urcorner$, we take any b such that $b\mathrm{F}\ulcorner A \to B\urcorner$, and then we must prove $\langle \mathsf{k}_{\pi} \cdot a, b\rangle \in \perp\!\!\!\perp$. By the supposition $b\mathrm{F}\ulcorner A \to B\urcorner$ and the axiom $(\mathsf{CT}_{\to})$, it follows that $(b)_0\mathrm{T}\ulcorner A\urcorner$. Thus, we obtain $(\mathsf{k}_{\pi} \cdot a) \cdot b \simeq \langle (b)_0, a\rangle \in \perp\!\!\!\perp$, which implies $\langle \mathsf{k}_{\pi} \cdot a, b\rangle \in \perp\!\!\!\perp$ by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$.
2. Assuming $a \in \perp\!\!\!\perp$, we define a number $\mathsf{k}_{\perp\!\!\!\perp} := \lambda a.\lambda b.a$. To show $(\mathsf{k}_{\perp\!\!\!\perp} \cdot a)\mathrm{T}\ulcorner A\urcorner$, we take any b such that $b\mathrm{F}\ulcorner A\urcorner$. Then, $(\mathsf{k}_{\perp\!\!\!\perp} \cdot a) \cdot b \simeq a \in \perp\!\!\!\perp$; thus, the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$ implies $\langle \mathsf{k}_{\perp\!\!\!\perp} \cdot a, b\rangle \in \perp\!\!\!\perp$. Therefore, $(\mathsf{k}_{\perp\!\!\!\perp} \cdot a)\mathrm{T}\ulcorner A\urcorner$ holds by the axiom $(\mathsf{Ax}_{\mathrm{T}})$.

■

Note that the above k_{π} is the CPS translation of *call with current continuation* (cf. [26, 48]). Using k_{π}, we can define a realiser for Peirce's law. In Krivine's formulation, Peirce's law is realised by the constant symbol cc. Thus, our formulation is more similar to Oliva and Streicher's formulation of classical realisability [52] in that these constants are definable, i.e., they are not introduced as primitive symbols.

With the above preparations, we can now show the formalised realisability of PA.

Theorem 96. We assume that PA is formulated in the language $\mathcal{L}$. For each $\mathcal{L}$-sentence A, if $\mathsf{PA} \vdash A$, then there exists a closed term s such that $\mathsf{CR} \vdash s\mathrm{T}\ulcorner A\urcorner$.

Moreover, this claim is formally expressible in CR, i.e., we can find a number k_{PA} such that:

$$\mathsf{CR} \vdash \forall x \forall \ulcorner A \urcorner \in \mathit{Sent}_{\mathcal{L}}.\ \mathit{Bew}_{\mathsf{PA}}(x, \ulcorner A \urcorner) \to (\mathsf{k}_{\mathsf{PA}} \cdot x)\mathrm{T}\ulcorner A \urcorner,$$

where $\mathit{Bew}_{\mathsf{PA}}(x, y)$ is a canonical provability predicate for PA expressing means that x is a code of the proof of a sentence y.

Proof. The proof is by induction on the length of the derivation of A in PA. Here, we divide the cases by the last axiom or rule.

Peirce's law. Assume that $A = ((B \to C) \to B) \to B$. According to Lemmas 94 and 95, we define a term s as follows:

$$s = \lambda b.\langle \mathsf{k}_{\to} \cdot \langle (b)_0, \mathsf{k}_{\pi} \cdot (b)_1 \rangle, (b)_1 \rangle.$$

We show that $s\mathrm{T}\ulcorner((B \to C) \to B) \to B\urcorner$. Thus, taking any b satisfying $b\mathrm{F}\ulcorner((B \to C) \to B) \to B\urcorner$, we prove $\langle s, b\rangle \in \perp\!\!\!\perp$. By the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$, it is sufficient to show $\langle \mathsf{k}_{\to} \cdot \langle (b)_0, \mathsf{k}_{\pi} \cdot (b)_1 \rangle, (b)_1 \rangle \in \perp\!\!\!\perp$. From the axiom $(\mathsf{CR}_{\to})$, we obtain $(b)_0\mathrm{T}\ulcorner(B \to C) \to B\urcorner$ and $(b)_1\mathrm{F}\ulcorner B\urcorner$. Thus, by Lemma 95, we obtain $(\mathsf{k}_{\pi} \cdot (b)_1)\mathrm{T}\ulcorner B \to C\urcorner$, which implies that $(\mathsf{k}_{\to} \cdot \langle (b)_0, \mathsf{k}_{\pi} \cdot (b)_1\rangle)\mathrm{T}\ulcorner B\urcorner$ by Lemma 94. Thus, by the axiom $(\mathsf{Ax}_{\mathrm{T}})$, we can derive the required formula $\langle \mathsf{k}_{\to} \cdot \langle (b)_0, (\mathsf{k}_{\pi} \cdot (b)_1)\rangle, (b)_1\rangle \in \perp\!\!\!\perp$.

Induction schema. Assume $A = B(0) \to (\forall x(B(x) \to B(x+1)) \to \forall x B(x))$. We take any $b\mathrm{F}\ulcorner B(0) \to (\forall x(B(x) \to B(x+1)) \to \forall x B(x))\urcorner$. Then, we obtain the following:

- $(b)_0\mathrm{T}\ulcorner B(0)\urcorner$;
- $((b)_1)_0\mathrm{T}\ulcorner\forall x(B(x) \to B(x+1))\urcorner$;
- $((b)_1)_1\mathrm{F}\ulcorner\forall x(B(x))\urcorner$.

By the recursion theorem, choose a number k such that

1. $(\mathsf{k} \cdot b) \cdot 0 \simeq (b)_0$
2. $(\mathsf{k} \cdot b) \cdot (n+1) \simeq \mathsf{k}_{\to} \cdot \langle \mathsf{k}_{sub} \cdot \langle ((b)_1)_0, n\rangle, (\mathsf{k} \cdot b) \cdot n\rangle$ for each n.

Then, we obtain $\forall x.\, ((\mathsf{k} \cdot b) \cdot x)\mathrm{T}\ulcorner B(\dot{x})\urcorner$; thus, Lemma 94 yields the formula $\langle \mathsf{k}_{\forall} \cdot (\mathsf{k} \cdot b), ((b)_1)_1\rangle \in \perp\!\!\!\perp$. Therefore, for the term $s := \lambda b.\langle \mathsf{k}_{\forall} \cdot (\mathsf{k} \cdot b), ((b)_1)_1\rangle$, we have $\langle s, b\rangle \in \perp\!\!\!\perp$ by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$, and thus $s\mathrm{T}\ulcorner A\urcorner$ follows.

Note that the other cases are treated in a similar manner. ∎

4.4 Proof-Theoretic Strength of Compositional Realisability

In Section 4.3, we observed that CR is expressively strong enough to formalise the classical number realisability. In the following, we turn to the proof-theoretic strength of CR and its relationship with CT. First, we show the conservativity of CR over PA.

Proposition 97. CR is conservative over PA.

Proof. We define a translation $\mathcal{T}\colon \mathcal{L}_R \to \mathcal{L}$ such that the vocabularies of $\mathcal{L}$ are unchanged, and then we show the following:

$$\text{for } \mathcal{L}_R\text{-formula } A, \text{ if } \mathsf{CR} \vdash A, \text{ then } \mathsf{PA} \vdash \mathcal{T}(A).$$

If $A \in \mathcal{L}$, we have $\mathcal{T}(A) = A$; thus, the conservativity follows.

The translation $\mathcal{T}$ is defined as follows:

- $\mathcal{T}(s = t) = s = t$;
- $\mathcal{T}(s \in \perp\!\!\!\perp) = \mathcal{T}(s\mathrm{F}t) = \mathcal{T}(s\mathrm{T}t) = (0 = 0)$;
- $\mathcal{T}$ commutes with the logical symbols.

Roughly speaking, each pair is contradictory, and each sentence is realised and refuted by every number under this interpretation. Therefore, we can easily see that the translation of each axiom of CR is derivable in PA. ∎

4.4.1 Compositional Realisability as Compositional Truth

Although CR itself is proof-theoretically weak, here, we show that some assumption on the pole provides CR with the same strength as CT.

Lemma 98. Let $\perp\!\!\!\perp = \emptyset$ denote the sentence $\neg\exists x(x \in \perp\!\!\!\perp)$, and let $\mathsf{CR}^{\emptyset}$ be CR augmented with $\perp\!\!\!\perp = \emptyset$. Then, $\mathsf{CR}^{\emptyset}$ can define the truth prediate of CT as, e.g., the predicate $0\mathrm{T}x$. In other words, $\mathsf{CR}^{\emptyset}$ derives the following:

$(\mathsf{CT}_=)'$ $\forall \ulcorner s \urcorner, \ulcorner t \urcorner (0\mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(\ulcorner s \urcorner, \ulcorner t \urcorner))$

$(\mathsf{CT}_\to)'$ $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in \mathit{Sent}_{\mathcal{L}}\, (0\mathrm{T}\ulcorner A \to B \urcorner \leftrightarrow (0\mathrm{T}\ulcorner A \urcorner \to 0\mathrm{T}\ulcorner B \urcorner))$

$(\mathsf{CT}_\forall)'$ $\forall \ulcorner A_x \urcorner \in \mathit{Sent}_{\mathcal{L}}(0\mathrm{T}\ulcorner \forall x A \urcorner \leftrightarrow \forall x\, 0\mathrm{T}\ulcorner A(\dot{x}) \urcorner)$.

Therefore, every $\mathcal{L}$-theorem of CT is derivable in $\mathsf{CR}^{\emptyset}$.

Proof. By $⫫ = \emptyset$ and the axiom $(\mathsf{Ax}_{\mathrm{T}})$, we easily obtain the following:

$$0\mathrm{T}\ulcorner A \urcorner \leftrightarrow \forall b(\neg b\mathrm{F}\ulcorner A \urcorner).$$

With this, we can derive the formulas $(\mathsf{CT}_=)'$, $(\mathsf{CT}_\to)'$, and $(\mathsf{CT}_\forall)'$.

$(\mathsf{CT}_=)'$ In $\mathsf{CR}^\emptyset$, we deduce as follows.

$$\begin{aligned} 0\mathrm{T}\ulcorner s = t \urcorner &\Leftrightarrow \forall b(\neg b\mathrm{F}\ulcorner s = t \urcorner) && \text{by } (\mathsf{Ax}_{\mathrm{T}}) \text{ and } ⫫ = \emptyset \\ &\Rightarrow Eq(s,t) && \text{by } (\mathsf{CR}_{=1}) \\ &\Rightarrow \forall b(\neg b\mathrm{F}\ulcorner s = t \urcorner) && \text{by } (\mathsf{CR}_{=2}) \text{ and } ⫫ = \emptyset \end{aligned}$$

$(\mathsf{CT}_\forall)'$

$$\begin{aligned} 0\mathrm{T}\ulcorner \forall x A \urcorner &\Leftrightarrow \forall b\, \neg b\mathrm{F}\ulcorner \forall x A \urcorner && \text{by } (\mathsf{Ax}_{\mathrm{T}}) \text{ and } ⫫ = \emptyset \\ &\Leftrightarrow \forall b\, \neg((b)_1\mathrm{F}(sub(\ulcorner A \urcorner, (b)_0))) && \text{by } (\mathsf{CR}_\forall) \\ &\Leftrightarrow \forall x \forall y\, \neg(y\mathrm{F}\ulcorner A(\dot{x}) \urcorner) && \\ &\Leftrightarrow \forall x\, 0\mathrm{T}\ulcorner A(\dot{x}) \urcorner && \text{by } (\mathsf{Ax}_{\mathrm{T}}) \text{ and } ⫫ = \emptyset \end{aligned}$$

The other cases are similar. ■

Thus, the assumption $⫫ = \emptyset$ reduces truth to realisability. Next, we give another characterisation of this assumption using a kind of reflection principle stating that realisability is subsumed by truth.

Lemma 99. Over CR, the following are equivalent.

1. The reflection schema: $\exists x(x\mathrm{T}\ulcorner A \urcorner) \to A$ for every $\mathcal{L}$-sentence A.
2. The axiom: $⫫ = \emptyset$.

Proof.

(1) ⇒ (2): Assume for a contradiction that $a \in ⫫$ for some a. Then, $\mathsf{k}_{⫫} \cdot a$ verifies every $\mathcal{L}$-sentence by Lemma 95. Thus, the schema (1) implies every sentence, a contradiction. Therefore, the axiom (2) : $⫫ = \emptyset$ follows.

(2) ⇒ (1): As shown in Lemma 98, CR with $⫫ = \emptyset$ can define the truth predicate of CT as $0\mathrm{T}x$. Thus, similar to Lemma 86, we obtain $0\mathrm{T}\ulcorner A \urcorner \to A$ for every $\mathcal{L}$-sentence A. Therefore, we also have schema (1).

■

Proposition 100. The theories $\mathsf{CR}^{\emptyset}$, CR with the reflection schema, and CT have exactly the same $\mathcal{L}$-consequences.

Proof. In Lemma 98, we observed that CT is interpretable in $\mathsf{CR}^{\emptyset}$; thus, it is interpretable in CR with the reflection schema (Lemma 99).

For the converse direction, we note that the model construction of CR in Proposition 93 is also applicable to $\mathsf{CR}^{\emptyset}$ and is formalisable in the theory ACA, the second-order system for arithmetical comprehension, which has the same $\mathcal{L}$-consequences as CT (for the proof, see, e.g., [30]). ■

4.4.2 Compositional Realisability with the Reflection Rule

Although $\mathsf{CR}^{\emptyset}$ (or equivalently CR with the reflection schema) and CT have the same proof-theoretic strength, $\mathsf{CR}^{\emptyset}$ is satisfied only when the pole $⫫$ is empty. Thus, our next goal is to find a principle that is compatible with any choice of the pole. Here, our suggestion is to weaken the reflection schema to the rule form.

Definition 101. (CR^{+}) The $\mathcal{L}_R$-theory CR^{+} is the extension of CR with the reflection rule:

$$\frac{s\mathrm{T}\ulcorner A\urcorner}{A}$$

for every closed term s and $\mathcal{L}$-sentence A.

Note that $s\mathrm{T}\ulcorner A\urcorner$ must be derived without assumptions.

Proposition 102. Let A be any $\mathcal{L}_R$-sentence and assume that $\mathsf{CR}^{+} \vdash A$. Then, we have $\langle \mathbb{N}, ⫫, \mathbb{T}_{⫫}, \mathbb{F}_{⫫}\rangle \models A$ for any pole $⫫$. Here, $\langle \mathbb{N}, ⫫, \mathbb{T}_{⫫}, \mathbb{F}_{⫫}\rangle$ is an $\mathcal{L}_R$-model in which the vocabularies of $\mathcal{L}$ are interpreted in the standard model $\mathbb{N}$, $⫫$ is for the symbol $x \in ⫫$, $\mathbb{T}_{⫫}$ is for $x\mathrm{T}y$, and $\mathbb{F}_{⫫}$ is for $x\mathrm{F}y$.

Proof. The proof is by induction on the derivation of A. Since the other cases are already contained in the proof of Proposition 93, it is sufficient to consider the case of the reflection rule. Therefore, we assume that an $\mathcal{L}$-sentence A is derived by the reflection rule from $s\mathrm{T}\ulcorner A\urcorner$ for some closed term s. By the induction hypothesis, it follows that $\langle \mathbb{N}, ⫫, \mathbb{T}_{⫫}, \mathbb{F}_{⫫}\rangle \models s\mathrm{T}\ulcorner A\urcorner$ for any pole $⫫$. Thus, particularly for the empty pole $⫫ = \emptyset$, we obtain $\langle \mathbb{N}, \emptyset, \mathbb{T}_{⫫}, \mathbb{F}_{⫫}\rangle \models A$ by Proposition 93 and Lemma 99. As A is an $\mathcal{L}$-sentence, we also have $\langle \mathbb{N}, ⫫, \mathbb{T}_{⫫}, \mathbb{F}_{⫫}\rangle \models A$ for any pole $⫫$. ■

In the remainder of this section, we prove that CR^{+} has the same proof-theoretic strength as CT and $\mathsf{CR}^{\emptyset}$. The upper bound of CR^{+} is obvious: Lemma 99 and Proposition 102 establish that CR^{+} is a *proper* subtheory of $\mathsf{CR}^{\emptyset}$. The lower-bound argument is more difficult because CR^+ is not expressively rich enough to interpret CT. Instead we can proceed directly through a well-ordering proof for CR^+ by showing that the principle of transfinite induction for $\mathcal{L}$-formulas is provable for each ordinal below $\varepsilon_{\varepsilon_0}$. The argument is, essentially, just the extraction of the computational content of the standard well-ordering proof for CT (for the detailed proof, see Section 4.4.3).

As a result, we can determine the proof-theoretic strength of CR^+.

Theorem 103. CR^+ has exactly the same $\mathcal{L}$-theorems as CT and $\mathsf{CR}^\emptyset$.

4.4.3 Well-Ordering Proof in Compositional Realisability

Here, we determine the proof-theoretic strength of CR^+. However, in contrast to $\mathsf{CR}^\emptyset$, CR^+ is not sufficiently expressively strong to relatively interpret CT. Thus, we provide a well-ordering proof of CR^+, from which we can conclude that CR^+ derives the same $\mathcal{L}$-consequences as both CT and $\mathsf{CR}^\emptyset$.

For this purpose, we require an ordinal notation system OT for predicative ordinal numbers. We use several facts about OT (for the proof, see, e.g., [54]). A formula $x \in OT$ is defined as meaning that x is a representation of an ordinal number in OT. Let α, β, and γ range over the ordinal numbers in OT. Thus, $\forall \alpha A(\alpha)$ abbreviates $\forall x(x \in OT \to A(x))$. By the standard method, we can define relations and operations on OT. Let $<$ be the less-than relation, 0 is zero as the ordinal number, $\alpha \in Suc$ means that α is a successor ordinal, $\alpha \in Lim$ says that α is a limit ordinal, $+$ is the ordinal addition, φxy is the Veblen function, and the binary primitive recursive function $[\alpha]_x$ returns the x-th element of the fundamental sequence for α. For these symbols, the following notations and facts are used in this paper:

- Let $\omega^\alpha := \varphi 0 \alpha$ and $1 := \omega^0$.
- $[\omega^\alpha]_n = \begin{cases} \omega^{\alpha-1} \times n & \text{if } \alpha \in Suc, \\ \omega^{[\alpha]_n} & \text{if } \alpha \in Lim. \end{cases}$
- Let $\omega_n(\alpha) := \begin{cases} \alpha & \text{if } n = 0, \\ \omega^{\omega_{n-1}(\alpha)} & \text{if } n > 0. \end{cases}$
- Let $\varepsilon_\alpha := \varphi 1 \alpha$.
- $[\varepsilon_\alpha]_n = \begin{cases} \omega_n(1) & \text{if } \alpha = 0, \\ \omega_n(\varepsilon_{\alpha-1} + 1) & \text{if } \alpha \in Suc, \\ \varepsilon_{[\alpha]_n} & \text{if } \alpha \in Lim. \end{cases}$

The T-free consequences of CT can be expressed using some transfinite induction schema. For an $\mathcal{L}$-formula $A(x)$ and an ordinal α, we define the formula $\mathsf{TI}(A,\alpha) = \mathsf{Prog}\lambda x A(x) \to A(\alpha)$, where $\mathsf{Prog}\lambda x A(x) = \forall\alpha(\forall\beta < \alpha(A(\beta)) \to A(\alpha))$. Then, we define the schema: $\mathsf{TI}(\alpha) = \{\mathsf{TI}(A,\beta) \mid A \text{ is an } \mathcal{L}\text{-formula}\}$. In addition, let $\mathsf{TI}(<\alpha) := \bigcup_{\beta<\alpha} \mathsf{TI}(\beta)$. Finally, let the $\mathcal{L}$-theory $\mathsf{PA}+\mathsf{TI}(<\alpha)$ be the extension of PA with the schema $\mathsf{TI}(<\alpha)$.

The following is well-known (see, e.g., [30, Theorem 8.35]):

Theorem 104. The theory CT derives the same $\mathcal{L}$-formulae as $\mathsf{PA}+\mathsf{TI}(<\varepsilon_{\varepsilon_0})$.

According to this fact, it is sufficient to show that the schema $\mathsf{TI}(<\varepsilon_{\varepsilon_0})$ is derivable in CR^+. To this end, in Lemma 107, we prove that each instance of $\mathsf{TI}(<\varepsilon_{\varepsilon_0})$ is realisable in CR. Then, CR^+ can derive $\mathsf{TI}(<\varepsilon_{\varepsilon_0})$ itself according to the reflection rule.

The proof is essentially based on the standard well-ordering proof of CT (cf. [51, Lemma 3.11]). So, we briefly sketch the outline of the proof in CT. Let $I_0(\alpha) = \forall\ulcorner A\urcorner \in \mathit{Sent}_{\mathcal{L}}.\ \mathrm{T}\ulcorner\mathsf{TI}(A,\dot{\alpha})\urcorner$, which expresses the schema $\mathsf{TI}(\alpha)$ as a single statement. In PA, we can derive the schemata $\mathsf{TI}(0)$, $\mathsf{TI}(\alpha+1)$ from $\mathsf{TI}(\alpha)$, and $\mathsf{TI}(\omega^\alpha)$ from $\mathsf{TI}(\alpha)$, respectively (e.g., see [54, Section 7.4]). By formalising these results, CT can derive $I_0(\varepsilon_0)$. Furthermore, by generalising this argument in CT, we can derive $I_0(\varepsilon_\alpha) \to I_0(\varepsilon_{\alpha+1})$. In addition, CT derives $\alpha \in \mathit{Lim} \to \{[\forall\beta < \alpha I_0(\varepsilon_\beta)] \to I_0(\varepsilon_\alpha)\}$. These facts together mean that $I_0(\varepsilon_x)$ is progressive, i.e., $\mathsf{CT} \vdash \mathsf{Prog}\lambda x I_0(\varepsilon_x)$. Since CT can derive the transfinite induction for $I_0(\varepsilon_x)$ up to ε_0, the schema $\mathsf{TI}(<\varepsilon_{\varepsilon_0})$ is obtained in CT, as required. To emulate this proof within CR, we must extract the computational content of the proof. For example, it is necessary to explicitly give a partial recursive function that computes a realiser of $\mathsf{TI}(\alpha+1)$ from that of $\mathsf{TI}(\alpha)$.

We define $I_0(e,\alpha) = \forall\ulcorner A\urcorner \in \mathit{Sent}_{\mathcal{L}}.\ (e\cdot\ulcorner A\urcorner)\mathrm{T}\ulcorner\mathsf{TI}(A,\dot{\alpha})\urcorner$, which means that e computes a realiser of the transfinite induction $\mathsf{TI}(A,\alpha)$ for each $\mathcal{L}$-sentence A.

Lemma 105. 1. There exists a number k_0 such that:

$$\mathsf{CR} \vdash I_0(\mathsf{k}_0, 0).$$

2. There exists a number k_{suc} such that:

$$\mathsf{CR} \vdash I_0(e,\alpha) \to I_0(\mathsf{k}_{suc}\cdot\langle e,\alpha\rangle, \alpha+1).$$

3. There exists a number k_ω such that:

$$\mathsf{CR} \vdash I_0(e,\alpha) \to I_0(\mathsf{k}_\omega\cdot\langle e,\alpha\rangle, \omega^\alpha).$$

4. There exists a number k_{lim} such that:

$$\mathsf{CR} \vdash \alpha \in \mathit{Lim} \to [\forall n I_0(e\cdot n, [\alpha]_n) \to I_0(\mathsf{k}_{lim}\cdot\langle e,\alpha\rangle, \alpha)].$$

Note that k_0 realises the schema $\mathsf{TI}(0)$; k_{suc} realises $\mathsf{TI}(\alpha+1)$ from $\mathsf{TI}(\alpha)$; k_ω realises $\mathsf{TI}(\omega^\alpha)$ from $\mathsf{TI}(\alpha)$; k_{lim} realises $\mathsf{TI}(\alpha)$ for a limit ordinal α if each $\mathsf{TI}([\alpha]_n)$ is realised.

Proof.

1. For every $\mathcal{L}$-formula A, we can primitive recursively find a proof of $\mathsf{TI}(A, 0)\urcorner)$ in PA. Thus, by Theorem 96, there exists a required k_0.
2. In PA, we can primitive recursively find a proof of $\forall\alpha(\mathsf{TI}(A,\alpha) \to \mathsf{TI}(A,\alpha+1))$ for each $\mathcal{L}$-formula A. Thus, by Theorem 96, there exists a number k such that $\mathsf{CR} \vdash ((\mathsf{k}\cdot\alpha)\cdot\ulcorner A\urcorner)\mathrm{T}\ulcorner\mathsf{TI}(A,\alpha) \to \mathsf{TI}(A,\alpha+1)\urcorner$. For this k, we define k_{suc} to be such that for any $\mathcal{L}$-sentence A,
$$(\mathsf{k}_{suc}\cdot\langle e,\alpha\rangle)\cdot\ulcorner A\urcorner \simeq \mathsf{k}_{\to}\cdot\langle(\mathsf{k}\cdot\alpha)\cdot\ulcorner A\urcorner, e\cdot\ulcorner A\urcorner\rangle.$$
Then, if e satisfies $I_0(e,\alpha)$, we have $((\mathsf{k}_{suc}\cdot\langle e,\alpha\rangle)\cdot\ulcorner A\urcorner)\mathrm{T}\ulcorner\mathsf{TI}(A,\dot{\alpha}+1)\urcorner$, as required.
3. For every $\mathcal{L}$-formula A, there exists an $\mathcal{L}$-formula A' such that we can primitive recursively find a proof of $\forall\alpha(\mathsf{TI}(A',\alpha) \to \mathsf{TI}(A,\omega^\alpha))$ in PA. Thus, similar to the proof of the item 2, there is a required function k_ω.
4. We assume $\forall n I_0(e\cdot n, [\alpha]_n)$. From this e, we can primitive recursively define $\mathsf{k}^\dagger$ such that:
$$\forall\ulcorner A\urcorner \in \mathit{Sent}_{\mathcal{L}}.\ ((\mathsf{k}^\dagger\cdot\langle e,\alpha\rangle)\cdot\ulcorner A\urcorner)\mathrm{T}\ulcorner\forall\beta < \dot{\alpha}\mathsf{TI}(A,\beta)\urcorner.$$
In addition, for each $\mathcal{L}$-formula A, we obtain $\forall\beta < \alpha\mathsf{TI}(A,\beta) \to \mathsf{TI}(A,\alpha)$ in PA; thus, according to Thorem 96, we take a number $\mathsf{k}^\ddagger$ such that:
$$((\mathsf{k}^\ddagger\cdot\alpha)\cdot\ulcorner A\urcorner)\mathrm{T}\ulcorner\forall\beta < \dot{\alpha}\mathsf{TI}(A,\beta) \to \mathsf{TI}(A,\dot{\alpha})\urcorner.$$
Now, we define k_{lim} such that:
$$(\mathsf{k}_{lim}\cdot\langle e,\alpha\rangle)\cdot\ulcorner A\urcorner \simeq \mathsf{k}_{\to}\cdot\langle(\mathsf{k}^\ddagger\cdot\alpha)\cdot\ulcorner A\urcorner, (\mathsf{k}^\dagger\cdot\langle e,\alpha\rangle)\cdot\ulcorner A\urcorner\rangle.$$
We then obtain $((\mathsf{k}_{lim}\cdot\langle e,\alpha\rangle)\cdot\ulcorner A\urcorner)\mathrm{T}\ulcorner\mathsf{TI}(A,\dot{\alpha})\urcorner$, as required.

■

The following lemma shows the progressiveness of the epsilon function.

Lemma 106. 1. There exists a number $\mathsf{k}_{\varepsilon\_0}$ such that:
$$\mathsf{CR} \vdash I_0(\mathsf{k}_{\varepsilon\_0}, \varepsilon_0).$$

2. There exists a number $\mathsf{k}_{\varepsilon\_suc}$ such that:

$$\mathsf{CR} \vdash I_0(e, \varepsilon_\alpha) \to I_0(\mathsf{k}_{\varepsilon\_suc} \cdot \langle e, \alpha \rangle, \varepsilon_{\alpha+1}).$$

3. There exists a number k_ε such that:

$$\mathsf{CR} \vdash \mathsf{Prog}\lambda\alpha I_0(\mathsf{k}_\varepsilon \cdot \alpha, \varepsilon_\alpha).$$

Proof.

1. We define a number k as follows:
$\begin{cases} \mathsf{k} \cdot 0 :\simeq \mathsf{k}_{suc} \cdot \langle \mathsf{k}_0, 0 \rangle, \\ \mathsf{k} \cdot (n+1) :\simeq \mathsf{k}_\omega \cdot \langle \mathsf{k} \cdot n, [\varepsilon_0]_n \rangle \quad \text{for each } n. \end{cases}$
Then, we clearly have $\mathsf{CR} \vdash \forall n(I_0(\mathsf{k} \cdot n, [\varepsilon_0]_n)$, hence by item 4 in Lemma 105, the number $\mathsf{k}_{\varepsilon\_0} := \mathsf{k}_{lim} \cdot \langle \mathsf{k}, \varepsilon_0 \rangle$ is the required one.
2. The proof is nearly the same as that of item 1.
3. For an ordinal α and a number e, let $e[\alpha]$ be such that $e[\alpha] \cdot n :\simeq e \cdot [\alpha]_n$. Here, k_ε is defined as follows:
$\begin{cases} \mathsf{k}_\varepsilon \cdot 0 :\simeq \mathsf{k}_{\varepsilon\_0} \\ \mathsf{k}_\varepsilon \cdot (\alpha+1) :\simeq \mathsf{k}_{\varepsilon\_suc} \cdot \langle \mathsf{k}_\varepsilon \cdot \alpha, \alpha \rangle \\ \mathsf{k}_\varepsilon \cdot \alpha :\simeq \mathsf{k}_{lim} \cdot \langle \mathsf{k}_\varepsilon[\alpha], \varepsilon_\alpha \rangle \qquad \text{for } \alpha \in \mathit{Lim}. \end{cases}$
Then, to show the claim,we take any α and assume $\forall \beta < \alpha I_0(\mathsf{k}_\varepsilon \cdot \beta, \varepsilon_\beta)$.
   - If $\alpha = 0$, then the conclusion $I_0(\mathsf{k}_\varepsilon \cdot \alpha, \varepsilon_\alpha)$ is clear by item 1.
   - If $\alpha \in \mathit{Suc}$, then by the assumption, we obtain $I_0(\mathsf{k}_\varepsilon \cdot (\alpha-1), \varepsilon_{\alpha-1})$; thus, we obtain the conclusion by item 2.
   - If $\alpha \in \mathit{Lim}$, then we obtain $I_0(\mathsf{k}_\varepsilon[\alpha] \cdot n, \varepsilon_{[\alpha]_n})$ for each n. Here, since $[\varepsilon_\alpha]_n = \varepsilon_{[\alpha]_n}$, Lemma 105 implies that $I_0(\mathsf{k}_{lim} \cdot \langle \mathsf{k}_\varepsilon[\alpha], \varepsilon_\alpha \rangle, \varepsilon_\alpha \rangle$; thus, it follows that $I_0(\mathsf{k}_\varepsilon \cdot \alpha, \varepsilon_\alpha)$.

■

Lemma 107. For each $\mathcal{L}$-formula A and for each ordinal number $\alpha < \varepsilon_{\varepsilon_0}$, we can find a term s such that $\mathsf{CR} \vdash s\mathrm{T}\ulcorner \mathsf{TI}(A, \alpha) \urcorner$.

Proof. We fix any ordinal $\alpha < \varepsilon_{\varepsilon_0}$, and then there exists an ordinal $\beta < \varepsilon_0$ such that $\alpha \leq \varepsilon_\beta < \varepsilon_{\varepsilon_0}$. Thus, by taking any $\mathcal{L}$-formula A, it is sufficient to prove $((\mathsf{k}_\varepsilon \cdot \beta) \cdot \ulcorner A \urcorner)\mathrm{T}\ulcorner \mathsf{TI}(A, \varepsilon_\beta) \urcorner$ because we can primitive recursively find a required term s from the term $(\mathsf{k}_\varepsilon \cdot \beta) \cdot \ulcorner A \urcorner$.

Since PA derives any transfinite induction for $\beta < \varepsilon_0$, we have $\mathsf{TI}(I_0(\mathsf{k}_\varepsilon \cdot x, \varepsilon_x), \beta)$. Therefore, according to item 3 in Lemma 106, we have $I_0(\mathsf{k}_\varepsilon \cdot \beta, \varepsilon_\beta)$. Thus, it follows that $((\mathsf{k}_\varepsilon \cdot \beta) \cdot \ulcorner A \urcorner)\mathrm{T}\ulcorner \mathsf{TI}(A, \varepsilon_\beta)\urcorner$, as desired. ■

By Lemma 107 and the reflection rule, CR^+ yields the formula $\mathsf{TI}(A, \alpha)$ for each $\mathcal{L}$-formula A and $\alpha < \varepsilon_{\varepsilon_0}$. Thus, CR^+ derives every theorem of $\mathsf{PA} + \mathsf{TI}(< \varepsilon_{\varepsilon_0})$. Therefore, according to Proposition 100 and Theorem 104, the proof of Theorem 103 is completed.

4.5 Friedman–Sheard-Style Realisability

Starting from the typed theory CR for realisability, it seems natural to extend it to type-free theories. In this section, we formulate a Friedman-Sheard-style theory FSR for classical realisability and study its proof-theoretic properties. In particular, we show that FSR has a nice property, *self realisability*, which states that if A is derived, then A is realisable provably in the same system.

4.5.1 Friedman–Sheard Realisability FSR

In this subsection, we want to formulate a self-referential system of classical realisability. For this purpose, let us refer to the case of truth theory. Recall that the system FS (Definition 7) consists of the compositional axioms, where the range of quantification of sentences is not only $\mathcal{L}$, but also $\mathcal{L}_\mathrm{T}$. In addition, FS has the introduction and elimination rules for the truth predicate. Here, we restate the definition for the current language:

Definition 108. (cf. Definition 7) The $\mathcal{L}_\mathrm{T}$-theory GCT (generalised compositional truth) consists of PA and the following axioms:

($\mathsf{GCT}_=$) $\forall \ulcorner s \urcorner, \ulcorner t \urcorner.\ \mathrm{T}\ulcorner s = t \urcorner \leftrightarrow Eq(\ulcorner s \urcorner, \ulcorner t \urcorner)$;

($\mathsf{GCT}_\rightarrow$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}_\mathrm{T}}.\ \mathrm{T}\ulcorner A \rightarrow B \urcorner \leftrightarrow (\mathrm{T}\ulcorner A \urcorner \rightarrow \mathrm{T}\ulcorner B \urcorner)$;

($\mathsf{GCT}_\forall$) $\forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}_\mathrm{T}}.\ \mathrm{T}\ulcorner \forall x A \urcorner \leftrightarrow \forall v \mathrm{T}\ulcorner A(\dot{v}) \urcorner$.

The $\mathcal{L}_\mathrm{T}$-theory FS (Friedman-Sheard theory of truth) is the extension of GCT with the following rules: for each $\mathcal{L}_\mathrm{T}$-sentence A,

$$\frac{A}{\mathrm{T}\ulcorner A \urcorner}\ (\mathsf{NEC}) \quad \frac{\mathrm{T}\ulcorner A \urcorner}{A}\ (\mathsf{CONEC}).$$

In contrast to the axiom ($\mathsf{GCT}_=$), the Tarski-biconditional for the truth predicate yields a paradox:

Fact 109. ([20]) 1. GCT with the following axiom is inconsistent:

$$\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}} (\mathrm{T}\ulcorner \mathrm{T}(\dot{\overbrace{\ulcorner A \urcorner}}) \urcorner \leftrightarrow \mathrm{T}\ulcorner A \urcorner).$$

2. GCT with the following schema is inconsistent:

$$\mathrm{T}\ulcorner A \urcorner \to A, \text{ for each } \mathcal{L}_{\mathrm{T}}\text{-sentence } A.$$

Based on GCT, we define its realisability version GCR :

Definition 110. (GCR) An $\mathcal{L}_R$-theory GCR (generalised compositional realisability) consists of PA and the universal closures of the following axioms:

($\mathsf{Ax}_{\perp\!\!\!\perp}$) $T(a,b,c) \to (U(c) \in \perp\!\!\!\perp \to \langle a,b \rangle \in \perp\!\!\!\perp)$

(Ax_{T}) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_R}.\ a\mathrm{T}\ulcorner A \urcorner \leftrightarrow \forall b(b\mathrm{F}\ulcorner A \urcorner \to \langle a,b \rangle \in \perp\!\!\!\perp)$

(GCR_{reg}) $\forall \ulcorner s \urcorner, \ulcorner t \urcorner \forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}_R}.\ Eq(\ulcorner s \urcorner, \ulcorner t \urcorner) \to [a\mathrm{F}\ulcorner A(s) \urcorner \to a\mathrm{F}\ulcorner A(t) \urcorner]$

(GCR_{P1}) $\neg P(\overrightarrow{x}) \to a\mathrm{F}\ulcorner P(\dot{\overrightarrow{x}}) \urcorner$

(GCR_{P2}) $P(\overrightarrow{x}) \to (a\mathrm{F}\ulcorner P(\dot{\overrightarrow{x}}) \urcorner \leftrightarrow a \in \perp\!\!\!\perp)$

($\mathsf{GCR}_{\to}$) $\forall \ulcorner A \urcorner, \ulcorner B \urcorner \in Sent_{\mathcal{L}_R}.\ a\mathrm{F}\ulcorner A \to B \urcorner \leftrightarrow [(a)_0\mathrm{T}\ulcorner A \urcorner \wedge (a)_1\mathrm{F}\ulcorner B \urcorner]$

($\mathsf{GCR}_{\forall}$) $\forall \ulcorner A_x \urcorner \in Sent_{\mathcal{L}_R}.\ a\mathrm{F}\ulcorner \forall x A \urcorner \leftrightarrow (a)_1\mathrm{F}\ulcorner A((\dot{a})_0) \urcorner$

Here, P is an atomic predicate of $\mathcal{L} \cup \{\perp\!\!\!\perp\}$.

Note that the above axiom ($\mathsf{Ax}_{\perp\!\!\!\perp}$) is exactly the same as in Definition 90, but we have chosen not to use abbreviations because its precise form is important to prove the realisability of ($\mathsf{Ax}_{\perp\!\!\!\perp}$) in Lemma 118. Similarly, to simplify the proof of Lemmata 121 and 122, we separate the axioms for atomic predicates into (GCR_{P1}) and (GCR_{P2}) (cf. Remark 91). Because of this choice, the term regularity principle (GCR_{reg}) (Proposition 92) needs to be explicitly added as an axiom.

As is clear from the above definition, GCR is just an generalisation of CR in that each axiom of GCR ranges over every $\mathcal{L}_R$-sentences. Thus, by the same proof, all the results in CR, except Lemma 99 and Proposition 100, hold also in GCR by replacing the range of quantification of sentences from $\mathcal{L}$ to $\mathcal{L}_R$. Because of that, even in GCR, we shall sloppily use the result for CR.

Similar to the case of truth theory (Fact 109), the next proposition shows that we need to treat T and F differently from the atomic predicates of $\mathcal{L}$. That is the reason why Lemma 99 cannot be generalised to $\mathcal{L}_R$-sentences.

Proposition 111. 1. Let GCR' be the extension of GCR with the following axioms:

($\mathsf{GCR}_{\mathrm{F1}}$) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_R}.\ \neg b\mathrm{F}\ulcorner A \urcorner \to a\mathrm{F}\ulcorner \dot{b}\mathrm{F}(\dot{\overbrace{\ulcorner A \urcorner}}) \urcorner$,

($\mathsf{GCR}_{\mathrm{F2}}$) $\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_R}.\ b\mathrm{F}\ulcorner A \urcorner \to (a\mathrm{F}\ulcorner \dot{b}\mathrm{F}(\dot{\overbrace{\ulcorner A \urcorner}}) \urcorner \leftrightarrow a \in \perp\!\!\!\perp)$.

Then, $\mathsf{GCR}' \vdash \perp\!\!\!\perp \neq \emptyset$.

2. The extension of GCR with the following reflection schema is inconsistent:

$$\exists x(x\mathrm{T}\ulcorner A \urcorner) \to A, \text{ for each } \mathcal{L}_R\text{-sentence } A.$$

Proof.

1. For a contradiction, we suppose $\perp\!\!\!\perp = \emptyset$ in GCR'. Then, in the same way as for the proof of Lemma 99, we can define the truth predicate $\mathrm{T}(x)$ of GCT as $0\mathrm{T}(x)$. In addition, from the axioms ($\mathsf{GCR}_{\mathrm{F1}}$) and ($\mathsf{GCR}_{\mathrm{F2}}$), we derive in GCR:

$$\forall \ulcorner A \urcorner \in Sent_{\mathcal{L}_{\mathrm{T}}}.\ 0\mathrm{T}\ulcorner 0\mathrm{T}(\dot{\overbrace{\ulcorner A \urcorner}}) \urcorner \leftrightarrow 0\mathrm{T}\ulcorner A \urcorner.$$

Thus, by Fact 109, we obtain a contradiction, and hence $\perp\!\!\!\perp \neq \emptyset$.

2. Similarly to Lemma 99, the reflection schema implies in GCR that $\perp\!\!\!\perp = \emptyset$. Thus, in the same way as for the item 1, the truth predicate $\mathrm{T}(x)$ of GCT is definable as $0\mathrm{T}(x)$. Furthermore, the reflection schema implies the schema:

$$0\mathrm{T}\ulcorner A \urcorner \to A, \text{ for each } \mathcal{L}_R\text{-sentence } A.$$

Therefore, by Fact 109, we obtain a contradiction.

∎

In the remaining part of this subsection, we formulate counterparts of the rules (NEC) and (CONEC). The rule (NEC) informally says that from the truth of A, we can infer the *formal* truth of A. Based on this idea, we firstly define realisability in the form of an explicit $\mathcal{L}_R$-formula $n \in |A|$. Then, the realisability version of (NEC) should allow an inference from the explicit realisability $n \in |A|$ to the formal realisability $n\mathrm{T}\ulcorner A \urcorner$. The rule ($\mathsf{CONEC}$) is similarly treated.

Definition 112. (Explicit realisation) For each term s and $\mathcal{L}_R$-formula A, we inductively define $\mathcal{L}_R$-formulas $s \in |A|$ and $s \in \|A\|$, with renaming bound variables in A if necessary.

- $s \in \|P\| :\equiv P \to s \in \perp\!\!\!\perp$, if $P \in \mathcal{L} \cup \{\perp\!\!\!\perp\}$;

- $s \in \|t\mathrm{F}u\| :\equiv s\mathrm{F}(t\dot{\in}\|u\|)$;
- $s \in \|t\mathrm{T}u\| :\equiv s\mathrm{F}(t\dot{\in}|u|)$;
- $s \in \|A \to B\| :\equiv (s)_0 \in |A| \wedge (s)_1 \in \|B\|$;
- $s \in \|\forall x A(x)\| :\equiv (s)_1 \in \|A((s)_0)\|$;
- $s \in |A| :\equiv \forall a(a \in \|A\| \to \langle s, a\rangle \in \perp\!\!\!\perp)$ for a fresh variable a.

Here, $x\dot{\in}|y|$ is a binary primitive recursive function symbol such that PA derives $x\dot{\in}|\ulcorner A \urcorner| = \ulcorner \dot{x} \in A \urcorner$ for any $\mathcal{L}_R$-sentence A. The existence of such a function is assured by the primitive recursion theorem. Another binary primitive recursive function symbol $x\dot{\in}\|y\|$ is similarly defined.

We can easily verify the next substitution lemma:

Lemma 113. 1. For every terms s, t and $\mathcal{L}_R$-formula A, we have:

   (a) $\mathsf{PA} \vdash (t \in \|A\|)[x/s] \;\leftrightarrow\; t[x/s] \in \|A[x/s]\|$,

   (b) $\mathsf{PA} \vdash (t \in |A|)[x/s] \;\leftrightarrow\; t[x/s] \in |A[x/s])|$.

   Here, $B[x/s]$ is a substitution of s into the free occurrences of x in B.

2. For every terms s, t and $\mathcal{L}_R$-formula A, we have:

   (a) $\mathsf{PA} \vdash s = t \to (x \in \|A(s)\| \to x \in \|A(t)\|)$.

   (b) $\mathsf{PA} \vdash s = t \to (x \in |A(s)| \to x \in |A(t)|)$.

The next lemma shows that explicit realisability and formal realisability are equivalent for $\mathcal{L}$.

Lemma 114. 1. Let $P(\overrightarrow{y})$ be an atomic formula of $\mathcal{L} \cup \{\perp\!\!\!\perp\}$.

   (a) $\mathsf{GCR} \vdash x\mathrm{F}\ulcorner P(\dot{\overrightarrow{y}})\urcorner \leftrightarrow x \in \|P(\overrightarrow{y})\|$.

   (b) $\mathsf{GCR} \vdash x\mathrm{T}\ulcorner P(\dot{\overrightarrow{y}})\urcorner \leftrightarrow x \in |P(\overrightarrow{y})|$.

2. For every terms s, t, we have:

   (a) $\mathsf{GCR} \vdash x \in |s\mathrm{F}t| \leftrightarrow x\mathrm{T}(s\dot{\in}\|t\|)$,

   (b) $\mathsf{GCR} \vdash x \in |s\mathrm{T}t| \leftrightarrow x\mathrm{T}(s\dot{\in}|t|)$.

Proof.

1. As the proof for the second item is similar, we consider only the first. So, taking any x and y, we derive in GCR as follows:

$$\begin{aligned} x\mathrm{F}\ulcorner P(\dot{\overrightarrow{y}})\urcorner &\Leftrightarrow P(\overrightarrow{y}) \to x \in \perp\!\!\!\perp && \text{(by } (\mathsf{GCR}_{P1}) \text{ and } (\mathsf{GCR}_{P2})\text{)} \\ &\Leftrightarrow x \in \|P(\overrightarrow{y})\|. && \text{(by def. of } \|P\|\text{)} \end{aligned}$$

2. We consider the first item. We derive in GCR as follows:

$$\begin{aligned} x \in |s\mathrm{F}t| &\Leftrightarrow \forall a(a \in \|s\mathrm{F}t\| \to \langle x, a\rangle \in \perp\!\!\!\perp) && \text{(by def. of } |\mathrm{F}|) \\ &\Leftrightarrow \forall a(a\mathrm{F}(s\dot{\in}\|t\|) \to \langle x, a\rangle \in \perp\!\!\!\perp) && \text{(by def. of } \|\mathrm{F}\|) \\ &\Leftrightarrow x\mathrm{T}(s\dot{\in}\|t\|). && \text{(by } (\mathsf{Ax_T})) \end{aligned}$$

■

Definition 115. (FSR) An $\mathcal{L}_R$-theory FSR (Friedman–sheard theory for realisability) consists of GCR with the two rules.

$$\frac{s \in |A|}{s\mathrm{T}\ulcorner A \urcorner}\ (\mathsf{NECR}) \qquad \frac{s\mathrm{T}\ulcorner A \urcorner}{s \in |A|}\ (\mathsf{CONECR}),$$

for every closed term s and every $\mathcal{L}_R$-sentence A.

4.5.2 Revision Model for FSR

Before the proof-theoretic consideration of FSR, we give a model of FSR. Similar to Proposition 8, we construct a revisional model, in which the evaluation of each sentence changes at each step.

First, we fix any pole $\perp\!\!\!\perp \subseteq \mathbb{N}$. Then, given a set $\mathbb{F}_{\perp\!\!\!\perp} \subseteq \mathbb{N} \times \mathbb{N}$ for the interpretation of F, the interpretation $\mathbb{T}_{\perp\!\!\!\perp} \subseteq \mathbb{N} \times \mathbb{N}$ of T can be defined by $\perp\!\!\!\perp$ and $\mathbb{F}_{\perp\!\!\!\perp}$:

$$\mathbb{T}_{\perp\!\!\!\perp} := \{(n, \ulcorner A \urcorner) \in \mathbb{N} \times Sent_{\mathcal{L}_R} \mid \forall m \in \mathbb{N}((m, \ulcorner A \urcorner) \in \mathbb{F}_{\perp\!\!\!\perp} \Rightarrow \langle n, m\rangle \in \perp\!\!\!\perp)\}.$$

Thus, for a set $\mathbb{F}_{\perp\!\!\!\perp}$, we shall express the above $\mathcal{L}_R$-model simply by $\langle \mathbb{F}_{\perp\!\!\!\perp}\rangle$.

For a set $\mathbb{F}_{\perp\!\!\!\perp} \subseteq \mathbb{N} \times \mathbb{N}$, the revision operator Γ is defined as follows:

$$\Gamma(\mathbb{F}_{\perp\!\!\!\perp}) := \{(n, \ulcorner A \urcorner) \in \mathbb{N} \times Sent_{\mathcal{L}_R} \mid \langle \mathbb{F}_{\perp\!\!\!\perp}\rangle \models n \in \|A\|\}.$$

Then, we inductively define a series of models $\Gamma^n(\mathbb{F}_{\perp\!\!\!\perp})$:

$$\begin{aligned} \Gamma^0(\mathbb{F}_{\perp\!\!\!\perp}) &:= \mathbb{F}_{\perp\!\!\!\perp}, \\ \Gamma^{n+1}(\mathbb{F}_{\perp\!\!\!\perp}) &:= \Gamma(\Gamma^n(\mathbb{F}_{\perp\!\!\!\perp})). \end{aligned}$$

Proposition 116. Take any $\mathbb{F}_{\perp\!\!\!\perp} \subseteq \mathbb{N} \times \mathbb{N}$ and assume $\mathsf{FSR} \vdash A$ for an $\mathcal{L}_R$-formula. Then, there exists a number n such that for all $m \geq n$, it holds that $\langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models A^\forall$ where $A^\forall$ is a universal closure of A.

Proof. The proof is by induction on the derivation, so assume $\mathsf{FSR} \vdash A$. We divide the cases by the last axiom or rule of the derivation.

$(\mathsf{Ax_T})$ Suppose that A is the axiom $(\mathsf{Ax_T})$:

$$\forall \ulcorner B \urcorner \in Sent_{\mathcal{L}_R}\{a\mathrm{T}\ulcorner B \urcorner \leftrightarrow \forall b(b\mathrm{F}\ulcorner B \urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp)\},$$

then for all $m \geq 1$, we can show $\langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models A^{\forall}$. Taking any $m \geq 1$, $\ulcorner B \urcorner \in Sent_{\mathcal{L}_R}$, and $a \in \mathbb{N}$, we have:

$$\begin{aligned}\langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models a\mathrm{T}\ulcorner B \urcorner &\Leftrightarrow \forall b \in \mathbb{N}((b, \ulcorner B \urcorner) \in \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp}) \Rightarrow \langle a, b\rangle \in \perp\!\!\!\perp) \\ &\Leftrightarrow \langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models \forall b(b\mathrm{F}\ulcorner B \urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp).\end{aligned}$$

Thus, it follows that $\langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models (\mathsf{Ax_T})^{\forall}$.

(NECR) Suppose that $A :\equiv s\mathrm{T}\ulcorner B \urcorner$ is obtained from $s \in |B|$. Then, by induction hypothesis we have a number n such that for all $m \geq n$,

$$\begin{aligned}&\langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models s \in |B| \\ &\Leftrightarrow \langle \Gamma^m(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models \forall b(b \in \|B\| \to \langle s, b\rangle \in \perp\!\!\!\perp) \\ &\Leftrightarrow \langle \Gamma^{m+1}(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models \forall b(b\mathrm{F}\ulcorner B \urcorner \to \langle s, b\rangle \in \perp\!\!\!\perp).\end{aligned}$$

Since $\langle \Gamma^{m+1}(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models (\mathsf{Ax_T})^{\forall}$, we also get $\langle \Gamma^{m+1}(\mathbb{F}_{\perp\!\!\!\perp})\rangle \models s\mathrm{T}\ulcorner B \urcorner$, which is true for all $m \geq n$.

The other cases are similarly proved. ■

4.5.3 Proof Theory of FSR

In this subsection, we show that FSR is *self-realisable*, that is, for any theorem A of FSR, we can find a term s such that $\mathsf{FSR} \vdash s \in |A|$ (Theorem 125). Finally, conservativity of FSR (Proposition 126) and some open problems are mentioned.

Lemma 117. (Realisation of PA) For every $\mathcal{L}_R$-formula A, if $\mathsf{PA} \vdash A$, then there exists a term s such that $\mathsf{GCR} \vdash s \in |A|$.

Proof. By induction on the length of the derivation of A, we can give the required term in the same way as the proof of Lemma 96. ■

Lemma 118. (Realisation of $(\mathsf{Ax}_{\perp\!\!\!\perp})$) There exists a term s such that $\mathsf{GCR} \vdash s \in |(\mathsf{Ax}_{\perp\!\!\!\perp})|$.

Proof. Taking any a, b and c, we want to find a term that realises the following:

$$T(a,b,c) \to (U(c) \in \perp\!\!\!\perp \to \langle a,b\rangle \in \perp\!\!\!\perp).$$

We will show that a term $f := \lambda u.\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle$ realises A. That is, we want to get:

$$f \in |T(a,b,c) \to (U(c) \in \perp\!\!\!\perp \to \langle a,b\rangle \in \perp\!\!\!\perp)|.$$

Take any term u which refutes A, that is,

$$u \in \|T(a,b,c) \to (U(c) \in \perp\!\!\!\perp \to \langle a,b\rangle \in \perp\!\!\!\perp)\|.$$

Thus, we need to show $\langle f, u\rangle \in \perp\!\!\!\perp$. For this, we prove that $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$. Then, we have $\langle f, u\rangle \in \perp\!\!\!\perp$ by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$.

By the definition of $\|A\|$, we have:

$$(u)_0 \in |T(a,b,c)|, \tag{4.1}$$
$$(u)_1 \in \|U(c) \in \perp\!\!\!\perp \to \langle a,b\rangle \in \perp\!\!\!\perp\|. \tag{4.2}$$

Similarly, (4.2) is equivalent to:

$$((u)_1)_0 \in |U(c) \in \perp\!\!\!\perp|, \tag{4.3}$$
$$((u)_1)_1 \in \|\langle a,b\rangle \in \perp\!\!\!\perp\|. \tag{4.4}$$

To verify $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, we divide the cases by whether $T(a,b,c)$ and $U(c) \in \perp\!\!\!\perp$ hold or not.

- If $T(a,b,c)$ is not true, then we have $\langle ((u)_1)_0, ((u)_1)_1\rangle \in \|T(a,b,c)\|$ and thus by (4.1), we also get $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, as required.
- Thus, we can assume that $T(a,b,c)$ is true. Then, by (4.1) we have $\forall w(w \in \perp\!\!\!\perp \to \langle (u)_0, w\rangle \in \perp\!\!\!\perp)$. In addition, if $U(c) \in \perp\!\!\!\perp$ is not true, then $\langle ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$ by (4.3), and hence $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$.
- Thus, we can further suppose that $U(c) \in \perp\!\!\!\perp$ is true, then we similarly have $\forall w(w \in \perp\!\!\!\perp \to \langle ((u)_1)_0, w\rangle \in \perp\!\!\!\perp)$. Moreover, it follows that $\langle a,b\rangle \in \perp\!\!\!\perp$ by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$. Thus, from (4.4) we have $((u)_1)_1 \in \perp\!\!\!\perp$, which implies $\langle ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, which in turn yields $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$.

Therefore, in any case, we obtain $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$. ■

Lemma 119. (Realisation of $(\mathsf{Ax}_{\mathrm{T}})$) There exists a term s such that $\mathsf{GCR} \vdash s \in |(\mathsf{Ax}_{\mathrm{T}})|$.

Proof. Taking any x, we want to find a realiser for the following:

$$Sent_{\mathcal{L}_R}(x) \to [a\mathrm{T}x \leftrightarrow \forall b(b\mathrm{F}x \to \langle a, b\rangle \in \perp\!\!\!\perp)].$$

If $Sent_{\mathcal{L}_R}(x)$ does not hold, the above formula is derivable in PA. Thus, by Lemma 117, we can primitive recursively give a realiser for the formula. Therefore, taking any code $\ulcorner A \urcorner$ of an $\mathcal{L}_R$-sentence, it suffices to consider the following:

$$a\mathrm{T}\ulcorner A \urcorner \leftrightarrow \forall b(b\mathrm{F}\ulcorner A \urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp).$$

As the realisability is closed under logic, we separate the formula into the following:

1. $a\mathrm{T}\ulcorner A \urcorner \to (b\mathrm{F}\ulcorner A \urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp)$ for any fixed b;
2. $\forall b(b\mathrm{F}\ulcorner A \urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp) \to a\mathrm{T}\ulcorner A \urcorner$.

1. Take any $u \in \|a\mathrm{T}\ulcorner A \urcorner \to (b\mathrm{F}\ulcorner A \urcorner \to \langle a, b\rangle \in \perp\!\!\!\perp)\|$, then we have:

$$(u)_0 \in |a\mathrm{T}\ulcorner A \urcorner|, \tag{4.5}$$
$$((u)_1)_0 \in |b\mathrm{F}\ulcorner A \urcorner|, \tag{4.6}$$
$$((u)_1)_1 \in \|\langle a, b\rangle \in \perp\!\!\!\perp\|. \tag{4.7}$$

Then, we want to find a procedure f such that $\langle f, u\rangle \in \perp\!\!\!\perp$.
From (4.5) and (4.6), we have:

$$(u)_0\mathrm{T}\ulcorner \forall b(b \in \|A\| \to \langle \dot{a}, b\rangle \in \perp\!\!\!\perp \urcorner, \tag{4.8}$$
$$((u)_1)_0\mathrm{T}\ulcorner \dot{b} \in \|A\|\urcorner. \tag{4.9}$$

Thus, by Lemma 94, we obtain:

$$(\mathsf{k}_{\to} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, b\rangle, ((u)_1)_0\rangle)\mathrm{T}\ulcorner \langle \dot{a}, \dot{b}\rangle \in \perp\!\!\!\perp\urcorner. \tag{4.10}$$

From (4.7) and Lemma 114, we have:

$$((u)_1)_1\mathrm{F}\ulcorner \langle \dot{a}, \dot{b}\rangle \in \perp\!\!\!\perp\urcorner. \tag{4.11}$$

Thus by (4.10), (4.11), and the axiom $(\mathsf{Ax}_{\mathrm{T}})$, we have

$$\langle \mathsf{k}_{\to} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, b\rangle, ((u)_1)_0\rangle, ((u)_1)_1\rangle \in \perp\!\!\!\perp.$$

Now, let f be such that $f \cdot b :\simeq \lambda u.\langle \mathsf{k}_{\to} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, b\rangle, ((u)_1)_0\rangle, ((u)_1)_1\rangle$, then we can show $\langle f \cdot b, u\rangle \in \perp\!\!\!\perp$ with the help of $(\mathsf{Ax}_{\perp\!\!\!\perp})$.

2. Take any u such that:

$$(u)_0 \in |\forall b(b\mathrm{F}\ulcorner A \urcorner \rightarrow \langle a, b \rangle \in \perp\!\!\!\perp)|, \tag{4.12}$$
$$(u)_1 \in \|a\mathrm{T}\ulcorner A \urcorner\|. \tag{4.13}$$

Then, we want to find a procedure g such that $\langle g, u \rangle \in \perp\!\!\!\perp$.
From (4.12), we have for all b and w,

$$w \in |b\mathrm{F}\ulcorner A \urcorner| \rightarrow (\mathsf{k}_{\rightarrow} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, b \rangle, w \rangle) \in |\langle a, b \rangle \in \perp\!\!\!\perp|. \tag{4.14}$$

From (4.13) we have:

$$(u)_1\mathrm{F}\ulcorner \forall b(b \in \|A\| \rightarrow \langle \dot{a}, b \rangle \in \perp\!\!\!\perp)\urcorner. \tag{4.15}$$

Thus, by the axioms $(\mathsf{GCR}_{\forall})$ and $(\mathsf{GCR}_{\rightarrow})$, it follows that:

$$(((u)_1)_1)_0\mathrm{T}\ulcorner ((\dot{u})_1)_0 \in \|A\|\urcorner, \tag{4.16}$$
$$(((u)_1)_1)_1\mathrm{F}\ulcorner \langle \dot{a}, ((\dot{u})_1)_0 \rangle \in \perp\!\!\!\perp\urcorner. \tag{4.17}$$

By (4.14) and (4.16), we obtain:

$$(\mathsf{k}_{\rightarrow} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, ((u)_1)_0 \rangle, (((u)_1)_1)_0 \rangle) \in |\langle a, ((u)_1)_0 \rangle \in \perp\!\!\!\perp|. \tag{4.18}$$

By (4.18) and Lemma 114, we also have:

$$(\mathsf{k}_{\rightarrow} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, ((u)_1)_0 \rangle, (((u)_1)_1)_0 \rangle)\mathrm{T}\ulcorner \langle \dot{a}, ((\dot{u})_1)_0 \rangle \in \perp\!\!\!\perp\urcorner. \tag{4.19}$$

Therefore, by (4.17) and (4.19), we obtain a contradictory pair:

$$\langle \mathsf{k}_{\rightarrow} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, ((u)_1)_0 \rangle, (((u)_1)_1)_0 \rangle, (((u)_1)_1)_1 \rangle \in \perp\!\!\!\perp.$$

Thus, define $g := \lambda u.\langle \mathsf{k}_{\rightarrow} \cdot \langle \mathsf{k}_{sub} \cdot \langle (u)_0, ((u)_1)_0 \rangle, (((u)_1)_1)_0 \rangle, (((u)_1)_1)_1 \rangle$.

■

Lemma 120. (Realisation of (GCR_{reg})) There exists a term s such that $\mathsf{GCR} \vdash s \in |(\mathsf{GCR}_{reg})|$.

Proof. Here, the exact form of the formula $Eq(y, z)$ is required. Since $Eq(y, z)$ is Σ_1-formula, we shall rephase it in the form $\exists x Eq'(x, y, z)$, where $Eq'(x, y, z)$ is some atomic predicate of $\mathcal{L}$.

Take any $x \in \mathbb{N}$, any $\ulcorner s = t \urcorner, \ulcorner A(0) \urcorner \in Sent_{\mathcal{L}_R}$, and any $a \in \mathbb{N}$. Then, similar to the proof of Lemma 119, it suffices to find a realiser of the following:

$$Eq'(x, \ulcorner s \urcorner, \ulcorner t \urcorner) \rightarrow (a\mathrm{F}\ulcorner A(s) \urcorner \rightarrow a\mathrm{F}\ulcorner A(t) \urcorner).$$

Thus, we take any u such that:

$$(u)_0 \in |Eq'(x, \ulcorner s \urcorner, \ulcorner t \urcorner)|, \tag{4.20}$$
$$((u)_1)_0 \in |a\mathrm{F}\ulcorner A(s)\urcorner|, \tag{4.21}$$
$$((u)_1)_1 \in \|a\mathrm{F}\ulcorner A(t)\urcorner\|. \tag{4.22}$$

We will show that $f := \lambda u.\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle$ satisfies $\langle f, u\rangle \in \perp\!\!\!\perp$. So, by the axiom $(\mathsf{Ax}_{\perp\!\!\!\perp})$, it is enough to prove $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$.

If the atomic $\mathcal{L}$-formula $Eq'(x, \ulcorner s \urcorner, \ulcorner t \urcorner)$ is false, then by (4.20), $(u)_0$ contradics any number, that is, we have $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, as required. Therefore, we assume $Eq'(x, \ulcorner s \urcorner, \ulcorner t \urcorner)$ is true. Hence, (4.20) implies that in order to show $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, we need to prove $\langle ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$.

By (4.21) and (4.22), we respectively obtain:

$$((u)_1)_0 \mathrm{T}\ulcorner \dot{a} \in \|A(s)\|\urcorner, \tag{4.23}$$
$$((u)_1)_1 \mathrm{F}\ulcorner \dot{a} \in \|A(t)\|\urcorner. \tag{4.24}$$

Since we now assume $Eq'(x, \ulcorner s \urcorner, \ulcorner t \urcorner)$, by (4.24) and the axiom GCR_{reg}, we obtain:

$$((u)_1)_1 \mathrm{F}\ulcorner \dot{a} \in \|A(s)\|\urcorner. \tag{4.25}$$

Therefore, from (4.23), (4.25), and the axiom $(\mathsf{GCR}_{\mathrm{T}})$, it follows that $\langle ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, which implies $\langle (u)_0, ((u)_1)_0, ((u)_1)_1\rangle \in \perp\!\!\!\perp$, as desired. ■

Lemma 121. (Realisation of (GCR_{P1})) There exists a term s such that $\mathsf{GCR} \vdash s \in |(\mathsf{GCR}_{P1})|$.

Proof. Take any u such that:

$$(u)_0 \in |\neg P(\overrightarrow{x})|, \tag{4.26}$$
$$(u)_1 \in \|a\mathrm{F}\ulcorner P(\dot{\overrightarrow{x}})\urcorner\|. \tag{4.27}$$

In order to show that $f := \lambda u.\langle (u)_0, (u)_1\rangle$ satisfies $\langle f, u\rangle \in \perp\!\!\!\perp$, we prove $\langle (u)_0, (u)_1\rangle \in \perp\!\!\!\perp$.

(4.26) is equivalent to:

$$\forall w((w)_0 \in |P(\overrightarrow{x})| \to [(w)_1 \in \|\bot\| \to \langle (u)_0, w\rangle \in \perp\!\!\!\perp]), \tag{4.28}$$

where recall that $\neg A :\equiv A \to \bot$, and the falsehood $\bot$ is defined as some false equation. Moreover, from (4.27), Lemm 114, and the axiom $(\mathsf{GCR}_{\to})$, we can deduce the following:

$$\begin{aligned}(u)_1 \in \|a\mathrm{F}\ulcorner P(\dot{\overrightarrow{x}})\urcorner\| &\Leftrightarrow (u)_1 \mathrm{F}\ulcorner P(\dot{\overrightarrow{x}}) \to \dot{a} \in \perp\!\!\!\perp\urcorner \\ &\Leftrightarrow ((u)_1)_0 \mathrm{T}\ulcorner P(\dot{\overrightarrow{x}})\urcorner \wedge ((u)_1)_1 \mathrm{F}\ulcorner \dot{a} \in \perp\!\!\!\perp\urcorner \\ &\Rightarrow ((u)_1)_0 \in |P(\overrightarrow{x})| \wedge ((u)_1)_1 \in \|\bot\|.\end{aligned} \tag{4.29}$$

Thus, by (4.28) and (4.29), we have $\langle (u)_0, (u)_1\rangle \in \perp\!\!\!\perp$. ■

Lemma 122. (Realisation of (GCR_{P2})) There exists a term s such that $\mathsf{GCR} \vdash s \in |(\mathsf{GCR}_{P2})|$.

Proof. The proof is similar to for (GCR_{P1}). ■

Lemma 123. (Realisation of $(\mathsf{GCR}_{\to})$) For any number a and any code $\ulcorner A \to B\urcorner$ of an $\mathcal{L}_R$-sentence, all the following are realisable in GCR:

1. $a\mathrm{F}\ulcorner A \to B\urcorner \to (a)_0\mathrm{T}\ulcorner A\urcorner$,
2. $a\mathrm{F}\ulcorner A \to B\urcorner \to (a)_1\mathrm{F}\ulcorner B\urcorner$,
3. $(a)_0\mathrm{T}\ulcorner A\urcorner \to [(a)_1\mathrm{F}\ulcorner B\urcorner \to a\mathrm{F}\ulcorner A \to B\urcorner]$.

Therefore, we can give a term that realises $(\mathsf{GCR}_{\to})$.

Proof.

1. Take any u such that $(u)_0 \in |a\mathrm{F}\ulcorner A \to B\urcorner|$ and $(u)_1 \in \|(a)_0\mathrm{T}\ulcorner A\urcorner\|$, then we have $(u)_0\mathrm{T}\ulcorner\dot{a} \in \|A \to B\|\urcorner$ and $(u)_1\mathrm{F}\ulcorner(\dot{a})_0 \in |A|\urcorner$. By Lemma 96, we obtain a number w such that $w\mathrm{T}\ulcorner(\dot{a})_0 \in |A| \wedge (\dot{a})_1 \in \|B\| \to (\dot{a})_0 \in |A|\urcorner$, from which and $(u)_0\mathrm{T}\ulcorner(\dot{a})_0 \in |A| \wedge (\dot{a})_1 \in \|B\|\urcorner$, we get $(\mathsf{k}_{\to} \cdot \langle w, (u)_0\rangle)\mathrm{T}\ulcorner(\dot{a})_0 \in |A|\urcorner$. Thus, $\langle\langle \mathsf{k}_{\to} \cdot \langle w, (u)_0\rangle, (u)_1\rangle \in \perp\!\!\!\perp$ by the axiom $(\mathsf{Ax}_{\mathrm{T}})$. Therefore, the term $f := \lambda u.\langle\langle \mathsf{k}_{\to} \cdot \langle w, (u)_0\rangle, (u)_1\rangle$ satisfies $\langle f, u\rangle \in \perp\!\!\!\perp$.
2. Similarly to the proof of the item 1.
3. Take any u such that:
$$(u)_0\mathrm{T}\ulcorner(\dot{a})_0 \in |A|\urcorner, \tag{4.30}$$
$$((u)_1)_0\mathrm{T}\ulcorner(\dot{a})_1 \in \|B\|\urcorner, \tag{4.31}$$
$$((u)_1)_1\mathrm{F}\ulcorner\dot{a} \in \|A \to B\|\urcorner. \tag{4.32}$$
By Lemma 96, we obtain a number w such that
$$w\mathrm{T}\ulcorner(\dot{a})_0 \in |A| \to ((\dot{a})_1 \in \|B\| \to [(\dot{a})_0 \in |A| \wedge (\dot{a})_1 \in \|B\|])\urcorner. \tag{4.33}$$
Thus, from (4.30) and (4.31), we obtain:
$$(\mathsf{k}_{\to} \cdot \langle \mathsf{k}_{\to} \cdot \langle w, (u)_0\rangle, ((u)_1)_0\rangle)\mathrm{T}\ulcorner(\dot{a})_0 \in |A| \wedge (\dot{a})_1 \in \|B\|\urcorner. \tag{4.34}$$
On the other hand, (4.32) is equivalent to:
$$((u)_1)_1\mathrm{F}\ulcorner(\dot{a})_0 \in |A| \wedge (\dot{a})_1 \in \|B\|\urcorner. \tag{4.35}$$
Thus, we have $\langle \mathsf{k}_{\to} \cdot \langle \mathsf{k}_{\to} \cdot \langle w, (u)_0\rangle, ((u)_1)_0\rangle, ((u)_1)_1\rangle \in \perp\!\!\!\perp$. Therefore, we can employ $g := \lambda u.\langle \mathsf{k}_{\to} \cdot \langle \mathsf{k}_{\to} \cdot \langle w, (u)_0\rangle, ((u)_1)_0\rangle, ((u)_1)_1\rangle$ as a required realiser.

■

Lemma 124. (Realisation of $(\mathsf{GCR}_\forall)$) There exists a term s such that $\mathsf{GCR} \vdash s \in |(\mathsf{GCR}_\forall)|$.

Proof. The proof is similar to for $(\mathsf{GCR}_\rightarrow)$. ■

Theorem 125. For each $\mathcal{L}_R$-formula A, if $\mathsf{FSR} \vdash A$, then there exists a term s such that $\mathsf{FSR} \vdash s \in |A|$.

Proof. The proof is by induction on the derivation length of A. Assume that $\mathsf{FSR} \vdash A$. If A is obtained by the axioms of GCR, then we already have an answer by the above lemmata. Thus, the remaining cases are (NECR) and (CONECR).

(NECR) Assume that $A :\equiv s\mathrm{T}\ulcorner B \urcorner$ is derived from $s \in |B|$. Then, by induction hypothesis, we have some term t such that $\mathsf{FSR} \vdash t \in |s \in |B||$. Thus, by (NECR) we deduce $t\mathrm{T}\ulcorner \dot{s} \in |B| \urcorner$, which is $t \in |s\mathrm{T}\ulcorner B \urcorner|$, and hence t is a required term.

(CONECR) Assume that $A :\equiv s \in |B|$ is derived from $s\mathrm{T}\ulcorner B \urcorner$. Then, by induction hypothesis, we have some term t such that $\mathsf{FSR} \vdash t \in |s\mathrm{T}\ulcorner B \urcorner|$. Thus, by (CONECR) we deduce $t \in |s \in |B||$, and hence, t is a required term.

■

Finally, we mention the proof-theoretic strength of FSR.

Proposition 126. FSR is conservative over PA.

Proof. The same translation $\mathcal{T}$ as in Proposition 97 gives an interpretation of FSR into PA. Note that we trivially have $\mathsf{PA} \vdash \mathcal{T}(s \in |A|)$ for every term s and formula A. ■

Definition 127. (GCR^+) An $\mathcal{L}_R$-theory GCR^+ is the extension of GCR with the reflection rule:

$$\frac{s\mathrm{T}\ulcorner A \urcorner}{A},$$

for every closed term s and every $\mathcal{L}$-sentence A.

Note that the sentence A in the reflection rule of GCR^+ is restricted to $\mathcal{L}$. Thanks to this condition, we can easily extend the soundness result (Proposition 116) to GCR^+. Now, let GCR^{++} be an extension of GCR^+ such that the reflection rule allows arbitrary $\mathcal{L}_R$-sentences. In contrast to GCR^+, it is no longer clear whether Proposition 116 holds for GCR^{++}, i.e., GCR^{++} is sound with respect to any choice of the pole, or not.

We conclude this chapter by listing two open problems on the proof-theoretic strength.

Conjecture 128.
1. GCR^+ is proof-theoretically equivalent to FS.
2. GCR^{++} is consistent. In particular, it is proof-theoretically equivalent to FS.

.

Chapter 5
Universes for Frege Structure

In this chapter, we turn to the question of how strong an axiomatic truth theory can be obtained in a bottom-up manner. For this purpose, this chapter considers theories of *Frege structure*. The framework of Frege structure, which is introduced by Aczel [1], is essentially a model of lambda calculus augmented with the notions of *truth* and *proposition*. While Aczel's study was model-theoretic, Frege structure can be seen as an axiomatic theory of truth formulated over an applicative theory (see Section 5.1). In addition, depending on what kinds of truth and proposition are assumed, various theories of Frege structure have been proposed (for an overview, see, e.g., [14]). To analyse Frege structure proof-theoretically, it is useful to use results of another applicative framework, *explicit mathematics* [17]. Explicit mathematics is classified as an intensional set theory, in which *types*, second-order objects for sets, are generated by individual terms, called *names*.

Although these two frameworks are different at a glance, there are, as Aczel anticipated [1, p. 34], various technical correspondences between them. In particular, for various systems of explicit mathematics, we can find proof-theoretically equivalent theories of Frege structure, as displayed in Table 5, where EM is a system of explicit mathematics (see Definition 131); EMU is an extension of EM by *universes* (cf. [60]); NEM is an extension of EM by *name induction* (cf. [43]); T_0 is an extension of EM by *inductive generations* (see Definition 133); KF and PT are theories of Frege structure based on Kripke-Feferman logic and Aczel-Feferman logic, respectively (see Definition 135 and 156); KFU and PTU are respectively extensions of KF and PT by *universes* (see Definition 137 and 159); and VF is a theory of Frege structure based on supervaluation logic (see Definition 171). The table also contains the corresponding subsystems of second-order arithmetic and their proof-theoretic ordinals. The system Σ^1_1-AC has the schema Σ^1_1 axiom of choice; $\mathsf{ATR} + (\Sigma^1_1\text{-}\mathsf{DC})$ consists of the arithmetical transfinite recursion ATR with the Σ^1_1 dependent choice; Π^1_1-CA^-_0 is the parameter-free Π^1_1 comprehension schema; and Δ^1_2-$\mathsf{CA} + \mathsf{BI}$ is the Δ^1_2 comprehension schema with the bar induction. For details of their proof-theoretic ordinals, see, e.g., [37, 53]).

Table 5.1: Applicative theories

Ordinal strength	Explicit mathematics	Frege structure	Second-order arithmetic
$\psi_\Omega(\varepsilon_{I+1})$	T_0	?	Δ^1_2-$\mathsf{CA}+\mathsf{BI}$
$\psi_\Omega(\varepsilon_{\Omega+1})$	NEM	VF	Π^1_1-CA^-_0
$\varphi 1 \varepsilon_0 0$	EMU	KFU, PTU	$\mathsf{ATR}+(\Sigma^1_1$-$\mathsf{DC})$
$\varphi \varepsilon_0 0$	EM	KF, PT	Σ^1_1-AC

As the table shows, the correspondence between explicit mathematics and Frege structure has so far been obtained only up to the strength of $(\Pi^1_1\text{-}\mathsf{CA})^-_0$. Therefore, this chapter aims to provide well-motivated theories of Frege structure proof-theoretically equivalent to T_0. This task is also in line with our purpose of obtaining strong truth theories. As far as the author knows, Fujimoto's system Aut(VF) [23], which is essentially a significant extension of VF in Definition 13, is so far the strongest among well-motivated truth theories. Since the strength of Aut(VF) lies strictly between Δ^1_2-CA and Δ^1_2-$\mathsf{CA}+\mathsf{BI}$, the author believes that the theories proposed in this chapter break the record, though our base theory is not PA. [1]

There is another reason to seek for stronger truth theories. Halbach [29] argues that an expressively strong truth theory can, to some extent, reduce ontological assumptions on sets to semantic assumptions. Thus, perhaps such a theory, if it is well motivated, can take the place of set theory as a foundation for a large part of mathematics. Moreover, since T_0 is expressively rich enough to interpret various set theories (cf. [44, 63, 64]), we can expect our theories of Frege structure to contribute to Halbach's programme.

The structure of this chapter is as follows: In Section 5.1, we define the total applicative theory TON as the common base theory of explicit mathematics and Frege structure. We also define the theories of explicit mathematics, EM and its extension T_0. In Section 5.2, following Kahle's formulation [40, 41], we introduce Kripke-Feferman-style theories of Frege structure, KF and its extension KFU. In Section 5.3, we expand KFU by the *proposition induction* schema, to obtain a theory KFUPI that is as strong as T_0. In Section 5.4, we consider Aczel-Feferman-style theories of Frege structure, PT and its extension PTU. In conclusion, we can obtain a theory PTUPI that has the same strength as T_0. In Section 5.5, we formulate an alternative principle *least universes* schema, and then we show that this also gives the strength of T_0 to both KFU and PTU. In Section 5.6, we introduce a supervaluation-style Frege structure essentially based on Kahle's

[1] However, the author must mention that Cantini gave a recursion-theoretically motivated system of Frege structure which is at least as strong as T_0 [14, p. 253]. Over Peano arithmetic, Schindler [56] formulated a disquotational truth theory of the same consistency strength as the full second-order arithmetic.

formulation [40]. Then, following Kahle's suggestion [40, p. 124], we extend VF by universes and prove that the resulting theory VFU is proof-theoretically equivalent to T_0. Therefore, one of Kahle's open questions is solved.

The content of this chapter is based on [32].

5.1 Explicit Mathematics

This section defines Feferman's system T_0 of explicit mathematics. We first define the total applicative theory TON, which, in this chapter, is used as the common base theory of explicit mathematics and Frege structure.

5.1.1 Total Applicative Theory

The first-order language $\mathcal{L}$ for the total operational theory (TON) (see, e.g. [41]) consists of the standard logical symbols, individual variables ($x, y, z, x_0, x_1, \cdots, a, b, c, f, g, h, \cdots$), individual constants k, s (combinators), $\mathsf{p}, \mathsf{p}_0, \mathsf{p}_1$ (pairing and projections), $\mathsf{0}$ (zero), $\mathsf{s_N}$ (successor), $\mathsf{p_N}$ (predecessor), $\mathsf{d_N}$ (definition by numerical cases), a binary function symbol $\mathsf{App}(x, y)$ (application), and a unary predicate $\mathrm{N}(x)$ (natural numbers). Formulae of $\mathcal{L}$ are constructed from atomic formulae by the logical symbols $\neg A$, $A \wedge B$, $A \to B$, and $\forall x.\ A$. We assume that $\vee$ and $\exists$ are defined in a standard manner, whereas $\to$ is given as a primitive symbol. The meaning of each symbol is clear from its defining axioms below.

We shall use the following abbreviations in this chapter:

1. $ab := (ab) := \mathsf{App}(a, b)$.
2. $a_1a_2a_3 \dots a_n :\equiv ((\cdots((a_1a_2)a_3\cdots)a_n)$.
3. $(x, y) := \mathsf{p}xy$.
4. $(x)_i := \mathsf{p}_i x$ for $i \in \{0, 1\}$.
5. $s \neq t := \neg(s = t)$.
6. $\forall x_0, x_1, \dots x_n.\ A :\equiv \forall x_0.\forall x_1. \dots \forall x_n.\ A$.

Definition 129. The $\mathcal{L}$-theory TON consists of the following axioms:

- $\mathsf{k}ab = a$,
- $\mathsf{s}abc = ac(bc)$,
- $(a, b)_0 = a \wedge (a, b)_1 = b$,
- $\mathrm{N}(\mathsf{0}) \wedge \forall x.\ \mathrm{N}(x) \to \mathrm{N}(\mathsf{s_N}x)$,
- $\forall x.\ \mathrm{N}(x) \to \mathsf{s_N}x \neq 0 \wedge \mathsf{p_N}(\mathsf{s_N}x) = x$,

- $\forall x.\ \mathrm{N}(x) \land x \neq \mathsf{0} \to \mathrm{N}(\mathsf{p_N} x) \land \mathsf{s_N}(\mathsf{p_N} x) = x$,
- $\mathrm{N}(a) \land \mathrm{N}(b) \land a = b \to \mathsf{d_N} uvab = u$,
- $\mathrm{N}(a) \land \mathrm{N}(b) \land a \neq b \to \mathsf{d_N} uvab = v$,
- $A(\mathsf{0}) \land [\forall x.\ \mathrm{N}(x) \land A(x) \to A(\mathsf{s_N} x)] \to \forall x.\ \mathrm{N}(x) \to A(x)$, for every $\mathcal{L}$-formula A.

In this chapter, we will repeatedly use the following well-known fact :

Fact 130. (cf. [14, p. 16])

λ-abstraction. For each variable x and an $\mathcal{L}$-term t, we can find a term $\lambda x.t$ such that $\mathsf{TON} \vdash (\lambda x.t)x = t$.

Recursion. There is a term rec such that $\mathsf{TON} \vdash \forall f.\ \mathsf{rec} f = f(\mathsf{rec} f)$.

5.1.2 Explicit Mathematics

We define systems of explicit mathematics over TON. Usually, theories of explicit mathematics are defined in second-order language. For simplicity, however, we shall formulate them as first-order ones, similar to [34].

The first-order language $\mathcal{L}_{EM}$ of explicit mathematics is an extension of $\mathcal{L}$ with two predicates: a unary predicate symbol $\mathrm{R}(x)$, meaning that x represents a set, and a binary predicate symbol $x \in y$, meaning that x is contained in y. We define $\forall x \in y.\ A(x)$ to be $\forall x.\ x \in y \to A(x)$. In addition, $\mathcal{L}_{EM}$ has individual constant symbols, called *generators*: int (intersection), j (join), nat (natural numbers), id (identity), co (complements), dom (domain), inv (inversion), and i (inductive generations).

Definition 131. The $\mathcal{L}_{EM}$-theory EM consists of the following axioms:

Natural numbers. $\mathrm{R}(\mathsf{nat}) \land \forall x.\ x \in \mathsf{nat} \leftrightarrow \mathrm{N}(x)$.

Identity. $\mathrm{R}(\mathsf{id}) \land \forall x.\ x \in \mathsf{id} \leftrightarrow \exists y.\ x = (y, y)$.

Complements. $\mathrm{R}(x) \to \mathrm{R}(\mathsf{co}(x)) \land \forall y.\ y \in \mathsf{co}(x) \leftrightarrow y \notin x$.

Intersections. $\mathrm{R}(x) \land \mathrm{R}(y) \to \mathrm{R}(\mathsf{int}(x, y)) \land \forall z.\ z \in \mathsf{int}(x, y) \leftrightarrow z \in x \land z \in y$.

Domains. $\mathrm{R}(a) \to \mathrm{R}(\mathsf{dom}(a)) \land \forall x.\ x \in \mathsf{dom}(a) \leftrightarrow \exists y.\ (x, y) \in a$.

Inverse images. $\mathrm{R}(a) \to \mathrm{R}(\mathsf{inv}(a, f)) \land \forall x.\ x \in \mathsf{inv}(a, f) \leftrightarrow fx \in a$.

Joins. $\mathrm{R}(x) \wedge [\forall y \in x.\ \mathrm{R}(fy)] \to \mathrm{R}(\mathsf{j}(x, f)) \wedge \Sigma(x, f, \mathsf{j}(x, f))$, where

$$\Sigma(x, f, y) :\equiv \forall u.\ u \in y \leftrightarrow \exists v, w.\ u = (v, w) \wedge v \in x \wedge w \in fv.$$

The above axioms explain how a new set is generated from existing ones. For instance, the join axiom says that given a set x and a function f whose domain is x, there exists the disjoint union $\mathsf{j}(x, f)$ of the range of f.

Fact 132. (cf. [6]) EM is proof-theoretically equivalent to Σ^1_1-AC.

5.1.3 Universes and Inductive Generations

A *universe* in explicit mathematics is a set that is closed under the name-generating operations in Definition 131. More formally, the fact that a is a universe is expressed by the formula:

$$\mathrm{U}(a) :\equiv [\forall b.\ \mathcal{C}(a, b) \to b \in a] \to \forall b \in a.\ \mathrm{R}(b),$$

where $\mathcal{C}(a, b)$ is the disjunction of the following:

1. $a = \mathsf{nat} \vee a = \mathsf{id}$,
2. $\exists x.\ b = \mathsf{co}(x) \wedge x \in a$,
3. $\exists x, y.\ b = \mathsf{int}(x, y) \wedge x \in a \wedge y \in a$,
4. $\exists x.\ b = \mathsf{dom}(x) \wedge x \in a$,
5. $\exists f, x.\ b = \mathsf{inv}(x, f) \wedge x \in a$,
6. $\exists f, x.\ b = \mathsf{join}(x, f) \wedge x \in a \wedge \forall y \in x.fy \in a$.

Of course, we require an additional axiom to assure the existence of universes. Here, we introduce the *limit axiom* [60]. For a new constant symbol l, the limit axiom is the following:

$$\mathrm{R}(x) \to \mathrm{R}(\mathsf{l}x) \wedge x \in \mathsf{l}x.$$

Then, the system EMU is defined as EM equipped with the limit axiom. From Strahm's proof [60, Theorem 15], it follows that EMU is proof-theoretically equivalent to $\widehat{\mathsf{ID}}_{<\varepsilon_0}$, the arithmetical fixed-point theory iterated up to ε_0. While the proof-theoretic strength of EMU is beyond the range of predicativity, that is, its proof-theoretic ordinal is larger than the Feferman–Schütte ordinal Γ_0, it is still weaker than impredicative theories, such as the theory ID_1 of arithmetical inductive definitions (for the definition, see [54]). Therefore, the strength of EMU is called *metapredicative* [60].

Feferman [17] formulated a highly strong principle for explicit mathematics called *inductive generation*:

$$\mathrm{R}(a) \wedge \mathrm{R}(b) \to \mathrm{R}(\mathsf{i}(a,b)) \wedge \mathsf{Closed}(a,b,\mathsf{i}(a,b)). \tag{IG.1}$$

$$\mathrm{R}(a) \wedge \mathrm{R}(b) \wedge \mathsf{Closed}(a,b,A) \to \forall x \in \mathsf{i}(a,b).\ A(x), \tag{IG.2}$$

where A is any formula, and

- $y <_b x :\equiv (y,x) \in b$;
- $\mathsf{Closed}(a,b,A(\bullet)) :\equiv \forall x \in a.\ [\forall y \in a.\ y <_b x \to A(y)] \to A(x)$.

Definition 133. The $\mathcal{L}_{EM}$-theory T_0 consists of EM with (IG.1) and (IG.2).

The proof-theoretic strength of T_0 is far beyond EMU.

Fact 134. ([35]) T_0 is proof-theoretically equivalent to Δ^1_2-$\mathsf{CA} + \mathsf{BI}$.

5.2 Frege Structure by Strong Kleene Schema

In this section, we introduce the Kripke–Feferman theory KF of Frege structure and its extension by universes. We first define the base language $\mathcal{L}_{FS}$ over which our theories of Frege structure are formulated.

The language $\mathcal{L}_{FS}$ is $\mathcal{L}$ augmented with the following symbols:

- individual constants $\dot{=}, \dot{\neg}, \dot{\wedge}, \dot{\to}, \dot{\forall}, \dot{\mathsf{N}}, \mathsf{I}$;
- unary predicates $\mathrm{T}(x), \mathrm{U}(x)$.

Here, $\mathrm{T}(x)$ is intended to mean the truth predicate; $\mathrm{U}(x)$ means that x is a *universe* (see below for more details); and the constant I is used to generate universes, similar to the one in explicit mathematics. The other individual constants are used as sentence-constructing operators. For example, the term $x\dot{\wedge}y := \dot{\wedge}(x,y)$ informally denotes (the code of) a conjunctive sentence consisting of x and y. We can similarly understand the terms $\dot{\neg}x$, $\dot{\mathsf{N}}x$, and $\dot{\forall}f$. We also use the notations $(x\dot{=}y) := (\dot{=}(x,y))$ and $(x\dot{\to}y) := (\dot{\to}(x,y))$, whose informal meaning should be clear.

5.2.1 System KF

The theory KF, as a theory of truth, is as defined in Definition 10. Cantini [13] later formulated KF as a theory of Frege structure. Recall that in KF, each sentence is monotonically evaluated based on the Strong Kleene evaluation, as displayed in the following truth table. Note that the conditional $A \to B$ is definable as $\neg A \vee B$.

$\neg$	
T	F
U	U
F	T

$\vee$	T	U	F
T	T	T	T
U	T	U	U
F	T	U	F

$\wedge$	T	U	F
T	T	U	F
U	U	U	F
F	F	F	F

$\rightarrow$	T	U	F
T	T	U	F
U	T	U	U
F	T	T	T

The Kripke–Feferman system, as an $\mathcal{L}_{FS}$-theory, is defined straightforwardly from the above table.

Definition 135. (cf. [14, 15, 40, 41]) The $\mathcal{L}_{FS}$-theory $\mathsf{KF}[\mathsf{TON}]$ has TON with the full induction schema for $\mathcal{L}_{FS}$ and (the universal closure of) the following axioms:

Compositional axioms.

($\mathsf{K}_{=}$) $\mathrm{T}(x\dot{=}y) \leftrightarrow x = y$

($\mathsf{K}_{\neq}$) $\mathrm{T}(\dot{\neg}(x\dot{=}y)) \leftrightarrow x \neq y$

(K_{N}) $\mathrm{T}(\dot{\mathsf{N}}x) \leftrightarrow \mathrm{N}(x)$

($\mathsf{K}_{\neg\mathrm{N}}$) $\mathrm{T}(\dot{\neg}(\dot{\mathsf{N}}x)) \leftrightarrow \neg\mathrm{N}(x)$

($\mathsf{K}_{\neg\neg}$) $\mathrm{T}(\dot{\neg}(\dot{\neg}x)) \leftrightarrow \mathrm{T}(x)$

($\mathsf{K}_{\wedge}$) $\mathrm{T}(x\dot{\wedge}y) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(y)$

($\mathsf{K}_{\neg\wedge}$) $\mathrm{T}(\dot{\neg}(x\dot{\wedge}y)) \leftrightarrow \mathrm{T}(\dot{\neg}x) \vee \mathrm{T}(\dot{\neg}y)$

($\mathsf{K}_{\rightarrow}$) $\mathrm{T}(x\dot{\rightarrow}y) \leftrightarrow \mathrm{T}(\dot{\neg}x) \vee \mathrm{T}(y)$

($\mathsf{K}_{\neg\rightarrow}$) $\mathrm{T}(\dot{\neg}(x\dot{\rightarrow}y)) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}y)$

($\mathsf{K}_{\forall}$) $\mathrm{T}(\dot{\forall}f) \leftrightarrow \forall x.\ \mathrm{T}(fx)$

($\mathsf{K}_{\neg\forall}$) $\mathrm{T}(\dot{\neg}(\dot{\forall}f)) \leftrightarrow \exists x.\ \mathrm{T}(\dot{\neg}(fx))$

Consistency.

($\mathsf{T\text{-}Cons}$) $\neg[\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}x)]$

In this chapter, we express $\mathsf{KF}[\mathsf{TON}]$ simply as KF, for there is no room for confusion with that of Definition 10.

The system KF is proof-theoretically stronger than TON. In particular, KF can relatively interpret EM. To see this, we define a translation $' : \mathcal{L}_{EM} \rightarrow \mathcal{L}_{FS}$. First, the vocabularies of $\mathcal{L}$ are unchanged. Second, we define $(x \in y)'$ to be $\mathrm{T}(yx)$, which can be read as 'y is true at x.' As for the translation of $\mathrm{R}(x)$, we define predicates $\mathrm{C}(f)$ and $\mathrm{P}(x)$, where $\mathrm{P}(x)$ means that x is a *proposition*, that is, x is determined to be true or false; and $\mathrm{C}(f)$ means that f is a *class* (propositional function), that is, fx is a proposition for every object x.

$$\mathrm{C}(f) := \forall x.\ \mathrm{P}(fx) := \forall x.\ \mathrm{T}(fx) \vee \mathrm{T}(\dot{\neg}(fx)).$$

Then, we let $\mathrm{R}'(x) :\equiv \mathrm{C}(x)$. Finally, as is remarked in [14, p. 59], for each generator of $\mathcal{L}_{EM}$, we can find a term such that the translation of the defining axiom of c in EM is derivable in KF (see also the proof of Lemma 177). To summarise, the following is obtained (for the proof, see [14, pp. 57–59]):

Fact 136. For each $\mathcal{L}_{EM}$-formula A, if $\mathsf{EM} \vdash A$, then $\mathsf{KF} \vdash A'$.

We also remark that EM and KF are proof-theoretically equivalent (for the proof, see, e.g., [14, Section 57]).

5.2.2 Universes for KF

In order to strengthen a given truth theory, iterating truth predicates is effective in many cases (cf. [18, 37, 23]). In the framework of Frege structure, an analogous idea is realisable by using the notion of *universes* (cf. [14, 41]). Following [41], we let $f \sqsubset g$ (g *reflects* on f) be the following formula:

$$[\forall x.\ \mathrm{T}(fx) \to \mathrm{T}(g(fx))]\ \wedge\ [\forall x.\ \mathrm{T}(\dot{\neg} fx) \to \mathrm{T}(g(\dot{\neg} fx))].$$

The predicate $f \sqsubset g$ informally means that both positive and negative facts about f are expressed as positive statements within g.

Definition 137. (cf. [40, 41]) The $\mathcal{L}_{FS}$-theory KFU has TON and (the universal closure of) the following axioms:

Compositional axioms in U**.**

($\mathsf{U}_=$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(x \dot{=} y)) \leftrightarrow x = y$

($\mathsf{U}_{\neq}$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(\dot{\neg}(x \dot{=} y))) \leftrightarrow x \neq y$

(U_{N}) $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(u(\dot{\mathsf{N}} x)) \leftrightarrow \mathrm{N}(x)$

($\mathsf{U}_{\neg\mathrm{N}}$) $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(u(\dot{\neg}(\dot{\mathsf{N}} x))) \leftrightarrow \neg\mathrm{N}(x)$

($\mathsf{U}_{\neg\neg}$) $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(u(\dot{\neg}(\dot{\neg} x))) \leftrightarrow \mathrm{T}(ux)$

($\mathsf{U}_{\wedge}$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(x \dot{\wedge} y)) \leftrightarrow \mathrm{T}(ux) \wedge \mathrm{T}(uy)$

($\mathsf{U}_{\neg\wedge}$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(\dot{\neg}(x \dot{\wedge} y))) \leftrightarrow \mathrm{T}(u(\dot{\neg} x)) \vee \mathrm{T}(u(\dot{\neg} y))$

($\mathsf{U}_{\to}$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(x \dot{\to} y)) \leftrightarrow \mathrm{T}(u(\dot{\neg} x)) \vee \mathrm{T}(uy)$

($\mathsf{U}_{\neg\to}$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(\dot{\neg}(x \dot{\to} y))) \leftrightarrow \mathrm{T}(ux) \wedge \mathrm{T}(u(\dot{\neg} y))$

($\mathsf{U}_{\forall}$) $\mathrm{U}(u) \to \forall f.\ \mathrm{T}(u(\dot{\forall} f)) \leftrightarrow \forall x.\ \mathrm{T}(u(fx))$

($\mathsf{U}_{\neg\forall}$) $\mathrm{U}(u) \to \forall f.\ \mathrm{T}(u(\dot{\neg}(\dot{\forall} f))) \leftrightarrow \exists x.\ \mathrm{T}(u(\dot{\neg}(fx)))$

Consistency in U**.**

(U-Cons) $\mathrm{U}(u) \to \forall x.\ \neg[\mathrm{T}(ux) \wedge \mathrm{T}(u(\dot{\neg} x))]$

Structural properties of U**.**

(U-Class) $\mathrm{U}(u) \to \mathrm{C}(u)$

(U-True) $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(ux) \to \mathrm{T}(x)$

(Lim) $\mathrm{C}(f) \to \mathrm{U}(\mathsf{l}f) \wedge f \sqsubset \mathsf{l}f$

In the above definition, the predicate $\mathrm{U}(u)$ indicates that u is a universe. The compositional axioms and the consistency axiom for U ensures that each universe satisfies the axioms of KF; thus, universes work as quasi-truth predicates. Of course, KF-axioms relativised to universes do not have any logical strength unless we do not further postulate the existence of universes. Thus, we added the axiom (Lim) in Definition 137, which generates an infinite series of universes: $u_0 \sqsubset u_1 \sqsubset u_2 \sqsubset \cdots \sqsubset u_n \sqsubset \cdots$. This roughly corresponds to a hierarchy of infinitely iterated truth predicates for KF, whence the strength of KFU exceeds that of KF. Moreover, Kahle [41, Proposition 14] showed that KFU is a proper extension of KF:

Fact 138. KF is a subtheory of KFU.

In particular, KFU is shown to be proof-theoretically equivalent to EMU ([41, Theorem 33]):

Fact 139. KFU and EMU have the same $\mathcal{L}$-theorems.

5.3 Proposition Induction for Strong Kleene: KFUPI

While adding universes makes KF stronger, KFU falls within metapredicativity in strength, so it is still much weaker than impredicative theories such as $\Pi^1_1\text{-}\mathsf{CA}^-_0$. This section aims to propose a stronger principle that gives the same strength as T_0.

For our first attempt, we shall borrow an idea from the system NAI of explicit mathematics in [38]. In the system EMU, names are inductively generated by several axioms, such as join and the limit axiom. The system NAI is then defined as an extension of EMU with induction principles requiring that the whole universe of names only contains those inductively generated by these axioms. Despite lacking inductive generations, NAI is shown to be proof-theoretically equivalent to T_0 [38, Conclusion 25]. Since *class* in Frege structure is an analogue of *name* in explicit mathematics, we can expect to get a theory of Frege structure equivalent to T_0 by considering an induction principle for classes or propositions.

5.3.1 Proposition Induction

How, then, should we inductively characterise *propositions* in KF? For example, let us consider the conjunction. In the Strong Kleene schema, $A \wedge B$ is a proposition when both A and B are propositions. In fact, we can derive $\mathrm{P}(x) \wedge \mathrm{P}(y) \to \mathrm{P}(x \dot{\wedge} y)$ in KF. On the other hand, once we establish the falsity of either A or B, the conjunction $A \wedge B$ is determined to be false regardless of the other conjunct. This could be expressed, for example, as $([\mathrm{P}(x) \wedge \mathrm{T}(\dot{\neg} x)] \vee [\mathrm{P}(y) \wedge \mathrm{T}(\dot{\neg} y)]) \to \mathrm{P}(x \dot{\wedge} y)$, which is indeed provable in KF. So, in KF, we can characterise conjunctive propositions in two ways. Understanding the other logical symbols similarly, we can characterise propositions for KF as follows:

Lemma 140. KF derives the following:

1. $\mathrm{P}(x \dot{=} y) \wedge \mathrm{P}(\dot{\mathsf{N}} x)$,
2. $\mathrm{P}(\dot{\neg} x) \leftrightarrow \mathrm{P}(x)$,
3. $\mathrm{P}(x \dot{\wedge} y) \leftrightarrow [\mathrm{P}(x) \wedge \mathrm{P}(y)] \vee [\mathrm{P}(x) \wedge \mathrm{T}(\dot{\neg} x)] \vee [\mathrm{P}(y) \wedge \mathrm{T}(\dot{\neg} y)]$,
4. $\mathrm{P}(x \dot{\to} y) \leftrightarrow [\mathrm{P}(x) \wedge \mathrm{P}(y)] \vee [\mathrm{P}(x) \wedge \mathrm{T}(\dot{\neg} x)] \vee [\mathrm{P}(y) \wedge \mathrm{T}(y)]$,
5. $\mathrm{P}(\dot{\forall} f) \leftrightarrow [\forall x.\ \mathrm{P}(fx)] \vee [\exists y.\ \mathrm{P}(fy) \wedge \mathrm{T}(\dot{\neg}(fy))]$.

Remark 141. This inductive characterisation of propositions is nearly the same as that of *determinateness* in [24], where Halbach and Fujimoto advocate the definition by appealing to the function of truth as a generalising device [24, pp. 222–223]. We also remark that in KF, it might also be natural to treat each proposition and its negation independently. For instance, in the Strong Kleene evaluation, $\neg(A \wedge B)$ is true when either $\neg A$ or $\neg B$ is true. So, whether $\neg(A \wedge B)$ is a proposition can be determined by $\neg A$ and $\neg B$, rather than by $A \wedge B$. Formally speaking, KF derives:

$$\mathrm{P}(\dot{\neg}(x \dot{\wedge} y)) \leftrightarrow [\mathrm{P}(\dot{\neg} x) \wedge \mathrm{P}(\dot{\neg} y)] \vee [\mathrm{P}(\dot{\neg} x) \wedge \mathrm{T}(\dot{\neg} x)] \vee [\mathrm{P}(\dot{\neg} y) \wedge \mathrm{T}(\dot{\neg} y)].$$

The same remark applies to the other negated propositions $\neg\neg A$, $\neg(A \to B)$, and $\neg\forall x.\ A$.

Finally, the axiom (Lim) generates a new class $\mathsf{l}f$ from a given class f. In particular, we have:

$$\mathsf{KFU} \vdash \mathrm{C}(f) \to \forall x.\ \mathrm{P}(\mathsf{l}fx).$$

In summary, we can express the above construction of propositions in KFU by a single operator $\mathscr{A}$. For each $\mathcal{L}_{FS}$-formula B and a free variable x, the formula $\mathscr{A}(B(\bullet), x)$ [2] is the disjunction of the following:

1. $\exists y, z.\ x = (y \dot{=} z) \wedge y = z$
2. $\exists y.\ x = (\dot{\mathsf{N}}y) \wedge \mathrm{N}(y)$
3. $\exists y.\ x = \dot{\neg} y \wedge B(y)$
4. $\exists y, z.\ x = y \dot{\wedge} z \wedge \{[B(y) \wedge B(z)] \vee [B(y) \wedge \mathrm{T}(\dot{\neg} y)] \vee [B(z) \wedge \mathrm{T}(\dot{\neg} z)]\}$
5. $\exists y, z.\ x = y \dot{\to} z \wedge \{[B(y) \wedge B(z)] \vee [B(y) \wedge \mathrm{T}(\dot{\neg} y)] \vee [B(z) \wedge \mathrm{T}(z)]\}$
6. $\exists g.\ x = \dot{\forall} g \wedge \{[\forall y.\ B(gy)] \vee \exists y.\ B(gy) \wedge \mathrm{T}(\dot{\neg}(gy))\}$
7. $\exists f, y.\ x = \mathsf{l}fy \wedge \forall z.\ B(fz)$

Then, the proposition induction schema (PI) is given by:

$$[\forall x.\ \mathscr{A}(B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{P}(x) \to B(x), \tag{PI}$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

The schema (PI) assures that all propositions are obtained by the inductive construction as explained above.

For the proof-theoretic purposes, we also need the following axioms.

Definition 142. The schema (UG) consists of axioms asserting that each constant symbol in $\{\dot{=}, \dot{\neg}, \dot{\wedge}, \dot{\to}, \dot{\forall}, \dot{\mathsf{N}}, \mathsf{l}\}$ can be uniquely decomposed. For example, (UG) has the following:

$$\dot{\wedge} \neq \dot{\neg},$$

$$\dot{\wedge} a = \dot{\wedge} b \to a = b.$$

With the help of (UG), we can verify in KFU that $\mathrm{P}(x)$ is an $\mathscr{A}$-closure:

$$\mathsf{KFU} + (\mathsf{UG}) \vdash \forall x.\ \mathscr{A}(\mathrm{P}(\bullet), x) \to \mathrm{P}(x).$$

Definition 143. The $\mathcal{L}_{FS}$-theory KFUPI is KFU with the schemata (UG) and (PI).

[2]Here, $\bullet$ means every occurrence of some fixed free variable (*placeholder*).

5.3.2 Lower Bound of KFUPI

The remainder of this section is devoted to the proof-theoretic analysis of KFUPI. In this subsection, we give a relative interpretation of T_0 into KFUPI by extending the translation $'$ used in Fact 136. For each symbol except i, we can employ the same interpretation as given in the proof of Fact 136. In particular, recall that the predicate $\mathrm{R}(f)$ is interpreted as $\mathrm{C}(f)$, namely the statement that f is a *class*. Moreover, $x \in y$ is interpreted as the formula $\mathrm{T}(yx)$. For readability, we sometimes refer to $\mathrm{T}(yx)$ as $x \dot{\in} y$. Also, $\forall x \dot{\in} y.\ A(x)$ means $\forall x.\ x \dot{\in} y \to A(x)$. Finally, we want to interpret the inductive generation i as an appropriate term acc. In other words, we need to find a term acc such that KFUPI derives the translation of the axioms (IG.1) and (IG.2) in Definition 133:

$(\mathsf{IG}.1)'$ $\mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathrm{C}(\mathsf{acc}(a,b)) \wedge \mathsf{Closed}'(a,b,\mathrm{T}(\mathsf{acc}(a,b)(\bullet)))$;

$(\mathsf{IG}.2)'$ $\mathrm{C}(a) \wedge \mathrm{C}(b) \wedge \mathsf{Closed}'(a,b,A(\bullet)) \to \forall x \dot{\in} \mathsf{acc}(a,b).\ A(x)$,

where A is any $\mathcal{L}_{FS}$-formula. We used the notations:

- $y \dot{<}_b x :\equiv (y,x) \dot{\in} b :\equiv \mathrm{T}(b(y,x))$;
- $\mathsf{Closed}'(a,b,A(\bullet)) :\equiv \forall x \dot{\in} a.\ [\forall y \dot{\in} a.\ y \dot{<}_b x \to A(y)] \to A(x)$.

We essentially follow the proof of the lower bound of LUN ([38, pp. 153-156]), to derive the above formulas. Roughly speaking, the interpretations of (IG.1) and (IG.2) are derivable by (L1) and (L2), respectively. Let $\oplus$ be a term such that

$$\oplus(u,v) = \lambda z.[((z)_0 \dot{=} 0 \dot{\wedge} u((z)_1)) \ \dot{\vee} \ (\dot{\neg}((z)_0 \dot{=} 0) \dot{\wedge} v((z)_1))],$$

which represents the disjoint sum of u and v. By recursion, we can construct a term t such that

$$t(u,v,w) = \dot{\forall} y(\oplus(u,v)(0,y) \dot{\to} [\oplus(u,v)(1,(y,w)) \dot{\to} t(u,v,y)]).$$

The term $t(u,v,w)$ informally says that in u, every v-predecessor y of w satisfies $t(u,v,y)$. Of course, $t(u,v,w)$ is not, in general, a proposition even if u and v are classes; hence, we cannot simply employ t as the interpretation of i. Nevertheless, by attaching the limit constant l to t, we can treat t as a proposition. Thus, for the interpretation of i, we further define a term acc such that

$$\mathsf{acc}(u,v)x = \big(\oplus(u,v)(0,x)\big) \dot{\wedge} \big(\dot{\forall} y(\oplus(u,v)(0,y) \dot{\to} [\oplus(u,v)(1,(y,x)) \dot{\to} (\mathsf{l}(\oplus(u,v))(t(u,v,y))])\big).$$

By the following lemma, the term $\mathsf{acc}(a,b)$ is indeed a class whenever a and b are classes, so the first conjunct of $(\mathsf{IG}.1)'$ is derived.

Lemma 144. $\mathsf{KFU} \vdash \mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathrm{C}(\oplus(a,b)) \wedge \mathrm{C}(\mathsf{acc}(a,b))$.

Proof. Assume $\mathrm{C}(a) \wedge \mathrm{C}(b)$. First, taking any z, we show that $\oplus(a,b)z$ is a proposition in order to show that $\oplus(a,b)$ is a class. By the definition of $\oplus$, $\oplus(a,b)z$ is of the form:

$$((z)_0\dot{=}0\dot{\wedge}a((z)_1))\ \dot{\vee}\ (\dot{\neg}((z)_0\dot{=}0)\dot{\wedge}b((z)_1)).$$

Here, $(z)_0\dot{=}0$ and $\dot{\neg}((z)_0\dot{=}0)$ are propositions by Lemma 140, and $a((z)_1)$ and $b((z)_1)$ are also propositions by the assumption $\mathrm{C}(a) \wedge \mathrm{C}(b)$. Therefore, $\oplus(a,b)z$ is a proposition, again by Lemma 140. Since z is arbitrary, $\oplus(a,b)$ is a class, and thus so is $\mathsf{l}(\oplus(a,b))$ by the axiom (Lim). Second, we deduce that $\mathsf{acc}(a,b)x$ is a proposition for any x. Recall that $\mathsf{acc}(a,b)x$ is of the form

$$\oplus(a,b)(0,x)\dot{\wedge}\dot{\forall}y(\oplus(a,b)(0,y)\dot{\to}[\oplus(a,b)(1,(y,x))\dot{\to}(\mathsf{l}(\oplus(a,b))(t(a,b,y)))]).$$

Since we have shown that $\oplus(a,b)$ and $\mathsf{l}(\oplus(a,b))$ are classes, Lemma 140 implies that $\mathsf{acc}(a,b)x$ is a propositon, as required. ■

The following is the second half of $(\mathsf{IG}.1)'$.

Lemma 145. Recall: $\mathsf{Closed}'(a,b,B(\bullet)) :\equiv \forall x\dot{\in}a.\ [\forall y\dot{\in}a.\ y\dot{<}_b x \to A(y)] \to A(x)$. Then, $\mathsf{KFU} \vdash \mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathsf{Closed}'(a,b,\mathrm{T}(\mathsf{acc}(a,b)(\bullet)))$.

Informally, this lemma says that if every b-predecessor of x is contained in the b-accessible part of a, then x is also contained in the b-accessible part of a.
Proof. Assume $\mathrm{C}(a)$ and $\mathrm{C}(b)$. Taking any c, we further assume that $\mathrm{T}(ac)$ and $\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,c)) \to \mathrm{T}(\mathsf{acc}(a,b)y)$, then we have to derive $\mathrm{T}(\mathsf{acc}(a,b)c)$. The first conjunct $\mathrm{T}(\oplus(a,b)(0,c))$ of $\mathrm{T}(\mathsf{acc}(a,b)c)$ is equivalent to $\mathrm{T}(ac)$ in KF and is nothing but one of the assumptions. To derive the second conjunct of $\mathrm{T}(\mathsf{acc}(a,b)c)$, we take any d, and we will derive:

$$\mathrm{T}((\oplus(a,b)(0,d))\dot{\to}(\oplus(a,b)(1,(d,c))\dot{\to}(l(\oplus(a,b))t(a,b,d)))). \tag{5.1}$$

Then, the axioms of KF yield the desired conclusion.

Since the term $\oplus(a,b)$ is a class by Lemma 144 and is reflected by $l(\oplus(a,b))$, the axioms of KF imply that the formula (5.1) is equivalent to:

$$\mathrm{T}(\oplus(a,b)(0,d)) \to \mathrm{T}(\oplus(a,b)(1,(d,c))) \to \mathrm{T}(l(\oplus(a,b))t(a,b,d)).$$

Hence, we suppose that $\mathrm{T}(\oplus(a,b)(0,d))$ and $\mathrm{T}(\oplus(a,b)(1,(d,c)))$ in order to derive $\mathrm{T}(l(\oplus(a,b))t(a,b,d))$. These suppositions are equivalent to $\mathrm{T}(ad)$ and $\mathrm{T}(b(d,c))$, respectively; thus, letting $y := d$, the initial assumption $\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,c)) \to \mathrm{T}(\mathsf{acc}(a,b)y)$ implies $\mathrm{T}(\mathsf{acc}(a,b)d)$, and hence

$$\mathrm{T}(\dot{\forall}y(\oplus(a,b)(0,y)\dot{\to}(\oplus(a,b)(1,(y,d))\dot{\to}(l(\oplus(a,b))t(a,b,y)))))$$

by the definition of acc. By the axioms ($\mathsf{K}_\forall$) and ($\mathsf{K}_\to$), we have

$$\forall y.\ \mathrm{T}(\oplus(a,b)(0,y)) \to \mathrm{T}(\oplus(a,b)(1,(y,d))) \to \mathrm{T}(l(\oplus(a,b))t(a,b,y)).$$

As $\oplus(a,b)$ is a class, the axioms ($\mathsf{U\text{-}True}$) and (Lim) imply that

$$\forall y.\ \mathrm{T}(l(\oplus(a,b))(\dot\neg(\oplus(a,b)(0,y)))) \vee \mathrm{T}(l(\oplus(a,b))(\dot\neg(\oplus(a,b)(1,(y,d)))))$$
$$\vee \mathrm{T}(l(\oplus(a,b))t(a,b,y)).$$

Therefore, by using the axioms ($\mathsf{U}_\to$) and ($\mathsf{U}_\forall$), we obtain

$$\mathrm{T}(l(\oplus(a,b))(\dot\forall y(\oplus(a,b)(0,y)\dot\to(\oplus(a,b)(1,(y,d))\dot\to t(a,b,y))))),$$

which is nothing but the desired formula $\mathrm{T}(l(\oplus(a,b))t(a,b,d))$. Therefore, the formula (5.1) is derived. ■

Finally, we want to derive $(\mathsf{IG}.2)'$:

$$\mathrm{C}(a) \wedge \mathrm{C}(b) \wedge \mathsf{Closed}'(a,b,A(\bullet)) \to \forall x.\ \mathrm{T}(\mathsf{acc}(a,b)x) \to A(x).$$

The author shall roughly explain the basic idea of the proof, which is essentially based on a standard technique found in, e.g., [10, 15, 38, 43]. Taking any x, we want to deduce $A(x)$ from the assumptions $\mathrm{C}(a), \mathrm{C}(b), \mathsf{Closed}'(a,b,A(\bullet))$ and $\mathrm{T}(\mathsf{acc}(a,b)x)$. By the definition of t and acc, the last assumption implies that $t(a,b,x)$ is true, and hence is a proposition. In addition, by Lemma 147 below, whether $A(x)$ holds or not can be reduced to whether $t(a,b,x)$ is a proposition. Therefore, $A(x)$ follows, as required.

We require the following lemma for the proof of Lemma 147.

Lemma 146. KFU derives the following:

$$\mathrm{C}(u) \wedge \mathrm{C}(v) \wedge w\dot\in\mathsf{acc}(u,v) \to \forall x\dot\in u.\ x\dot<_v w \to x\dot\in\mathsf{acc}(u,v).$$

Informally, it says that if w is contained in the v-accessible part of u, then its every v-predecessor x in u also belongs in the v-accessible part of u.

Proof. Assume $\mathrm{C}(u), \mathrm{C}(v)$, and $w\dot\in\mathsf{acc}(u,v)$. In addition, we take any x such that $x\dot\in u$ and $x\dot<_v w$. Then, our purpose is to derive $x\dot\in\mathsf{acc}(u,v)$. Recall that $w\dot\in\mathsf{acc}(u,v)$ is the following formula:

$$\mathrm{T}(\oplus(u,v)(0,w))\dot\wedge\dot\forall y(\oplus(u,v)(0,y)\dot\to[\oplus(u,v)(1,(y,w))\dot\to(l(\oplus(u,v))(t(u,v,y)))]).$$

Thus, in a similar way as the proof of Lemma 145, we obtain:

$$\forall y.\ \mathrm{T}(\oplus(u,v)(0,y)) \to \mathrm{T}(\oplus(u,v)(1,(y,w))) \to \mathrm{T}(l(\oplus(u,v))(t(u,v,y)))).$$

For $y := x$, it follows from the assumptions that $\mathrm{T}(l(\oplus(u,v))(t(u,v,x)))$. Therefore, from the definition of $t(u,v,x)$, we similarly have:

$$\forall y.\ \mathrm{T}(\oplus(u,v)(0,y)) \to \mathrm{T}(\oplus(u,v)(1,(y,x))) \to \mathrm{T}(l(\oplus(u,v))(t(u,v,y))).$$

Since $\oplus(u,v)$ is a class, we obtain:

$$\mathrm{T}(\dot{\forall} y(\oplus(u,v)(0,y)\dot{\to}[\oplus(u,v)(1,(y,x))\dot{\to}(l(\oplus(u,v))(t(u,v,y)))])).$$

Moreover, by the assumption, we also have $\mathrm{T}(\oplus(u,v)(0,x))$; thus $(\mathsf{K}_\wedge)$ yields $x\dot{\in}\mathsf{acc}(u,v)$, as required. ■

Lemma 147. For an $\mathcal{L}_{FS}$-formula $A(x)$, define $B_A(u,v,w)$ to be the formula $\forall x.\ \mathrm{T}(\mathsf{acc}(u,v)x) \to F_A(u,v,w,x)$, where $F_A(u,v,w,x)$ is the conjunction of the following formulas:

1. $w = t(u,v,x) \ \to\ A(x)$,
2. $\forall z.\ w = \{(\oplus(u,v)(1,(z,x)))\dot{\to}(t(u,v,z))\} \to \mathrm{T}(\oplus(u,v)(1,(z,x))) \to A(z)$,
3. $\forall z.\ w = \{(\oplus(u,v)(0,z))\dot{\to}(\oplus(u,v)(1,(z,x))\dot{\to}t(u,v,z))\} \to \mathrm{T}(\oplus(u,v)(0,z)) \to \mathrm{T}(\oplus(u,v)(1,(z,x))) \to A(z)$.

Then, KFUPI derives the following:

$$\mathrm{C}(u) \wedge \mathrm{C}(v) \wedge \mathsf{Closed}'(u,v,A(\bullet)) \to \forall w.\ \mathscr{A}(B_A(u,v,\bullet),w) \to B_A(u,v,w).$$

Proof. Assume $\mathrm{C}(u), \mathrm{C}(v), \mathsf{Closed}'(u,v,A(\bullet))$ and $\mathscr{A}(B_A(u,v,\bullet),w)$. Then, we need to prove $B_A(u,v,w)$. So, taking any x such that $\mathrm{T}(\mathsf{acc}(u,v)x)$, we show $F_A(u,v,w,x)$. The proof is mainly divided into three cases according to the form of w.

1. Assume $w = t(u,v,x) := \dot{\forall} z(\oplus(u,v)(0,z)\dot{\to}(\oplus(u,v)(1,(z,x))\dot{\to}t(u,v,z)))$. Then, we have to show $A(x)$. Now, w is of the universal form $\dot{\forall} f$, thus by the definition of $\mathscr{A}(B_A(u,v,\bullet),w)$ and (UG), we have:

$$[\forall z.\ B_A(u,v,fz)] \vee \exists z.\ B_A(u,v,fz) \wedge \mathrm{T}(\dot{\neg}(fz)).$$

However, in a similar manner to the proof of Lemma 145, the assumption $\mathrm{T}(\mathsf{acc}(u,v)x)$ implies, in KFU, $\forall z.\ \mathrm{T}(fz)$; that is,

$$\forall z.\ \mathrm{T}(\oplus(u,v)(0,z)\dot{\to}(\oplus(u,v)(1,(z,x))\dot{\to}t(u,v,z))).$$

Thus, $\exists z.\ \mathrm{T}(\dot{\neg}(fz))$ is false by $(\mathsf{T\text{-}Cons})$. Therefore, it follows that $\forall z.\ B_A(u,v,fz)$. Then, the assumption $\mathsf{acc}(u,v)x$ yields $\forall z.\ F_A(u,v,fz,x)$. Since fz is of the conditional form, the third conjunct of $F_A(u, v, fz, x)$ is applied, and thus, for all z, we have:

$$\mathrm{T}(\oplus(u, v)(0, z)) \to \mathrm{T}(\oplus(u, v)(1, (z, x))) \to A(z),$$

which is equivalent to $\forall z \dot{\in} u.\ z \dot{<}_v x \to A(z)$. Because $\mathrm{T}(\mathsf{acc}(u, v)x)$ implies $x \dot{\in} u$, the assumption $\mathsf{Closed}'(u, v, A(\bullet))$ implies the formula $A(x)$, as required.

2. Taking any z, we assume $w = (\oplus(u, v)(1, (z, x))) \dot{\to} t(u, v, z)$. Then, we will deduce $\mathrm{T}(\oplus(u, v)(1, (z, x))) \to A(z)$. Thus, similar to the case of $w = t(u, v, x)$, the assumption $\mathscr{A}(B_A(u, v, \bullet), w)$ implies the following:

$$\mathrm{T}(\oplus(u, v)(1, (z, x))) \to B_A(u, v, t(u, v, z)). \tag{5.2}$$

By the assumption $\mathrm{T}(\mathsf{acc}(u, v)x)$ and Lemma 146, we have $\mathrm{T}(\mathsf{acc}(u, v)z)$. Therefore, (5.2) implies:

$$\mathrm{T}(\oplus(u, v)(1, (z, x))) \to F_A(u, v, t(u, v, z), z).$$

Thus, the first conjunct of $F_A(u, v, t(u, v, z), z)$ is applied, and we get

$$\mathrm{T}(\oplus(u, v)(1, (z, x))) \to A(z).$$

3. The case $w = \oplus(u, v)(0, z) \dot{\to} (\oplus(u, v)(1, (z, x)) \dot{\to} t(u, v, z))$ is similarly proved as the second case.

4. If w is of the other form, we trivially have $B_A(u, v, w)$.

■

Using the above lemma, we can derive $(\mathsf{IG}.2)'$.

Lemma 148. KFUPI derives the following:

$$\mathrm{C}(a) \wedge \mathrm{C}(b) \wedge \mathit{Closed}(a, b, A) \to \forall x.\ \mathrm{T}(\mathsf{acc}(a, b)x) \to A(x).$$

Proof. We assume $\mathrm{C}(a), \mathrm{C}(b), \mathit{Closed}(a, b, A)$ and $\mathrm{T}(\mathsf{acc}(a, b)c)$ for an arbitrary c. Then, we have to derive $A(c)$. From the assumptions and Lemma 147, we obtain $\forall w.\ \mathscr{A}(B_A(u, v, \bullet), w) \to B_A(u, v, w)$. Thus, (PI) yields:

$$\forall x.\ \mathrm{P}(x) \to B_A(a, b, x).$$

Letting $x := t(a, b, c)$, we want to show $\mathrm{P}(t(a, b, c))$, from which $B_A(a, b, t(a, b, c))$ follows. Then, we can immediately obtain the conclusion $A(c)$ by the assumption $\mathrm{T}(\mathsf{acc}(a, b)c)$ and the definition of B_A.

In order to prove $\mathrm{P}(t(a,b,c))$, it suffices to derive the following:

$$\forall y.\ \mathrm{T}(\oplus(a,b)(0,y)) \to \mathrm{T}(\oplus(a,b)(1,(y,c))) \to \mathrm{T}(t(a,b,y)). \tag{5.3}$$

In fact, this formula implies, in KFU, $\mathrm{T}(t(a,b,c))$ and thus $\mathrm{P}(t(a,b,c))$ because $\oplus(a,b)$ is a class. However, the formula (5.3) follows from the assumption $\mathrm{T}(\mathsf{acc}(a,b)c)$. ■

To summarise, we obtain the interpretation of T_0 into KFUPI, in which $\mathcal{L}$-vocabularies are preserved.

Theorem 149. For each $\mathcal{L}_{EM}$-sentence A, if $\mathsf{T}_0 \vdash A$, then $\mathsf{KFUPI} \vdash A'$.

5.3.3 Kahle's Model for KFU

In this subsection, we introduce Kahle's model construction for KFU [41, pp. 214-215] and observe that this model also satisfies KFUPI. The upper-bound proof of KFUPI in Section 5.3.4 is essentially based on this model.

First, the base theory TON is interpreted by the closed total term model $\mathcal{CTT}$ (e.g. [14, p. 26]); the domain of $\mathcal{CTT}$ is the set of all closed $\mathcal{L}$-terms; the constant symbols are interpreted as themselves; the application function $\mathsf{App}(x,y)$ is defined to be the juxtaposition of x and y; and it is defined that the equation $x = y$ holds when x and y are β-equivalent in the standard sense. Then, $\mathrm{N}(x)$ holds when x reduces to a numeral. To interpret the predicates T and U, we define an operator $\Phi(X,Y,a,\alpha)$ for an ordinal number α :

1. $a \in Y$,
2. $\exists b,c.\ \alpha = 0 \land a = (b \dot{=} c) \land b = c$,
3. $\exists b,c.\ \alpha = 0 \land a = \dot{\neg}(b \dot{=} c) \land b \neq c$,
4. $\exists b.\ \alpha = 0 \land a = \dot{\mathsf{N}} b \land \mathrm{N}(b)$,
5. $\exists b.\ \alpha = 0 \land a = \dot{\neg}(\dot{\mathsf{N}} b) \land \neg\mathrm{N}(b)$,
6. $\exists b.\ a = \dot{\neg}(\dot{\neg} b) \land a \notin Y \land b \in X$,
7. $\exists b,c.\ a = (b \dot{\land} c) \land a \notin Y \land b \in X \land c \in X$,
8. $\exists b,c.\ a = \dot{\neg}(b \dot{\land} c) \land a \notin Y \land [\dot{\neg} b \in X \lor \dot{\neg} c \in X]$,
9. $\exists b,c.\ a = (b \dot{\to} c) \land a \notin Y \land [\dot{\neg} b \in X \lor c \in X]$,
10. $\exists b,c.\ a = \dot{\neg}(b \dot{\to} c) \land a \notin Y \land b \in X \land \dot{\neg} c \in X$,
11. $\exists f.\ a = (\dot{\forall} f) \land a \notin Y \land \forall x.\ fx \in X$,
12. $\exists f.\ a = \dot{\neg}(\dot{\forall} f) \land a \notin Y \land \exists x.\ \dot{\neg}(fx) \in X$,

13. $\exists b, c.\ \alpha \in \mathrm{Suc} \wedge a = \mathsf{I}bc \wedge a \notin Y \wedge \dot{\neg} a \notin Y \wedge [\forall x.\ bx \in Y \vee \dot{\neg}(bx) \in Y] \wedge c \in Y$,
14. $\exists b, c.\ \alpha \in \mathrm{Suc} \wedge a = \dot{\neg}(\mathsf{I}bc) \wedge a \notin Y \wedge \dot{\neg} a \notin Y \wedge [\forall x.\ bx \in Y \vee \dot{\neg}(bx) \in Y] \wedge c \notin Y$,

where $\alpha \in \mathrm{Suc}$ means that α is a successor ordinal.

A Φ-sequence (Z_α) of least fixed points Z_α is defined as follows:

1. $Z_0 := \emptyset$.
2. $Z_{\alpha+1} :=$ the least fixed point of $\Phi(X, Z_\alpha, x, \alpha + 1)$.
3. $Z_\alpha :=$ the least fixed point of $\Phi(X, \bigcup_{\beta<\alpha} Z_\beta, x, \alpha)$, for a limit ordinal α.

Since this sequence (Z_α) is monotonically increasing, we eventually reach the least ordinal ι such that $Z_\iota = Z_{\iota+1}$.[3]

Then, we interpret $\mathrm{T}(t)$ as $t \in Z_\iota$, and $\mathrm{U}(u)$ as $\exists f.\ u = \mathsf{I}f \wedge \forall x.\ fx \in Z_\iota \vee \dot{\neg}(fx) \in Z_\iota$. Combining this with the above interpretation, we obtain an $\mathcal{L}_{FS}$-model $\mathcal{M}_{\mathsf{KF}}$.

Proposition 150. $\mathcal{M}_{\mathsf{KF}} \models \mathsf{KFUPI}$.

Proof. Kahle [41, p. 215] already showed $\mathcal{M}_{\mathsf{KF}} \models \mathsf{KFU}+(\mathsf{UG})$; thus, we concentrate on the schema (PI):

$$[\forall x.\ \mathscr{A}(B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{P}(x) \to B(x).$$

We take any $\mathcal{L}_{FS}$-formula B such that $\mathcal{M}_{\mathsf{KF}} \models \forall x.\ \mathscr{A}(B(\bullet), x) \to B(x)$. Then, by induction on $\alpha \leq \iota$, we prove that if $x \in Z_\alpha$ or $\dot{\neg} x \in Z_\alpha$, then $\mathcal{M}_{\mathsf{KF}} \models B(x)$.

For example, we consider the case $x = \mathsf{I}fy$ with $\dot{\neg}(\mathsf{I}fy) \in Z_\alpha$. Then, we must show $\mathcal{M}_{\mathsf{KF}} \models B(\mathsf{I}fy)$. By the definition of the operator Φ, the crucial case is when α is the least successor ordinal such that $\forall z.\ fz \in \bigcup_{\beta<\alpha} Z_\beta \vee \dot{\neg}(fz) \in \bigcup_{\beta<\alpha} Z_\beta$. Thus, by the induction hypothesis, we obtain $\mathcal{M}_{\mathsf{KF}} \models \forall x.\ B(fz)$. Since B is $\mathscr{A}$-closed, it follows that $\mathcal{M}_{\mathsf{KF}} \models B(\mathsf{I}fy)$. The other cases are similarly proved by using subinduction on the construction of Z_α. ■

Remark 151. As Kahle remarks [41, p. 215], the model $\mathcal{M}_{\mathsf{KF}}$ also satisfies the following principles:

($\mathsf{U\text{-}Tran}$) $\mathrm{U}(u) \wedge \mathrm{U}(v) \to [\mathrm{T}(v(ux)) \to \mathrm{T}(vx)]$.

($\mathsf{U\text{-}Dir}$) $\mathrm{U}(u) \wedge \mathrm{U}(v) \to \exists w.\ \mathrm{U}(w) \wedge u \sqsubset w \wedge v \sqsubset w$.

[3] In fact, this ordinal ι is identical to the first recursively inaccessible ordinal, for the definition of which, see, e.g. [5].

(U-Nor) $\mathrm{U}(u) \to \exists f.\ \mathrm{C}(f) \land u = \mathsf{I}f$.

(U-Lin) $\mathrm{U}(u) \land \mathrm{U}(v) \to u \sqsubset v \lor v \sqsubset u \lor \forall x. x \dot{\in} u \leftrightarrow x \dot{\in} v$.

Therefore, Proposition 150 also verifies the consistency of KFUPI + (U-Tran) + (U-Dir) + (U-Nor) + (U-Lin).

5.3.4 Upper Bound of KFUPI

In order to obtain the upper bound of KFUPI, we interpret the truth predicate of KFUPI as the least fixed point of a non-monotone operator which is almost the same as Φ in Section 5.3.3. In the operator Φ, the truth condition of $\dot{\neg}(\mathsf{I}bc)$ for a class b was characterised as the *non*-truth of c, where non-monotonicity concerns. That is why monotone inductive operators are not enough to formalise $\mathcal{M}_{\mathsf{KF}}$. Thus, we use the theory FID([POS, QF]) of non-monotone inductive definition in [36].

Let $\mathcal{L}'$ be an language of Peano arithmetic (PA) that contains symbols for all primitive recursive functions. Let $\mathcal{L}'(X)$ be the extension of $\mathcal{L}'$ with a new unary predicate symbol $x \in X$. An *operator form* $\mathfrak{B}(X, u)$ is an $\mathcal{L}'(X)$-formula in which at most one variable u occurs freely. We write $\mathfrak{B}(A, u)$ as the result of replacing each occurrence of $t \in X$ by a formula $A(t)$. An operator form $\mathfrak{B}(X, u)$ is in POS if X occurs only positively in $\mathfrak{B}$; an operator form $\mathfrak{B}(X, u)$ is in QF if it contains no quantifier. We consider the particular operators $\mathfrak{A}_0 \in$ POS and $\mathfrak{A}_1 \in$ QF, which are defined below. Then, the operator $\mathfrak{A}$ is defined to be the following formula:

$$\mathfrak{A}(X, u) := \mathfrak{A}_0(X, u) \lor ([\forall x.\ \mathfrak{A}_0(X, x) \to x \in X] \land \mathfrak{A}_1(X, u)).$$

Let the two-sorted language $\mathcal{L}_{\mathcal{K}}$ be the extension of $\mathcal{L}'$ with ordinal variables $\alpha, \beta, \ldots$, a binary relation symbol $<$ on the ordinals, and the binary relation symbol $\mathrm{P}_{\mathfrak{A}}$ for the above operator $\mathfrak{A}$.

We shall use the following notations:

- $\mathrm{P}^{\alpha}_{\mathfrak{A}}(s) := \mathrm{P}_{\mathfrak{A}}(\alpha, s)$,
- $\mathrm{P}^{<\alpha}_{\mathfrak{A}}(s) := \exists \xi < \alpha.\ \mathrm{P}^{\xi}_{\mathfrak{A}}(s)$,
- $\mathrm{P}_{\mathfrak{A}}(s) := \exists \alpha.\ \mathrm{P}^{\alpha}_{\mathfrak{A}}(s)$.

Definition 152. The $\mathcal{L}_{\mathcal{K}}$-theory FID([POS, QF]) consists of PA with the full induction schema for $\mathcal{L}_{\mathcal{K}}$ and the following axioms:[4]

1. $\alpha \not< \alpha \land (\alpha < \beta \land \beta < \gamma \to \alpha < \gamma) \land (\alpha < \beta \lor \alpha = \beta \lor \alpha > \beta)$.
2. (OP.1) $\mathrm{P}^{\alpha}_{\mathfrak{A}}(s) \leftrightarrow \mathrm{P}^{<\alpha}_{\mathfrak{A}}(s) \lor \mathfrak{A}(\mathrm{P}^{<\alpha}_{\mathfrak{A}}, s)$,

[4] In contrast to Jäger and Studer's formulation of FID([POS, QF]) in [36], we now consider only one operator form $\mathfrak{A}$ for simplicity.

(OP.2) $\mathfrak{A}(\mathrm{P}_{\mathfrak{A}}, s) \to \mathrm{P}_{\mathfrak{A}}(s)$.

3. $[\forall \xi.\ \{\forall \eta < \xi.\ A(\eta)\} \to A(\xi)] \to \forall \xi.\ A(\xi)$, for all $\mathcal{L}_{\mathcal{K}}$-formulae $A(\alpha)$.

Based on [41, Section 5.2.2], we now formalise the closed total term model of KFUPI within $\mathsf{FID}([\mathsf{POS}, \mathsf{QF}])$. First, we define a translation $^\star : \mathcal{L} \to \mathcal{L}'$. Fixing some Gödel-numbering, each closed term t is assigned the corresponding natural number $\ulcorner t \urcorner$; thus, the domain of $\mathcal{CTT}$ can be seen as a subset of $\mathbb{N}$, which is expressed by some arithmetical formula. Thus, the quantifier symbols are relativised to this formula. The translation $\mathsf{App}^\star$ of the application function is defined to be the standard Kleene bracket $\{x\}(y)$, which returns the result of the partial recursive function $\{x\}$ at y if it has a value. Each constant symbol c is interpreted as a code e such that $\{e\}(\ulcorner t \urcorner) \simeq \ulcorner \mathsf{c}t \urcorner$ for each closed term t. Similarly, $\mathrm{N}^\star(x)$ is an arithmetical formula expressing that x is the code of a numeral, and an arithmetical formula $x =^\star y$ is true when x and y have the same reduct. Under this interpretation, every theorem of TON is derivable in PA (cf. [14, Theorem 4.13]). Next, to expand * to the language $\mathcal{L}_{FS}$, we need to define the extension of the truth predicate T.

We define the operator form $\mathfrak{A}_0(X, a) \in \mathsf{POS}$ to be the disjunction of the following:

1. $\exists b, c.\ a = (b \dot{=} c) \land b = c$
2. $\exists b, c.\ a = \dot{\neg}(b \dot{=} c) \land b \neq c$
3. $\exists b.\ a = (\dot{\mathsf{N}} b) \land \mathrm{N}(b)$
4. $\exists b.\ a = \dot{\neg}(\dot{\mathsf{N}} b) \land \neg \mathrm{N}(b)$
5. $\exists b.\ a = \dot{\neg}(\dot{\neg} b) \land b \in X$
6. $\exists b, c.\ a = (b \dot{\land} c) \land b \in X \land c \in X$
7. $\exists b, c.\ a = \neg(b \dot{\land} c) \land [\dot{\neg} b \in X \lor \dot{\neg} c \in X]$
8. $\exists b, c.\ a = (b \dot{\to} c) \land [\dot{\neg} b \in X \lor c \in X]$
9. $\exists b, c.\ a = \dot{\neg}(b \dot{\to} c) \land b \in X \land \dot{\neg} c \in X$
10. $\exists f.\ a = \dot{\forall} f \land \forall b.\ fb \in X$
11. $\exists f.\ a = \dot{\neg}(\dot{\forall} f) \land \exists b.\ \dot{\neg}(fb) \in X$
12. $\exists f.\ a = \mathsf{I} f(0 \dot{=} 0) \land \forall b.\ fb \in X \lor \dot{\neg} fb \in X$

Note that the last clause is given to record that f is a class at the stage. Next, let an operator form $\mathfrak{A}_1(X, a) \in \mathsf{QF}$ be the disjunction of the following:

1. $\exists f, b.\ a = \mathsf{I} fb \land \mathsf{I} f(0 \dot{=} 0) \in X \land \mathsf{I} f(\dot{\mathsf{N}} 0) \notin X \land b \in X$.
2. $\exists f, b.\ a = \dot{\neg}(\mathsf{I} fb) \land \mathsf{I} f(0 \dot{=} 0) \in X \land \mathsf{I} f(\dot{\mathsf{N}} 0) \notin X \land b \notin X$.

It should be clear that $\mathfrak{A}_1$ can be given as a quantifier-free formula. The condition that f is a class is now given by the quantifier-free formula $\mathsf{I}f(0\dot{=}0) \in X$, so $\mathfrak{A}_1(X,a)$ can dispense with quantifier symbols. The condition $\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}) \notin X$ is used to ensure that every term of the form $\mathsf{I}fu$ or $\dot{\neg}(\mathsf{I}fu)$ is simultaneously determined to be true or false at some stage (see Lemma 153). Finally, the operator form $\mathfrak{A}$ is defined, as explained above:

$$\mathfrak{A}(X,a) := \mathfrak{A}_0(X,a) \vee ([\forall x.\ \mathfrak{A}_0(X,x) \rightarrow x \in X] \wedge \mathfrak{A}_1(X,a)).$$

To complete the definition of the translation *, let $\mathrm{T}^\star(t)$ be $\mathrm{P}_{\mathfrak{A}}(t)$, and $\mathrm{U}^\star(u)$ be $\exists f.\ u = \mathsf{I}f \wedge \forall x. \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. Under the interpretation *, we want to derive all the theorems of KFUPI in $\mathsf{FID}([\mathsf{POS},\mathsf{QF}])$. For that purpose, we use the following lemma, which shows that when f is a class, $\mathsf{I}fb$ is determined to be true or false at some stage α, and the truth value never changes at the later stages.

Lemma 153. We work in $\mathsf{FID}([\mathsf{POS},\mathsf{QF}])$.

Assume $\forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. Then, there exists an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0})) \wedge \neg \mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$. Furthermore, for this α, we have the following:

1. $\forall b.\ \mathrm{P}_{\mathfrak{A}}(\mathsf{I}fb) \leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}fb)$,
2. $\forall b.\ \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(\mathsf{I}fb)) \leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\dot{\neg}(\mathsf{I}fb))$.

Proof. We assume that $\forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. Firstly, we prove $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$. For a contradiction, we suppose $\neg\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$. Since $(\mathsf{OP}.2)$ implies that $\mathrm{P}_{\mathfrak{A}}$ is closed under $\mathfrak{A}_0$, we have $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\mathsf{0}\dot{=}\mathsf{0}))$. Again by $(\mathsf{OP}.2)$, it follows that $\forall a.\ \mathfrak{A}_1(\mathrm{P}_{\mathfrak{A}},a) \rightarrow \mathrm{P}_{\mathfrak{A}}(a)$. Therefore, by the supposition $\neg\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$, we obtain $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$, a contradiction. Thus, $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$ is proved. By the transfinite induction schema, we also have an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0})) \wedge \neg\mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$.

Next, we show $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}fb) \rightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}fb)$. Note that the right-to-left direction is obvious. Since $\mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{0}\dot{=}\mathsf{0})$, it suffices to consider the case $b \neq (\mathsf{0}\dot{=}\mathsf{0})$. Now, we assume $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}fb)$, and hence we can take a least β such that $\mathrm{P}^{\beta}_{\mathfrak{A}}(\mathsf{I}fb)$. Then, β is clearly equal to α, and thus we have $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}fb)$, as required. We can similarly verify the second item $\mathrm{P}_{\mathfrak{A}}(\dot{\neg}(\mathsf{I}fb)) \leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\dot{\neg}(\mathsf{I}fb))$. ∎

Lemma 154. For every $\mathcal{L}_{FS}$-formula A, if $\mathsf{KFUPI} \vdash A$, then $\mathsf{FID}([\mathsf{POS},\mathsf{QF}]) \vdash A^\star$.

Proof. If A is an axiom of TON or (UG), we already have $\mathsf{PA} \vdash A^\star$. Thus, it suffices to deal with the axioms displayed in Definition 137 and the schema (PI).

Compositional axioms in U. We consider the axiom $(\mathsf{U}_\forall)$:

$$\mathrm{U}^\star(u) \rightarrow \forall g.\ \mathrm{T}^\star(u(\dot{\forall}g)) \leftrightarrow \forall x.\ \mathrm{T}^\star(u(gx)).$$

Thus, we assume $\mathrm{U}^\star(u)$, so we can take a closed term f such that $u = \mathsf{I}f \wedge \forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. By Lemma 153, there exists an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0})) \wedge \neg \mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$. Moreover, for any closed term g, we have:

$$\begin{aligned}
\mathrm{P}_{\mathfrak{A}}(u(\dot{\forall} g)) &\leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(u(\dot{\forall} g)) && (\because \text{Lemma } 153)\\
&\leftrightarrow \mathrm{P}^{<\alpha}_{\mathfrak{A}}(\dot{\forall} g) && (\because \text{def. of } \mathfrak{A})\\
&\leftrightarrow \forall x \mathrm{P}^{<\alpha}_{\mathfrak{A}}(gx) && (\because \mathrm{P}^{<\alpha} \text{ is } \mathfrak{A}_0\text{-closed})\\
&\leftrightarrow \forall x \mathrm{P}^{\alpha}_{\mathfrak{A}}(u(gx)) && (\because \text{def. of } \mathfrak{A})\\
&\leftrightarrow \forall x \mathrm{P}_{\mathfrak{A}}(u(gx)). && (\because \text{Lemma } 153)
\end{aligned}$$

Therefore, we obtain $\forall g.\ \mathrm{T}^\star(u(\dot{\forall} g)) \leftrightarrow \forall x.\ \mathrm{T}^\star(u(gx))$. The other compositional axioms are similarly treated.

Consistency in U**.** We consider the axiom ($\mathsf{U\text{-}Cons}$):

$$\mathrm{U}^\star(u) \to \forall x.\ \neg[\mathrm{T}^\star(ux) \wedge \mathrm{T}^\star(u(\dot{\neg} x))].$$

As above, from the assumption $\mathrm{U}^\star(u)$, we take a closed term f such that $u = \mathsf{I}f \wedge \forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. Also, we get an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0})) \wedge \neg \mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$. In order to get a contradiction, we further take any closed term x such that $\mathrm{P}_{\mathfrak{A}}(ux)$ and $\mathrm{P}_{\mathfrak{A}}(u(\dot{\neg} x))$. Then, Lemma 153 implies $\mathrm{P}^{\alpha}_{\mathfrak{A}}(ux)$ and $\mathrm{P}^{\alpha}_{\mathfrak{A}}(u(\dot{\neg} x))$, and thus we also have the inconsistency $\mathrm{P}^{<\alpha}_{\mathfrak{A}}(x) \wedge \neg \mathrm{P}^{<\alpha}_{\mathfrak{A}}(x)$ by the definition of $\mathfrak{A}_1$.

Structural properties of U**.** We consider the axiom (Lim):

$$\mathrm{C}^\star(f) \to \mathrm{U}^\star(\mathsf{I}f) \wedge (f \sqsubset \mathsf{I}f)^\star.$$

We assume $\mathrm{C}^\star(f)$, then $\mathrm{U}^\star(\mathsf{I}f)$ is clear from the definition. Thus, we show the translation of $f \sqsubset \mathsf{I}f$:

1. $\forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \to \mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(fx))$,
2. $\forall x.\ \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx)) \to \mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(\dot{\neg}(fx)))$.

As to the first item, we take any x satisfying $\mathrm{P}_{\mathfrak{A}}(fx)$, and show $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(fx))$. By the assumption $\mathrm{C}^\star(f)$, Lemma 153 implies that there exists a least α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathsf{N}}\mathsf{0}))$. Thus, by the definition of $\mathfrak{A}$, it follows that $\forall y.\ \mathrm{P}^{<\alpha}_{\mathfrak{A}}(fy) \vee \mathrm{P}^{<\alpha}_{\mathfrak{A}}(\dot{\neg}(fy))$. By the consistency of $\mathrm{P}_{\mathfrak{A}}$, we obtain $\mathrm{P}^{<\alpha}_{\mathfrak{A}}(fx)$. Therefore, we get $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(fx))$, and thus $\mathrm{P}_{\mathfrak{A}}(\mathsf{I}f(fx))$. The second item is similar.

The other structural axioms are similarly proved.

Propositon induction. The proof is almost the same as that of Proposition 150.

■

Theorem 155. KFUPI and T_0 are proof-theoretically equivalent.

Proof. The lower bound of KFUPI is given by Theorem 149. As for the upper bound, Lemma 154 showed that KFUPI is proof-theoretically reducible to FID([POS, QF]), which is known to be proof-theoretically equivalent to T_0 [36]. ■

5.4 Proposition Induction for Aczel-Feferman: DTUPI

In the present section, we consider another theory, PT, based on Aczel's original Frege structure [1]. Since the notion of propositions in PT is different from that in KF, we need to change the definition of proposition induction slightly. Nevertheless, almost the same proof-theoretic analysis as for KFUPI can be applied to PT, and hence we obtain the theory PTUPI, which is proof-theoretically equivalent to T_0 (Theorem 162).

5.4.1 System PT and Universes

Whereas KF is based on Strong Kleene logic, Aczel's original Frege structure [1] is essentially based on *Aczel–Feferman* logic, a variant of Weak Kleene logic which has the following truth tables, as given in Definition 11:

$\neg$	
T	F
U	U
F	T

$\vee$	T	U	F
T	T	U	T
U	U	U	U
F	T	U	F

$\wedge$	T	U	F
T	T	U	F
U	U	U	U
F	F	U	F

$\rightarrow$	T	U	F
T	T	U	F
U	U	U	U
F	T	T	T

Note that while $A \vee B$ is definable as $\neg(\neg A \wedge \neg B)$, $\rightarrow$ cannot be defined by $\neg$ and $\wedge$.

Similar to Definition 11, the theory PT (proposition and truth) is defined as follows.

Definition 156. (cf. [19]) The $\mathcal{L}_{FS}$-theory PT has TON and (the universal closure of) the following axioms:

Compositional axioms.

- $\mathrm{T}(x \dot{=} y) \leftrightarrow x = y$

- $\mathrm{T}(\dot\neg(x \dot= y)) \leftrightarrow x \neq y$
- $\mathrm{T}(\dot{\mathsf{N}}x) \leftrightarrow \mathrm{N}(x)$
- $\mathrm{T}(\dot\neg(\dot{\mathsf{N}}x)) \leftrightarrow \neg\mathrm{N}(x)$
- $\mathrm{T}(\dot\neg(\dot\neg x)) \leftrightarrow \mathrm{T}(x)$
- $\mathrm{T}(x \dot\wedge y) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(y)$
- $\mathrm{T}(\dot\neg(x \dot\wedge y)) \leftrightarrow [\mathrm{T}(\dot\neg x) \wedge \mathrm{T}(\dot\neg y)] \vee [\mathrm{T}(\dot\neg x) \wedge \mathrm{T}(y)] \vee [\mathrm{T}(x) \wedge \mathrm{T}(\dot\neg y)]$
- $\mathrm{T}(x \dot\to y) \leftrightarrow [\mathrm{T}(x) \wedge \mathrm{T}(y)] \vee \mathrm{T}(\dot\neg x)$
- $\mathrm{T}(\dot\neg(x \dot\to y)) \leftrightarrow [\mathrm{T}(x) \wedge \mathrm{T}(\dot\neg y)]$
- $\mathrm{T}(\dot\forall f) \leftrightarrow \forall x.\ \mathrm{T}(fx)$
- $\mathrm{T}(\dot\neg(\dot\forall f)) \leftrightarrow [\forall x.\ \mathrm{T}(fx) \vee \mathrm{T}(\dot\neg(fx))] \wedge \exists x.\ \mathrm{T}(\dot\neg(fx))$

Consistency.

- $\neg[\mathrm{T}(x) \wedge \mathrm{T}(\dot\neg x)]$

In Aczel's Frege structure, propositions can be naturally characterised inductively. Recall the notation $\mathrm{P}(x) := \mathrm{T}(x) \vee \mathrm{T}(\dot\neg x)$. Then, the following holds.

Lemma 157. (cf. [1]) PT derives the following:

1. $\mathrm{P}(x \dot= y) \wedge \mathrm{P}(\dot{\mathsf{N}}x)$,
2. $\mathrm{P}(\dot\neg x) \leftrightarrow \mathrm{P}(x)$,
3. $\mathrm{P}(x \dot\wedge y) \leftrightarrow \mathrm{P}(x) \wedge \mathrm{P}(y)$,
4. $\mathrm{P}(x \dot\to y) \leftrightarrow \mathrm{P}(x) \wedge (\mathrm{T}(x) \to \mathrm{P}(y))$,
5. $\mathrm{P}(\dot\forall f) \leftrightarrow \forall x.\ \mathrm{P}(fx)$.

The proof-theoretic analysis of PT is found in, e.g., [6, 15, 34, 22]:

Fact 158. PT and EM have the same $\mathcal{L}$-theorems.

Definition 159. The $\mathcal{L}_{FS}$-system PTU consists of TON and the following axioms:

Compositional axioms in U**.**

- $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(x \dot= y) \leftrightarrow x = y$
- $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(x \dot\neq y) \leftrightarrow x \neq y$
- $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(\dot{\mathsf{N}}x) \leftrightarrow \mathrm{N}(x)$
- $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(\dot\neg(\dot{\mathsf{N}}x)) \leftrightarrow \neg\mathrm{N}(x)$
- $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(\dot\neg(\dot\neg x)) \leftrightarrow \mathrm{T}(x)$
- $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(x \dot\wedge y) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(y)$

- $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(\dot\neg(x \dot\wedge y)) \leftrightarrow [\mathrm{T}(\dot\neg x) \wedge \mathrm{T}(\dot\neg y)] \vee [\mathrm{T}(\dot\neg x) \wedge \mathrm{T}(y)] \vee [\mathrm{T}(x) \wedge \mathrm{T}(\dot\neg y)]$
- $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(x \dot\to y) \leftrightarrow [\mathrm{T}(x) \wedge \mathrm{T}(y)] \vee \mathrm{T}(\dot\neg x)$
- $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(\dot\neg(x \dot\to y)) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(\dot\neg y)$
- $\mathrm{U}(u) \to \forall f.\ \mathrm{T}(\dot\forall f) \leftrightarrow \forall x.\ \mathrm{T}(fx)$
- $\mathrm{U}(u) \to \forall f.\ \mathrm{T}(\dot\neg(\dot\forall f)) \leftrightarrow [\forall x.\ \mathrm{T}(fx) \vee \mathrm{T}(\dot\neg(fx))] \wedge \exists x.\ \mathrm{T}(\dot\neg(fx))$

Consistency in U**.**

- $\mathrm{U}(u) \to \forall x.\ \neg[\mathrm{T}(x) \wedge \mathrm{T}(\dot\neg x)]$

Structural properties of U**.**

- $\mathrm{U}(u) \to \mathrm{C}(u)$
- $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(ux) \to \mathrm{T}(x)$
- $\mathrm{C}(f) \to \mathrm{U}(\mathsf{I}f) \wedge f \sqsubset \mathsf{I}f$

As far as the author knows, the system PTU has not been explicitly formulated in the literature, though Fujimoto [23, pp. 933-935] formulated and analysed transfinite iterations of the truth theory DT (Definition 11), which is essentially PT formulated over Peano arithmetic. Fujimoto's proof can be adapted to the proof-theoretic analysis of PTU. As a result, we obtain the following:

Corollary 160. PTU and EMU have the same $\mathcal{L}$-theorems.

5.4.2 PTU with Proposition Induction

In order to define the proposition induction for PT, we now use another operator $\mathscr{A}^{\mathsf{PT}}$, which is based on the inductive characterisation of propositions in Lemma 157. For each $\mathcal{L}_{FS}$-formula B and a free variable x, the formula $\mathscr{A}^{\mathsf{PT}}(B(\bullet), x)$ is the disjunction of the following:

1. $\exists y, z.\ x = (y \dot= z) \wedge y = z$
2. $\exists y.\ x = (\dot{\mathsf{N}} y) \wedge \mathrm{N}(y)$
3. $\exists y.\ x = \dot\neg y \wedge B(y)$
4. $\exists y, z.\ x = y \dot\wedge z \wedge B(y) \wedge B(z)$
5. $\exists y, z.\ x = y \dot\to z \wedge B(y) \wedge [\mathrm{T}(y) \to B(z)]$
6. $\exists g.\ x = \dot\forall g \wedge \forall y.\ B(gy)$
7. $\exists f, y.\ x = \mathsf{I} f y \wedge \forall z.\ B(fz)$

Then, the proposition induction schema ($\mathsf{PI}^{\mathsf{PT}}$) is given by:

$$[\forall x.\ \mathscr{A}^{\mathsf{PT}}(B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{P}(x) \to B(x),$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

Definition 161. The $\mathcal{L}_{FS}$-theory PTUPI is PTU with the schemata (UG) and ($\mathsf{PI}^{\mathsf{PT}}$).

As for the proof-theoretic strength of PTUPI, exactly the same lower-bound proof of Section 5.3.2 can be given in PTUPI; thus, we have $\mathsf{T}_0 \leq \mathsf{PTUPI}$. Moreover, Kahle's model construction in Section 5.3.3 and, thus, the upper-bound proof of KFUPI in Section 5.3.4 are easily modified for PTUPI. As a result, we also obtain the upper bound.

Theorem 162. PTUPI and T_0 are proof-theoretically equivalent.

5.5 Extension by Least Universes

In Sections 5.3 and 5.4, we extended theories of Frege structure by using the induction principle on propositions, in analogy with name induction in [38]. In the same paper, Jäger, Kahle and Studer further proposed the idea of *least universes*, based on which they formulated the theory LUN. As LUN is proof-theoretically equivalent to T_0 [38, Conclusion 25], we can expect to obtain a strong theory of Frege structure by requiring some kind of leastness on universes. In fact, Burgess [10] proposed a truth theory KF_μ with impredicative strength, in which the truth predicate is intended to denote the *least* fixed point of the Kripke operator. Therefore, in this section, we aim to extend KF by least universes.

In Kahle's model $\mathcal{M}_{\mathsf{KF}}$ in Section 5.3.3, each universe of the form $\mathsf{I}f$ reflects a class f and is a fixed point of the Kripke operator. Thus, one could naturally introduce leastness by defining a universe $\mathsf{I}f$ as the least set that reflects f and is closed under the Kripke operator.[5] However, we can easily observe that universes so defined violate natural structural properties in Remark 151, similar to [38, Theorem 14]. This means that such universes are incompatible with $\mathcal{M}_{\mathsf{KF}}$. To make matters worse, these universes are not generally fixed points of the Kripke operator, so they do not sufficiently serve as truth predicates. This definition of leastness does not capture universes in $\mathcal{M}_{\mathsf{KF}}$ because universes in $\mathcal{M}_{\mathsf{KF}}$ may contain other small universes. Therefore, in order to define a least universe $\mathsf{I}f$ compatible with $\mathcal{M}_{\mathsf{KF}}$, we also need to accommodate smaller universes $\mathsf{I}g$ such that $\mathsf{I}g \sqsubset \mathsf{I}f$

[5] In LUN, the least universe $\mathsf{lt}(a)$ for a name a is defined to be the least set that contains a and is closed under the set constructions of EMU.

holds. Taking these into consideration, we shall define the least universe $\mathsf{l}f$ for a class f as the least set that reflects f and other smaller universes, and is closed under the Kripke operator. In addition, to get a strong system, we characterise least universes in terms of *proposition* as in Lemma 140, rather than *truth*.

Let $\mathrm{C}_{\mathsf{l}f}(x) := \forall y.\ \mathrm{P}_{\mathsf{l}f}(xy) := \forall y.\ \mathrm{T}(\mathsf{l}f(xy)) \vee \mathrm{T}(\mathsf{l}f(\dot\neg(xy)))$. Thus, this means that x is a class within $\mathsf{l}f$. Then, for each term f, $\mathcal{L}_{FS}$-formula B and a free variable x, the formula $\mathscr{A}^{\mathsf{LU}}(f, B(\bullet), x)$ is defined to be the disjunction of the following:

1. $\exists y, z.\ x = (y \dot= z) \wedge y = z$
2. $\exists y.\ x = (\dot{\mathsf{N}}y) \wedge \mathrm{N}(y)$
3. $\exists y.\ x = \dot\neg y \wedge B(y)$
4. $\exists y, z.\ x = (y \dot\wedge z) \wedge \{[B(y) \wedge B(z)] \vee [B(y) \wedge \mathrm{T}(\mathsf{l}f(\dot\neg y))] \vee [B(z) \wedge \mathrm{T}(\mathsf{l}f(\dot\neg z))]\}$
5. $\exists y, z.\ x = (y \dot\to z) \wedge \{[B(y) \wedge B(z)] \vee [B(y) \wedge \mathrm{T}(\mathsf{l}f(\dot\neg y))] \vee [B(z) \wedge \mathrm{T}(\mathsf{l}fz)]\}$
6. $\exists g.\ x = (\dot\forall g) \wedge \{[\forall y.\ B(gy)] \vee \exists y.\ B(gy) \wedge \mathrm{T}(\mathsf{l}f(\dot\neg(gy))\}$
7. $\exists y.\ x = fy$
8. $\exists g, y.\ x = \mathsf{l}gy \wedge \mathrm{P}_{\mathsf{l}f}(x)$

The least universe schema (LU) is given by:

$$[\mathrm{C}(f) \wedge \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{P}_{\mathsf{l}f}(x) \to B(x), \qquad (\mathsf{LU})$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

Definition 163. The $\mathcal{L}_{FS}$-theory KFLU is KFU with the schemata (UG) and (LU).

As an immediate consequence of the definition, we can show that each least universe is closed under the operator $\mathscr{A}^{\mathsf{LU}}$:

Corollary 164. $\mathsf{KFLU} \vdash \mathrm{C}(f) \to \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, \mathrm{P}_{\mathsf{l}f}(\bullet), x) \to \mathrm{P}_{\mathsf{l}f}(x)$.

For the lower bound of KFLU, we can interpret T_0 into KFLU using the same translation as for KFUPI. Moreover, the proof is almost the same as for KFUPI, except that Lemmata 147 and 148 for the derivation of $(\mathsf{IG}.2)'$ are now replaced by the following lemmata:

Lemma 165. For any $\mathcal{L}_{FS}$-formula A, recall that $B_A(u, v, w)$ is the formula $\forall x.\ \mathrm{T}(\mathsf{acc}(u, v)x) \to F_A(u, v, w, x)$, where $F_A(u, v, w, x)$ is the conjunction of the following formulas:

1. $w = t(u, v, x) \ \to\ A(x)$,
2. $\forall z.\ w = \{(\oplus(u, v)(1, (z, x))) \dot\to t(u, v, z)\} \to \mathrm{T}(\oplus(u, v)(1, (z, x))) \to A(z)$,

3. $\forall z.\ w = \{\oplus(u,v)(0,z) \dot{\to} (\oplus(u,v)(1,(z,x)) \dot{\to} t(u,v,z))\} \to \mathrm{T}(\oplus(u,v)(0,z)) \to \mathrm{T}(\oplus(u,v)(1,(z,x))) \to A(z)$.

Then, KFLU derives the following:

$$\mathrm{C}(u) \wedge \mathrm{C}(v) \wedge \mathsf{Closed}'(u,v,A(\bullet)) \to \forall w.\ \mathscr{A}^{\mathsf{LU}}(\oplus(u,v), B_A(u,v,\bullet), w) \to B_A(u,v,w).$$

Lemma 166. KFUPI derives the following:

$$\mathrm{C}(a) \wedge \mathrm{C}(b) \wedge \mathit{Closed}(a,b,A) \to \forall x.\ \mathrm{T}(\mathsf{acc}(a,b)x) \to A(x).$$

Proof. We assume $\mathrm{C}(a), \mathrm{C}(b), \mathit{Closed}(a,b,A)$ and $\mathrm{T}(\mathsf{acc}(a,b)c)$ for an arbitrary c. Then, we have to derive $A(c)$. From the assumptions and Lemma 165, we obtain $\mathrm{C}(\mathsf{l}(\oplus(a,b)))$ and $\forall w.\ \mathscr{A}^{\mathsf{LU}}(\oplus(u,v), B_A(u,v,\bullet), w) \to B_A(u,v,w)$. Thus, ($\mathsf{LU}$) yields:

$$\forall x.\ \mathrm{P}_{\mathsf{l}(\oplus(a,b))}(x) \to B_A(a,b,x).$$

Letting $x := t(a,b,c)$, we want to show $\mathrm{P}_{\mathsf{l}(\oplus(a,b))}(t(a,b,c))$, from which we have $B_A(a,b,t(a,b,c))$, and thus we can immediately obtain the conclusion $A(c)$ by the assumption $\mathrm{T}(\mathsf{acc}(a,b)c)$ and the definition of B_A.

In order to prove $\mathrm{P}_{\mathsf{l}(\oplus(a,b))}(t(a,b,c))$, it suffices to derive the following:

$$\forall y.\ \mathrm{T}(\oplus(a,b)(0,y)) \to \mathrm{T}(\oplus(a,b)(1,(y,c))) \to \mathrm{T}(\mathsf{l}(\oplus(a,b))(t(a,b,y))). \tag{5.4}$$

In fact, this formula implies $\mathrm{T}(\mathsf{l}(\oplus(a,b))(t(a,b,c)))$ and $\mathrm{P}_{\mathsf{l}(\oplus(a,b))}(t(a,b,c))$, because $\oplus(a,b)$ is a class within $\mathsf{l}(\oplus(a,b))$.

However, the formula (5.4) follows from the assumption $\mathrm{T}(\mathsf{acc}(a,b)c)$. ∎

As for the upper bound, we can interpret KFLU into $\mathsf{FID}([\mathsf{POS},\mathsf{QF}])$ in a similar manner as for KFUPI. So, we only observe that the schema (KFLU) is satisfied in $\mathcal{M}_{\mathsf{KF}}$.

Proposition 167. $\mathcal{M}_{\mathsf{KF}} \models \mathsf{KFLU}$.

Proof. By Proposition 150, we concentrate on the schema (LU):

$$[\mathrm{C}(f) \wedge \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{P}_{\mathsf{l}f}(x) \to B(x).$$

So, we take any class f and assume that $\mathcal{M}_{\mathsf{KF}} \models \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, B(\bullet), x) \to B(x)$. Now, we prove by induction on α that if $x \in Z_\alpha$ or $\dot{\neg} x \in Z_\alpha$, then $B(x)$.

For example, consider the case where x is of the form $\mathsf{l}gy$. Then, since B is $\mathscr{A}^{\mathsf{LU}}$-closed, it follows that $B(\mathsf{l}gy)$. The other cases are shown by the subinduction on the construction of Z_α (cf. Proposition 150). ∎

In conclusion, we obtain the proof-theoretic strength of KFLU :

Theorem 168. KFLU and T_0 are proof-theoretically equivalent.

Similarly to KFLU, we can also consider least universes for PT. As for the Aczel-Feferman schema, we can naturally characterise least universes in terms of *truth* only; thus, we define an operator $\mathscr{A}^{\mathsf{PTLU}}$ as follows: For each term f, $\mathcal{L}_{\mathrm{T}}$-formula B and a free variable x, the formula $\mathscr{A}^{\mathsf{PTLU}}(f, B(\bullet), x)$ is the disjunction of the following:

1. $\exists y, z.\ x = (y \dot{=} z) \wedge y = z$
2. $\exists y, z.\ x = (\dot{\neg}(y \dot{=} z)) \wedge y \neq z$
3. $\exists y.\ x = (\dot{\mathsf{N}} y) \wedge \mathrm{N}(y)$
4. $\exists y.\ x = (\dot{\neg}(\dot{\mathsf{N}} y)) \wedge \neg \mathrm{N}(y)$
5. $\exists y.\ x = \dot{\neg}(\dot{\neg} y) \wedge B(y)$
6. $\exists y, z.\ x = (y \dot{\wedge} z) \wedge B(y) \wedge B(z)$
7. $\exists y, z.\ x = (\dot{\neg}(y \dot{\wedge} z)) \wedge \{[B(\dot{\neg} y) \wedge B(\dot{\neg} z)] \vee [B(y) \wedge B(\dot{\neg} z)] \vee [B(\dot{\neg} y) \wedge B(z)]\}$
8. $\exists y, z.\ x = (y \dot{\rightarrow} z) \wedge [\mathrm{T}(\mathsf{I} f y) \vee \mathrm{T}(\mathsf{I} f (\dot{\neg} y))] \wedge [\mathrm{T}(\mathsf{I} f(y)) \rightarrow B(z)]$
9. $\exists y, z.\ x = (\dot{\neg}(y \dot{\rightarrow} z)) \wedge B(y) \wedge B(\dot{\neg} z)$
10. $\exists g.\ x = (\dot{\forall} g) \wedge \forall y. B(g y)$
11. $\exists g.\ x = (\dot{\neg}(\dot{\forall} g)) \wedge [\forall y. B(g y) \vee B(\dot{\neg}(g y))] \wedge \exists y. B(\dot{\neg}(g y))$
12. $\exists y.\ [x = f y \vee x = \dot{\neg}(f y)] \wedge \mathrm{T}(x)$
13. $\exists g, y.\ [x = \mathsf{I} g y \vee x = \dot{\neg}(\mathsf{I} g y)] \wedge \mathrm{T}(\mathsf{I} f x)$

The least universe schema ($\mathsf{LU}^{\mathsf{PT}}$) is given by:

$$[\mathrm{C}(f) \wedge \forall x.\ \mathscr{A}^{\mathsf{PTLU}}(f, B(\bullet), x) \rightarrow B(x)] \rightarrow \forall x.\ \mathrm{T}(\mathsf{I} f x) \rightarrow B(x), \qquad (\mathsf{LU}^{\mathsf{PT}})$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

Thus, the schema ($\mathsf{LU}^{\mathsf{PT}}$) informally says that $\mathrm{T}(\mathsf{I} f x)$ is the least truth set satisfying each clause of $\mathscr{A}^{\mathsf{PTLU}}$.

Definition 169. The $\mathcal{L}_{FS}$-theory PTLU is PTU with the schemata (UG) and ($\mathsf{LU}^{\mathsf{PT}}$).

In exactly the same way as for KFLU, we can determine the proof-theoretic strength of PTLU:

Theorem 170. PTLU and T_0 are proof-theoretically equivalent.

5.6 Universes for Supervaluation: VFU

In this section, we consider the supervaluational Frege structure. Kripke initially sketched a semantic theory of truth based on the supervaluation schema [46, p. 711]. Based on Kripke's semantic definition, Cantini [12] defined and studied a formal theory VF (van Fraassen) over Peano arithmetic. In particular, Cantini [12] proved that VF is proof-theoretically equivalent to the theory ID_1 of arithmetical monotone inductive definition. In [14, 40], the system VF as a theory of Frege structure is formulated and found to be proof-theoretically equivalent to the one over Peano arithmetic. For future research, Kahle [40, p. 124] suggested extending VF by adding universes, similar to KFU and PTU. Given that KFU is roughly a transfinite iteration of KF, it is natural to expect VF with universes to have the strength of a transfinite iteration of VF. Thus, Kahle conjectured that such a theory would have at least the strength of ID_α for some ordinal α (cf. [40, 42]). The purpose of this section is to implement the idea of universes for VF and to verify Kahle's conjecture. In particular, we will show that VF with universes is proof-theoretically equivalent to T_0.

5.6.1 System VF and Universes

In this section, we add a constant symbol $\dot{\mathrm{T}}$ to $\mathcal{L}_{FS}$ for technical reasons (see Remark 172). Following [40, 41], we inductively define a corresponding term $\dot{A}$ for each formula A.

- $\dot{\overbrace{s = t}} := (s \dot{=} t)$, $\dot{\overbrace{\mathrm{N}(s)}} := \dot{\mathsf{N}}s$, $\dot{\overbrace{\mathrm{T}(s)}} := \dot{\mathsf{T}}s$;
- $\dot{\overbrace{\neg A}} := \dot{\neg}\dot{A}$, $\dot{\overbrace{A \wedge B}} := \dot{A}\dot{\wedge}\dot{B}$, $\dot{\overbrace{A \to B}} := \dot{A}\dot{\to}\dot{B}$;
- $\dot{\overbrace{\forall x A}} := \dot{\forall}(\lambda x.\dot{A})$.

Similar to Definition 13, we define the $\mathcal{L}_{FS}$-theory VF[TON], the formulation of which is essentially based on [12, 14, 40].

Definition 171. The $\mathcal{L}_{FS}$-theory VF[TON] consists of TON and the following axioms:

(T-Out) $\mathrm{T}(\dot{A}) \to A$;

(T-Elem) $P \to \mathrm{T}(\dot{P})$, for any $\mathcal{L}$-literal formula P;

(T-Imp) $\mathrm{T}(\dot{\overbrace{A \to B}}) \to (\mathrm{T}(\dot{A}) \to \mathrm{T}(\dot{B}))$;

(T-Univ) $\forall x \mathrm{T}(\dot{A}) \to \mathrm{T}(\dot{\overbrace{\forall x A}})$;[6]

(T-Log) $\mathrm{T}(\dot{C})$ for any logical theorem C;

(T-Cons) $\neg[\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg} x)]$;

(T-Self) $[\mathrm{T}(x) \leftrightarrow \mathrm{T}(\dot{\mathsf{T}} x)] \ \wedge \ [\mathrm{T}(\dot{\neg} x) \leftrightarrow \mathrm{T}(\dot{\neg}(\dot{\mathsf{T}} x))]$.

Here, A and B are any $\mathcal{L}_{FS}$-formulas.

In this chapter, we express VF[TON] simply as VF, for there is no room for confusion with that of Definition 13.

Remark 172. In Kahle's formulation of VF (called SON in [40]), the constant $\dot{\mathrm{T}}$ is defined to be an identity function $\lambda x.x$, so $\dot{\mathrm{T}}s$ is β-equivalent to s itself. Consequently, the axiom (T-Self) becomes trivial [40, pp. 111–112]. While this definition is not problematic in KF (Definition 135) and PT (Definition 156), it causes a contradiction in VF. In fact, for every $\mathcal{L}_{FS}$-sentence A, we have $\mathrm{T}(\dot{A}\dot{\vee}(\dot{\neg}\dot{A}))$ by (T-Log), which is, according to Kahle's definition, equivalent to $\mathrm{T}((\dot{\mathsf{T}}\dot{A})\dot{\vee}(\dot{\mathsf{T}}(\dot{\neg}\dot{A})))$. Therefore, (T-Out) implies $\mathrm{T}(\dot{A}) \vee \mathrm{T}(\dot{\neg}\dot{A})$. However, it is well known that VF with the schema $\mathrm{T}(\dot{A}) \vee \mathrm{T}(\dot{\neg}\dot{A})$ is inconsistent (see, e.g., [20]). That is the reason we explicitly introduced the constant $\dot{\mathsf{T}}$ and instead required the axiom (T-Self).

As is often said, VF has a non-compositional nature, and thus, it is not suitable for the inductive characterisation of proposition or truth, unlike KF and PT.[7] Alternatively, we show in the next subsection that simply adding the axiom (Lim) (Definition 137) to the universe-relative version of VF gives the same strength as T_0. So, analogously to KFU and PTU, we propose the system VFU.

Recall that $\mathrm{C}(f) :\equiv \forall x.\ \mathrm{T}(fx) \vee \mathrm{T}(\dot{\neg}(fx))$.

Definition 173. The $\mathcal{L}_{FS}$-theory VFU consists of TON and the following axioms:

VF-axioms in U.

(U-Out) $\mathrm{U}(u) \to [\mathrm{T}(u(\dot{A})) \to A]$;

(U-Elem) $\mathrm{U}(u) \to [P \to \mathrm{T}(u(\dot{P}))]$, for any $\mathcal{L}$-literal formula P;

(U-Imp) $\mathrm{U}(u) \to [\mathrm{T}(u(\dot{\overbrace{A \to B}})) \to \{\mathrm{T}(u\dot{A}) \to \mathrm{T}(u\dot{B})\}]$;

[6]One might prefer the single axiom $(\forall x.\mathrm{T}(fx)) \to \mathrm{T}(\dot{\forall} f)$ similarly to $(\mathsf{K}_\forall)$ in Definition 135. The reason the author choses the schematic form is just to simplify the upper-bound proof in Section 5.6.4. Nevertheless, the author believes that this single axiom does not affect the proof-theoretic strength of both VF and VFU (Definition 173).

[7]Note that Stern's supervaluation-style truth [59] is an attempt to overcome this difficulty of VF.

(U-Univ) $\mathrm{U}(u) \to [\forall x \mathrm{T}(u(\dot{A})) \to \mathrm{T}(u(\overset{\cdot}{\overbrace{\forall x A}}))]$;

(U-Log) $\mathrm{U}(u) \to \mathrm{T}(u(\dot{C}))$ for any logical theorem C;

(U-Cons) $\mathrm{U}(u) \to \forall x.\ \neg[\mathrm{T}(ux) \wedge \mathrm{T}(u(\dot{\neg} x))]$;

(U-Self) $\mathrm{U}(u) \to [\mathrm{T}(ux) \leftrightarrow \mathrm{T}(u(\dot{\mathsf{T}}x))] \ \wedge \ [\mathrm{T}(u(\dot{\neg}x)) \leftrightarrow \mathrm{T}(u(\dot{\neg}(\dot{\mathsf{T}}x)))]$;

Here, A and B are any $\mathcal{L}_{FS}$-formulas.

Structural properties of U.

(U-Class) $\mathrm{U}(u) \to \mathrm{C}(u)$;

(U-True) $\mathrm{U}(u) \to \forall x.\ \mathrm{T}(ux) \to \mathrm{T}(x)$;

(Lim) $\mathrm{C}(f) \to \mathrm{U}(\mathsf{I}f) \wedge f \sqsubset \mathsf{I}f$.

Similarly to Fact 138, we can prove the following:

Lemma 174. VF is a subtheory of VFU.

5.6.2 Lower Bound of VFU

In this subsection, we determine the lower bound of VFU. The proof proceeds in a similar way as for KFUPI and PTUPI. Thus, we define a translation $' : \mathcal{L}_{EM} \to \mathcal{L}_{FS}$. The interpretation of $\mathcal{L}$ is exactly the same as in Theorem 149. Each generator of T_0 is interpreted as a term of $\mathcal{L}_{FS}$. In particular, the interpretation of the inductive generation i is essentially a generalisation of Cantini's original lower-bound proof of VF [12, 14].

The following two lemmata are essentially by way of [14, Lemma 59.2].

Lemma 175. An $\mathcal{L}_{FS}$-formula A is T-*positive* if each truth predicate T in A occurs only positively. Then, for every T-positive $\mathcal{L}_{FS}$-formula A,

$$\mathsf{VFU} \vdash \mathrm{T}(\dot{A}) \leftrightarrow A.$$

Lemma 176. In VFU, the following are derivable:

1. $\mathrm{T}(\dot{A}) \wedge \mathrm{T}(\dot{B}) \leftrightarrow \mathrm{T}(\overset{\cdot}{\overbrace{A \wedge B}})$;
2. $\mathrm{T}(\dot{A}) \vee \mathrm{T}(\dot{B}) \to \mathrm{T}(\overset{\cdot}{\overbrace{A \vee B}})$;
3. $(\mathrm{T}(\dot{A}) \vee \mathrm{T}(\overset{\cdot}{\overbrace{\neg A}})) \to [\{\mathrm{T}(\dot{A}) \to \mathrm{T}(\dot{B})\} \leftrightarrow \mathrm{T}(\overset{\cdot}{\overbrace{A \to B}})]$;

4. $[\forall x \mathrm{T}(\dot{A})] \leftrightarrow \mathrm{T}(\overset{\cdot}{\overbrace{\forall x A}});$
5. $[\exists x \mathrm{T}(\dot{A})] \rightarrow \mathrm{T}(\overset{\cdot}{\overbrace{\exists x A}});$
6. $[\mathrm{T}(a) \vee \mathrm{T}(\dot{\neg} a)] \rightarrow [\neg \mathrm{T}(\dot{\mathrm{T}} a) \leftrightarrow \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}} a))].$

Using these lemmata, VFU can interpret each generator except i. Here, as an example, we only deal with join j, but the other generators can be similarly treated.

Lemma 177. Let a term $\underline{\mathsf{j}}$ be such that $\underline{\mathsf{j}}(x, f) = \lambda z. \overset{\cdot}{\overbrace{\exists v, w.\ z = (v, w) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw)}}$. Then, $\mathsf{VFU} \vdash (\text{join})'$, that is,

$$\mathsf{VFU} \vdash \mathrm{C}(x) \wedge [\forall y.\ \mathrm{T}(xy) \rightarrow \mathrm{C}(fy)] \rightarrow \mathrm{C}(\underline{\mathsf{j}}(x, f)) \wedge \Sigma'(x, f, \underline{\mathsf{j}}(x, f)),$$

where, $\Sigma'(x, f, \underline{\mathsf{j}}(x, f))$ is the following:

$$\forall z.\ \mathrm{T}(\underline{\mathsf{j}}(x, f)z) \leftrightarrow \exists v, w.\ z = (v, w) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw).$$

Therefore, the term $\underline{\mathsf{j}}$ interprets the join axiom in VFU, where recall that $\mathrm{R}(x)$ and $x \in y$ are interpreted as $\mathrm{C}(x)$ and $\mathrm{T}(yx)$, respectively.

Proof. Suppose that $\mathrm{C}(x)$ and $\forall y(\mathrm{T}(xy) \rightarrow \mathrm{C}(fy))$. As the second conjunct is obvious from Lemma 175, we show that $\mathrm{C}(\underline{\mathsf{j}}(x, f))$. Take any z and assume that $\neg \mathrm{T}(\underline{\mathsf{j}}(x, f)z)$, then $\mathrm{T}(\dot{\neg}(\underline{\mathsf{j}}(x, f)z))$ is derived with the help of Lemma 176:

$$\begin{aligned}
&\neg \mathrm{T}(\overset{\cdot}{\overbrace{\exists v, w.\ z = (v, w) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw)}}) \\
&\Longrightarrow\ \neg \exists v, w.\ \mathrm{T}(\overset{\cdot}{\overbrace{z = (v, w) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw)}}) \\
&\Longleftrightarrow \neg \exists v, w.\ \mathrm{T}(z = (v, w)) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw) \\
&\Longleftrightarrow \forall v, w.\ \neg \mathrm{T}(z = (v, w)) \vee \neg \mathrm{T}(xv) \vee \neg \mathrm{T}(fvw) \\
&\Longleftrightarrow \forall v, w.\ \mathrm{T}(\dot{\neg}(z \dot{=} (v, w))) \vee \mathrm{T}(\dot{\neg}(xv)) \vee \mathrm{T}(\dot{\neg}(fvw)) \\
&\Longrightarrow \forall v, w.\ \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}}(z \dot{=} (v, w)))) \vee \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}}(xv))) \vee \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}}(fvw))) \\
&\Longrightarrow \forall v, w.\ \mathrm{T}(\overset{\cdot}{\overbrace{\neg(\mathrm{T}(z = (v, w)) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw))}}) \\
&\Longleftrightarrow \mathrm{T}(\dot{\neg}(\underline{\mathsf{j}}(x, f)z)).
\end{aligned}$$

■

To complete the interpretation of T_0, we have to give the interpretation of inductive generation i. We define $\underline{\mathsf{i}}$ as the following term acc :

$\mathsf{acc}(a,b) := \lambda z.\overbrace{TI[a,b,z]}^{\cdot}$, where
$WF[a,b,f] := \forall x.\ \mathrm{T}(ax) \to [\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))] \to \mathrm{T}(\mathsf{I}_{a,b}(fx))$;
$TI[a,b,z] := \mathrm{T}(az) \land \forall f.\ WF[a,b,f] \to \mathrm{T}(\mathsf{I}_{a,b}(fz))$.

The term $\mathsf{I}_{a,b}$ is defined below. Informally speaking, $\mathsf{acc}(a,b)$ is the intersection of every set f that includes the $<_b$-accessible part of a.

Lemma 178. For each class u and v, we can take a universe $\mathsf{I}_{u,v}$ that reflects on both u and v. That is,

$$\mathsf{VFU} \vdash \forall u,v.\ \mathrm{C}(u) \land \mathrm{C}(v) \to [\mathrm{U}(\mathsf{I}_{u,v}) \land u \sqsubset \mathsf{I}_{u,v} \land v \sqsubset \mathsf{I}_{u,v}].$$

Proof. Taking any classes u and v, we define a term $\oplus$ such that

$$u \oplus v = \lambda x.\overbrace{\mathrm{N}((x)_0) \to \mathrm{T}(\mathsf{d}_{\mathsf{N}}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0})}^{\cdot},$$

for which we further let $\mathsf{I}_{u,v} := \mathsf{I}(u \oplus v)$. First, to show that $\mathsf{I}_{u,v}$ is a class, we prove that $u \oplus v$ is a class. Thus, taking any object x, we prove that $(u \oplus v)x$ is a proposition. If $\neg\mathrm{N}((x)_0)$, then we have $\mathrm{T}(\overbrace{\mathrm{N}((x)_0) \to \mathrm{T}(\mathsf{d}_{\mathsf{N}}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0})}^{\cdot})$, hence $\mathrm{P}((u \oplus v)x)$. Thus, we can assume $\mathrm{N}((x)_0)$. If $(x)_0 = \mathsf{0}$, then we have $\mathsf{d}_{\mathsf{N}}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0} = u(x)_1$. As u is a class, $(u \oplus v)x$ is a proposition, as required. Similarly, if $(x)_0 \neq \mathsf{0}$, then $\mathsf{d}_{\mathsf{N}}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0} = v(x)_1$, and thus $(u \oplus v)x$ is a proposition. Thus, in any case, $(u \oplus v)x$ is a proposition. Therefore, $u \oplus v$ is a class, and thus (Lim) yields that $\mathsf{I}_{u,v}$ is a class. Second, we show that $u \sqsubset \mathsf{I}_{u,v}$. For an arbitrary object x, assume $\mathrm{T}(ux)$. Then, since $(u \oplus v)(\mathsf{0},x) = \overbrace{\mathrm{N}(\mathsf{0}) \to \mathrm{T}(ux)}^{\cdot}$, it follows that $\mathrm{T}((u\oplus v)(\mathsf{0},x))$; thus, we obtain $\mathrm{T}(\mathsf{I}_{u,v}((u\oplus v)(\mathsf{0},x)))$, which, by ($\mathsf{U\text{-}Log}$), ($\mathsf{U\text{-}Imp}$), and ($\mathsf{U\text{-}Self}$), implies $\mathrm{T}(\mathsf{I}_{u,v}(ux))$. Similarly, $\mathrm{T}(\dot\neg(ux))$ implies $\mathrm{T}(\mathsf{I}_{u,v}(\dot\neg(ux)))$. Thus, the conclusion $u \sqsubset \mathsf{I}_{u,v}$ is obtained. In the same way, we also have $v \sqsubset \mathsf{I}_{u,v}$. ■

Lemma 179. 1. $\mathsf{VFU} \vdash \mathrm{C}(a) \land \mathrm{C}(b) \to \mathrm{C}(\mathsf{acc}(a,b))$;

2. $\mathsf{VFU} \vdash \mathrm{C}(a) \land \mathrm{C}(b) \to \mathsf{Closed}'(a,b,\mathsf{acc}(a,b))$;
3. $\mathsf{VFU} \vdash \mathrm{C}(a) \land \mathrm{C}(b) \land \mathsf{Closed}'(a,b,A(\bullet)) \to \forall x.\ \mathrm{T}(\mathsf{acc}(a,b)x) \to A(x)$, for each $\mathcal{L}_{FS}$-formula A.

Proof. We assume that $\mathrm{C}(a)$ and $\mathrm{C}(b)$.

1. Take any z; then we have to prove $\mathrm{P}(\mathsf{acc}(a,b)z)$. Therefore, supposing $\neg\mathrm{T}(\mathsf{acc}(a,b)z)$, we show $\mathrm{T}(\dot\neg(\mathsf{acc}(a,b)z))$. For that purpose, we prove that $WF[a,b,f]$ is a proposition for any f. First, by repeated use of Lemma 176, we observe that $\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot}$ is a proposition:

$$
\begin{aligned}
&\neg\mathrm{T}(\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})\\
&\iff \neg\forall y.\ \mathrm{T}(\overbrace{\mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})\\
&\iff \neg\forall y.\ \mathrm{T}(\overbrace{\mathrm{T}(ay)}^{\cdot}) \to \mathrm{T}(\overbrace{\mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})\\
&\iff \neg\forall y.\ \mathrm{T}(\overbrace{\mathrm{T}(ay)}^{\cdot}) \to \mathrm{T}(\overbrace{\mathrm{T}(b(y,x))}^{\cdot}) \to \mathrm{T}(\overbrace{\mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})\\
&\iff \exists y.\ \mathrm{T}(\overbrace{\mathrm{T}(ay)}^{\cdot}) \land \mathrm{T}(\overbrace{\mathrm{T}(b(y,x))}^{\cdot}) \land \neg\mathrm{T}(\overbrace{\mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})\\
&\iff \exists y.\ \mathrm{T}(\overbrace{\mathrm{T}(ay)}^{\cdot}) \land \mathrm{T}(\overbrace{\mathrm{T}(b(y,x))}^{\cdot}) \land \mathrm{T}(\overbrace{\neg\mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})\\
&\iff \exists y.\ \mathrm{T}(\overbrace{\neg(\mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy)))}^{\cdot})\\
&\implies \mathrm{T}(\overbrace{\neg\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot}).
\end{aligned}
$$

Therefore, it follows that $WF[a,b,f]$ is a proposition:

$$
\begin{aligned}
&\neg\mathrm{T}(\overbrace{WF[a,b,f]}^{\cdot})\\
&\iff \neg\forall x.\ \mathrm{T}(\overbrace{\mathrm{T}(ax) \to [\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))] \to \mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})\\
&\iff \neg\forall x.\ \mathrm{T}(\overbrace{\mathrm{T}(ax)}^{\cdot}) \to \mathrm{T}(\overbrace{[\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))] \to \mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})\\
&\iff \exists x.\ \mathrm{T}(\overbrace{\mathrm{T}(ax)}^{\cdot}) \land \neg\mathrm{T}(\overbrace{[\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))] \to \mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})\\
&\iff \exists x.\ \mathrm{T}(\overbrace{\mathrm{T}(ax)}^{\cdot}) \land \mathrm{T}(\overbrace{[\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))] \land \neg\mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})\\
&\iff \exists x.\ \mathrm{T}(\overbrace{\mathrm{T}(ax) \land [\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))] \land \neg\mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})\\
&\implies \mathrm{T}(\overbrace{\neg WF[a,b,f]}^{\cdot}).
\end{aligned}
$$

Using this, we can similarly get $\mathrm{T}(\dot\neg(\mathsf{acc}(a,b)z))$ from $\neg\mathrm{T}(\mathsf{acc}(a,b)z)$.

2. Assume that $\mathrm{T}(ax)$ and $\forall y[\mathrm{T}(ay) \to (\mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{acc}(a,b)y))]$; then we want to derive $\mathrm{T}(\mathsf{acc}(a,b)x)$:

$$
\begin{aligned}
&\mathrm{T}(\overbrace{\forall f.\ WF[a,b,f] \to \mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot}) \\
&\iff \forall f.\ \mathrm{T}(\overbrace{WF[a,b,f] \to \mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot}) \\
&\iff \forall f.\ \mathrm{T}(\overbrace{WF[a,b,f]}^{\cdot}) \to \mathrm{T}(\overbrace{\mathrm{T}(l_{a,b}(fx))}^{\cdot}).
\end{aligned}
$$

To show the last formula, we take any f and suppose $\mathrm{T}(\overbrace{WF[a,b,f]}^{\cdot})$. Then, we need to derive $\mathrm{T}(\overbrace{\mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})$. By the assumption, for any y such that $\mathrm{T}(ay)$ and $\mathrm{T}(b(y,x))$ we have $\mathrm{T}(\mathsf{acc}(a,b)y)$. Thus, the supposition $\mathrm{T}(\overbrace{WF[a,b,f]}^{\cdot})$ implies that $\mathrm{T}(\overbrace{\mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot})$. As y is arbitrary, this implies that

$$\mathrm{T}(\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(fy))}^{\cdot}).$$

Combining this with the assumption $\mathrm{T}(ax)$, the desired conclusion $\mathrm{T}(\overbrace{\mathrm{T}(\mathsf{I}_{a,b}(fx))}^{\cdot})$ follows from $\mathrm{T}(\overbrace{WF[a,b,f]}^{\cdot})$.

3. Take any x and assume that $\mathsf{Closed}'(a,b,A(\bullet))$ and $\mathrm{T}(\mathsf{acc}(a,b)x)$; we show $A(x)$. Let $A'(x) := \mathsf{Closed}'(a,b,A(\bullet)) \to A(x)$, then we easily have $\mathsf{Closed}'(a,b,A'(\bullet))$ by logic. Thus, in VFU, we also obtain $\mathrm{T}(\mathsf{I}_{a,b}(\overbrace{\mathsf{Closed}'(a,b,A'(\bullet))}^{\cdot}))$. Next, we can derive $\mathsf{Closed}'(a,b,\mathrm{T}(\mathsf{I}_{a,b}(A'(\bullet))))$ in the following way:

$$
\begin{aligned}
&\mathrm{T}(\mathsf{I}_{a,b}(\overbrace{\mathsf{Closed}'(a,b,A'(\bullet))}^{\cdot})) \\
&\implies \forall x.\ \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{\mathrm{T}(ax)}^{\cdot})) \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to A'(y)}^{\cdot})) \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(x)}^{\cdot})) \\
&\iff \forall x.\ \mathrm{T}(ax) \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to A'(y)}^{\cdot})) \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(x)}^{\cdot})) \\
&\iff \forall x.\ \mathrm{T}(ax) \to [\forall y.\ \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{\mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to A'(y)}^{\cdot}))] \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(x)}^{\cdot})) \\
&\iff \forall x.\ \mathrm{T}(ax) \to [\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(y)}^{\cdot}))] \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(x)}^{\cdot})).
\end{aligned}
$$

Letting $f := \lambda x.\overbrace{A'(x)}^{\cdot}$, the assumption $\mathrm{T}(\mathsf{acc}(a,b)x)$ implies in VFU the formula $\mathsf{Closed}'(a,b,\mathrm{T}(\mathsf{I}_{a,b}(A'(\bullet)))) \to \mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(x)}^{\cdot}))$. Therefore, we obtain $\mathrm{T}(\mathsf{I}_{a,b}(\overbrace{A'(x)}^{\cdot}))$, which yields $A'(x)$ by ($\mathsf{U\text{-}Out}$). Finally, combining this with the assumption $\mathsf{Closed}'(a,b,A(\bullet))$, the conclusion $A(x)$ follows.

■

Theorem 180. For each $\mathcal{L}_{FS}$-sentence A, if $\mathsf{T}_0 \vdash A$, then $\mathsf{VFU} \vdash A'$. In particular, every $\mathcal{L}$-theorem of T_0 is derivable in VFU.

5.6.3 Truth-as-Provability Interpretation of VFU

In this subsection, we give a model of VFU by generalising Cantini's *truth-as-provability* interpretation for VF [12, 14], which is formalisable in a suitable set theory, and thus the upper-bound of VFU is obtained (see subsection 5.6.4). The idea of our truth-as-provability interpretation is that the truth predicate $\mathrm{T}(x)$ is intepreted as the derivability of x in the indexed infinitary sequent calculus, as is displayed on the table below. Then, each axiom of VFU is shown to be true under this interpretation. Here, a *sequent* Γ is a finite set of closed terms, each of which is β-equivalent to $\dot{A}$ for some sentence A. To present the system in the form of Tait calculus, we consider only negation normal sentences. Therefore, the negation symbol $\neg$ may come only in front of atomic sentences, and then the global negation $\neg A$ becomes a defined expression with the help of De Morgan's law. Note that the conditional $A \to B$ is defined by $\neg A \vee B$. For simplicity, we do not distinguish terms which have the same reduct, thus we can suppose that every term is of the form $\dot{A}$ for some negation normal sentence A. For readability, we often simply write A instead of $\dot{A}$. Moreover, since we also consider terms of the form $\mathsf{l}ab$, it is useful to treat them as if they were sentences. Thus, we introduce a new binary predicate symbol $\mathrm{L}_x(y)$ and we let $\overbrace{\mathrm{L}_a(b)}^{\cdot} := \mathsf{l}ab$. Similarly, let $\overbrace{\neg\mathrm{L}_a(b)}^{\cdot} := \dot{\neg}(\mathsf{l}ab)$.

Next, we explain the calculus in more detail. In the calculus, the predicate $\vdash^{\alpha,\beta,\gamma} \Gamma$ means that Γ is derived in the system with the U-rank α, the T-rank β, and the derivation length γ. We introduce several notations:

- $\vdash^{\alpha} \Gamma$ means $\vdash^{\alpha,\beta,\gamma} \Gamma$ for some α, β, γ;
- $\vdash^{<\alpha} \Gamma$ means $\vdash^{\alpha_0} \Gamma$ for some $\alpha_0 < \alpha$;
- $\vdash^{\alpha,\beta} \Gamma$ means $\vdash^{\alpha,\beta,\gamma} \Gamma$ for some γ;
- $\vdash^{\alpha,\beta,<\gamma} \Gamma$ means $\vdash^{\alpha,\beta,\gamma_0} \Gamma$ for some $\gamma_0 < \gamma$;
- $\vdash^{\alpha,<\beta,<\gamma} \Gamma$ means $\vdash^{\alpha,\beta_0,\gamma_0} \Gamma$ for some $\beta_0 < \beta$ and $\gamma_0 < \gamma$;
- $\vdash^{\leq\alpha,\leq\beta,\leq\gamma} \Gamma$ means $\vdash^{\alpha_0,\beta_0,\gamma_0} \Gamma$ for some $\alpha_0 \leq \alpha$, $\beta_0 \leq \beta$ and $\gamma_0 \leq \gamma$;
- $\vdash \Gamma$ means $\vdash^{\alpha} \Gamma$ for some α;

- if $\neg$ comes to the left of one of the above expressions, it negates the whole expression. For example, $\neg \vdash \Gamma$ means that it is not the case that $\vdash \Gamma$.

Under these conventions, each rule is explained as follows. The rule (Lit) says that an $\mathcal{L}$-literal P that is true in the closed term model $\mathcal{CTT}$ is derivable. The rules (Log), $(\wedge), (\vee)_i, (\exists), (\forall)$ are given similarly to the standard sequent calculus. In particular, $(\forall)$ has infinitely many premises for each closed term a. The rules (T) and ($\neg$T) respectively introduce T and $\neg$T, with an increase in the T-rank. Note that the context of the premise of (T) and ($\neg$T) must be empty; otherwise, the system would be inconsistent according to the liar paradox. The rule (Weak) assures the monotonicity of the derivability with respect to the U-rank. Similarly to the operator Φ in $\mathcal{M}_{\mathsf{KF}}$, the rules (U) and ($\neg$U) have the side condition †, which consists of the following conditions. First, α needs to be a successor ordinal (cf. the operator Φ in $\mathcal{M}_{\mathsf{KF}}$). Thanks to this condition, we can assure that $\vdash^{<\alpha}$ is closed under (Lit), (Log), $(\wedge)$, $(\vee)_i$, $(\exists)$ and $(\forall)$. In particular, if $\vdash^{<\alpha} \Gamma, A(b)$ for all b, then $\vdash^{<\alpha} \Gamma, \forall x.\ A(x)$ holds. Second, a must satisfy the following:

$$\vdash^{<\alpha} ac \text{ or } \vdash^{<\alpha} \dot{\neg}(ac) \text{ holds for all closed terms } c.$$

Thus, it roughly says that a is a class, provably in $\vdash^{<\alpha}$. We express this property as $\vdash^{<\alpha} a : \text{Class}$. The third condition is that neither $\mathrm{L}_a(b)$ nor $\neg\mathrm{L}_a(b)$ are derived in $\vdash^{<\alpha}$:

$$\neg \vdash^{<\alpha} \mathrm{L}_a(b), \text{ and } \neg \vdash^{<\alpha} \neg\mathrm{L}_a(b).$$

Thus, it roughly says that $\mathrm{L}_a(b)$ is not a proposition in $\vdash^{<\alpha}$. We express this as $\neg \vdash^{<\alpha} \mathrm{L}_a(b) : \text{Prop}$.

By transfinite induction, we easily obtain the following:

Lemma 181. (Consistency) The empty sequent is not derivable: $\neg \vdash \emptyset$.

(Weakening) If $\vdash^{\leq\alpha,\leq\beta,\leq\gamma} \Gamma$, then $\vdash^{\alpha,\beta,\gamma} \Gamma, \Delta$.

We now show the cut-admissibility of the calculus. For a formula A, the logical complexity $\mathrm{co}(A)$ is defined as usual: if P is any literal of $\mathcal{L} \cup \{\mathrm{T}(x), \mathrm{L}_x(y)\}$, then $\mathrm{co}(P) := 0$; $\mathrm{co}(A \wedge B) := \mathrm{co}(A \vee B) := \max(\mathrm{co}(A), \mathrm{co}(B)) + 1$; $\mathrm{co}(\forall x.A(x)) := \mathrm{co}(\exists x.A(x)) := \mathrm{co}(A(x)) + 1$.

Lemma 182. (Cut-admissibility) If $\vdash^{\alpha,\beta,\gamma} \Gamma, A$ and $\vdash^{\delta,\varepsilon,\zeta} \Delta, \neg A$, then $\vdash^{\max(\alpha,\delta)} \Gamma, \Delta$.

Proof. We show the claim by septuple induction on α, δ, β, ε, $\mathrm{co}(A)$, γ, and ζ. The case where either Γ, A or $\Delta, \neg A$ is obtained by (Weak) is clear by the induction

Table 5.2: Sequent system $\vdash^{\alpha,\beta,\gamma}$

$$\frac{\mathcal{CTT} \models P}{\vdash^{\alpha,\beta,\gamma} \Gamma, P}\ (\mathrm{Lit}) \qquad \frac{}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{T}(a), \neg\mathrm{T}(a)}\ (\mathrm{Log})$$

$$\frac{\vdash^{\alpha,\beta,<\gamma} \Gamma, A_0 \wedge A_1, A_0 \quad \vdash^{\alpha,\beta,<\gamma} \Gamma, A_0 \wedge A_1, A_1}{\vdash^{\alpha,\beta,\gamma} \Gamma, A_0 \wedge A_1}\ (\wedge) \qquad \frac{\vdash^{\alpha,\beta,<\gamma} \Gamma, A_0 \vee A_1, A_i\ (i \leq 1)}{\vdash^{\alpha,\beta,\gamma} \Gamma, A_0 \vee A_1}\ (\vee)_i$$

$$\frac{\vdash^{\alpha,\beta,<\gamma} \Gamma, \exists x A(x), A(a)}{\vdash^{\alpha,\beta,\gamma} \Gamma, \exists x A(x)}\ (\exists) \qquad \frac{\vdash^{\alpha,\beta,<\gamma} \Gamma, \forall x A(x), A(a),\ \text{for all } a}{\vdash^{\alpha,\beta,\gamma} \Gamma, \forall x A(x)}\ (\forall)$$

$$\frac{\vdash^{\alpha,<\beta,<\gamma} A}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{T}(\dot{A})}\ (\mathrm{T}) \qquad \frac{\vdash^{\alpha,<\beta,<\gamma} \neg A}{\vdash^{\alpha,\beta,\gamma} \Gamma, \neg\mathrm{T}(\dot{A})}\ (\neg\mathrm{T})$$

$$\frac{\vdash^{<\alpha} \Gamma}{\vdash^{\alpha,\beta,\gamma} \Gamma}\ (\mathrm{Weak})$$

$$\frac{\vdash^{<\alpha} b}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{L}_a(b)}\ (\mathrm{U})^\dagger \qquad \frac{\neg \vdash^{<\alpha} b}{\vdash^{\alpha,\beta,\gamma} \Gamma, \neg\mathrm{L}_a(b)}\ (\neg\mathrm{U})^\dagger$$

hypothesis. Thus, we can rule out such a case. If A or $\neg A$ is not principal in the last rule, then the conclusion follows by the induction hypothesis. For example, assume that $\Delta, \neg A$ is derived by $(\forall)$ from the premises $\vdash^{\delta,\varepsilon,\zeta_a} \Delta_a, \neg A$ with $\zeta_a < \zeta$ for all closed terms a, then the induction hypothesis yields $\vdash^{\max(\alpha,\delta)} \Gamma, \Delta_a$. Then, $(\forall)$ derives $\vdash^{\max(\alpha,\delta)} \Gamma, \Delta$, as required.

Finally, we consider the case where both A and $\neg A$ are principal. The inductive case $\mathrm{co}(A) > 0$ is proved by a standard cut-elimination argument (cf. [14, Theorem 62.1]). Thus, we confine ourselves to the base cases $A \equiv \mathrm{T}(\dot{B})$ and $A \equiv \mathrm{L}_a(\dot{B})$.

$A \equiv \mathrm{T}(\dot{B})$ First, if Γ, A is an instance of (Log), then $\neg\mathrm{T}(\dot{B})$ is contained in Γ, hence we have $\neg\mathrm{T}(\dot{B}), \Delta \subseteq \Gamma, \Delta$. Therefore, Lemma 181 implies the conclusion. The case where $\Delta, \neg A$ is (Log) is similar. Secondly, we assume that Γ, A and $\Delta, \neg A$ are respectively obtained by (T) and $(\neg\mathrm{T})$:

$$\frac{\mid\!\!\frac{\alpha,\beta',\gamma'}{}\ B}{\mid\!\!\frac{\alpha,\beta,\gamma}{}\ \Gamma, \mathrm{T}(\dot{B})}\ (\mathrm{T}) \qquad \frac{\mid\!\!\frac{\delta,\varepsilon',\zeta'}{}\ \neg B}{\mid\!\!\frac{\delta,\varepsilon,\zeta}{}\ \Delta, \neg\mathrm{T}(\dot{B})}\ (\neg\mathrm{T}),$$

where $\beta' < \beta, \gamma' < \gamma, \varepsilon' < \varepsilon$ and $\zeta' < \zeta$. Since $\beta' < \beta$, the induction hypothesis for the premises yields that $\mid\!\!\frac{\max(\alpha,\delta)}{}\ \emptyset$, which contradicts Lemma 181. Thus, this case cannot occur.

$A \equiv \mathrm{L}_a(b)$ As a crucial case, we suppose that $\Gamma, \mathrm{L}_a(b)$ and $\Delta, \neg\mathrm{L}_a(b)$ are obtained by (U) and ($\neg$U), respectively:

$$\frac{\mid\!\!\frac{<\alpha}{}\ b}{\mid\!\!\frac{\alpha,\beta,\gamma}{}\ \Gamma, \mathrm{L}_a(b)}\ (\mathrm{U})^{\dagger} \qquad \frac{\neg \mid\!\!\frac{<\delta}{}\ b}{\mid\!\!\frac{\delta,\varepsilon,\zeta}{}\ \Delta, \neg\mathrm{L}_a(b)}\ (\neg\mathrm{U})^{\dagger}.$$

Here, we have the following side conditions:

$$\neg \mid\!\!\frac{<\alpha}{}\ \mathrm{L}_a(b) : \mathrm{Prop} \tag{5.5}$$

$$\neg \mid\!\!\frac{<\delta}{}\ \mathrm{L}_a(b) : \mathrm{Prop} \tag{5.6}$$

By (5.5) and (5.6), we clearly have $\alpha = \delta$. Therefore, the premise $\mid\!\!\frac{<\alpha}{}\ b$ is identical with $\mid\!\!\frac{<\delta}{}\ b$, which contradicts the other premise $\neg \mid\!\!\frac{<\delta}{}\ b$. Thus, this case cannot occur.

■

Using the cut-admissibility, we can give a model of VFU. The $\mathcal{L}_{FS}$-model $\mathcal{M}_{\mathsf{VFU}}$ is an expansion of $\mathcal{CTT}$, in which the vocabularies of $\mathcal{L}$ and the additional constant symbols of $\mathcal{L}_{FS}$ are interpreted in the same way as for $\mathcal{M}_{\mathsf{KF}}$ in Section 5.3.3. Then, $\mathrm{T}(x)$ is interpreted as $\vdash \{x'\}$, where $\{x'\}$ is the singleton of the negation normal form x' of x. For simplicity, we write $\vdash x$ instead of $\vdash \{x'\}$. Similarly, $\mathrm{U}(x)$ is interpreted as the statement that for some closed term a, x is of the form $\mathsf{l}a$ and $\vdash a : \mathrm{Class}$ holds.

The next lemma verifies the VF-axioms in U of Definition 173.

Lemma 183. Let a and b be any closed terms and suppose $\vdash a : \mathrm{Class}$. Then, the following hold in $\mathcal{M}_{\mathsf{VFU}}$:

($\mathsf{U\text{-}Elem}$) If $\mathcal{CTT} \models P$, then $\vdash \mathrm{L}_a(\dot{P})$, for each $\mathcal{L}$-literal sentence P.

($\mathsf{U\text{-}Imp}$) If $\vdash \mathrm{L}_a(\overbrace{A \to B}^{\cdot})$ and $\vdash \mathrm{L}_a(\dot{A})$, then $\vdash \mathrm{L}_a(\dot{B})$.

(U-Univ) If $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{A(c)}})$ for every closed term c, then $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{\forall x.\ A(x)}})$.

(U-Log) $\vdash \mathrm{L}_a(\dot{A})$ for each logical theorem A.

(U-Cons) It is not the case that both $\vdash \mathrm{L}_a(b)$ and $\vdash \mathrm{L}_a(\dot{\neg} b)$.

(U-Self) $\vdash \mathrm{L}_a(b)$ if and only if $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{\mathrm{T}(b)}})$. Similarly, $\vdash \mathrm{L}_a(\dot{\neg} b)$ if and only if $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{\neg\mathrm{T}(b)}})$.

Proof.

(U-Elem) Assuming $\mathcal{M}_{\mathsf{VFU}} \models P$, we show that $\vdash \mathrm{L}_a(\dot{P})$. From the supposition, we can take the least successor ordinal α such that $\vdash^{<\alpha} a$: Class. Then, we obviously have $\neg \vdash^{<\alpha} \mathrm{L}_a(\dot{P})$: Prop. In addition, we have $\vdash^{0} P$ by (Lit); thus, by (U), we can deduce $\vdash^{\alpha} \mathrm{L}_a(\dot{P})$, as required.

(U-Imp) Assuming $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{A \to B}})$ and $\vdash \mathrm{L}_a(\dot{A})$, we have to prove $\vdash \mathrm{L}_a(\dot{B})$. Similarly to the above, we take the least successor ordinal α such that $\vdash^{<\alpha} a$: Class. Then, we clearly have $\neg \vdash^{<\alpha} \mathrm{L}_a(\dot{A})$: Prop and $\vdash^{\alpha} \mathrm{L}_a(\dot{A})$: Prop. Here, if $\vdash^{\alpha} \neg\mathrm{L}_a(\dot{A})$, then the assumption $\vdash \mathrm{L}_a(\dot{A})$ implies $\vdash \emptyset$ by Lemma 182, which contradicts Lemma 181. Thus, $\vdash^{\alpha} \mathrm{L}_a(\dot{A})$, which yields $\vdash^{<\alpha} A$. Similarly, we also have $\vdash^{<\alpha} A \to B$. Therefore, again by Lemma 182, it follows that $\vdash^{<\alpha} B$, and thus we obtain $\vdash^{\alpha} \mathrm{L}_a(\dot{B})$ by (U).

(U-Univ) Assuming $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{A(c)}})$ for any closed term c, we show $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{\forall x A(x)}})$. Similarly to the above, we take the least successor ordinal α such that $\vdash^{<\alpha} a$: Class. Then, for all c, we clearly have $\vdash^{\alpha} \mathrm{L}_a(\overset{\cdot}{\overbrace{A(c)}})$, and thus $\vdash^{<\alpha} A(c)$. Since α is a successor ordinal, we also have $\vdash^{\alpha-1} A(c)$ for all c, which implies $\vdash^{\alpha-1} \forall x A(x)$ by the rule ($\forall$). Thus, it follows that $\vdash \mathrm{L}_a(\overset{\cdot}{\overbrace{\forall x A(x)}})$.

The other cases are similarly proved. ∎

Similarly to Lemma 183, the structural properties of VFU are satisfied in $\mathcal{M}_{\mathsf{VFU}}$:

Lemma 184. Let a and b be any closed terms and suppose $\vdash a$: Class. Then, the following hold in $\mathcal{M}_{\mathsf{VFU}}$:

($\mathsf{U\text{-}Class}$) $\vdash \mathrm{L}_a(b)$ or $\vdash \neg\mathrm{L}_a(b)$.

($\mathsf{U\text{-}True}$) If $\vdash \mathrm{L}_a(\dot{A})$, then $\vdash A$.

(Lim) If $\vdash ab$, then $\vdash \mathrm{L}_a(ab)$. If $\vdash \dot{\neg}(ab)$, then $\vdash \mathrm{L}_a(\dot{\neg}(ab))$.

Finally, we verify the axiom ($\mathsf{U\text{-}Out}$) :

Lemma 185. Let a be any closed term and suppose $\vdash a$: Class. Then, the following is satisfied in $\mathcal{M}_{\mathsf{VFU}}$:

($\mathsf{U\text{-}Out}$) $\vdash \mathrm{L}_a(\dot{A})$ implies $\mathcal{M}_{\mathsf{VFU}} \models A$ for any $\mathcal{L}_{FS}$-sentence A.

Proof. Since we observed that ($\mathsf{U\text{-}True}$) is satisfied in $\mathcal{M}_{\mathsf{VFU}}$, it suffices to show that $\vdash A$ implies $\mathcal{M}_{\mathsf{VFU}} \models A$. Then, in exactly the same way as for [14, Theorem 63.4], we can prove, by transfinite induction, that if $\vdash \Gamma$ for a sequent Γ which consists only of $\mathcal{L}_{FS}$-sentences, then at least one sentence of Γ is true in $\mathcal{M}_{\mathsf{VFU}}$. ■

In conclusion, every theorem of VFU is satisfied in $\mathcal{M}_{\mathsf{VFU}}$.

Theorem 186. $\mathcal{M}_{\mathsf{VFU}} \models \mathsf{VFU}$.

Remark 187. Similar to Remark 151, we can also verify the axioms ($\mathsf{U\text{-}Tran}$), ($\mathsf{U\text{-}Dir}$), ($\mathsf{U\text{-}Nor}$) and ($\mathsf{U\text{-}Lin}$) in $\mathcal{M}_{\mathsf{VFU}}$.

5.6.4 Upper Bound of VFU

To determine the upper bound of VFU, we want to formalise the model $\mathcal{M}_{\mathsf{VFU}}$ of the previous subsection. However, the theory $\mathsf{FID}([\mathsf{POS},\mathsf{QF}])$ of Section 5.3.4 is not expressive enough to formalise the cut-admissibility argument of Lemma 182. Thus, we will construct $\mathcal{M}_{\mathsf{VFU}}$ within the Kripke-Platek set theory KPi, which is proof-theoretically equivalent to $\mathsf{FID}([\mathsf{POS},\mathsf{QF}])$, and hence it follows that $\mathsf{VFU} \leq \mathsf{T}_0$.

For the formulation of KPi, we follow [36]. For the language $\mathcal{L}'$ of first-order Peano arithmetic, let $\mathcal{L}^* := \mathcal{L}' \cup \{\in, \mathsf{N}, \mathrm{S}, \mathrm{Ad}\}$, where $\in$ is the membership relation symbol; N is the set constant for the natural numbers; S is the unary predicate symbol, expressing that a given object is a set; and the unary predicate symbol Ad says that an object is an admissible set. Moreover, we assume that $\mathcal{L}^*$ contains restricted quantifiers $\forall x \in y$ and $\exists x \in y$ as primitive symbols. In $\mathcal{L}^*$, the equality symbol $=$ is defined as the following formula: $(a = b) := [a \in \mathsf{N} \wedge b \in \mathsf{N} \wedge a =_{\mathsf{N}} b] \vee [\mathrm{S}(a) \wedge \mathrm{S}(b) \wedge (\forall x \in a.\, x \in b) \wedge (\forall x \in b.\, x \in a)]$, where $=_{\mathsf{N}}$ is a primitive recursive equality on natural numbers, and thus is contained in $\mathcal{L}'$. An $\mathcal{L}^*$-formula A is Δ_0

if A contains no unrestricted quantifiers. Let $\mathrm{Tran}(x)$ be a defined Δ_0-predicate that expresses that x is a transitive set. For the $\mathcal{L}^*$-formula A, let A^a be the result of replacing each unrestricted quantifier $\exists x.()$ and $\forall x.()$ in A by $\exists x \in a.()$ and $\forall x \in a.()$, respectively.

Definition 188. (cf. [36]) The $\mathcal{L}^*$-theory KPi consists of the following axioms:

N-Induction and foundation. For all $\mathcal{L}^*$-formulas A,

- $A(0) \wedge [\forall x \in \mathsf{N}.\ A(x) \to A(x+1)] \to \forall x \in \mathsf{N}.\ A(x)$;
- $[\forall x.\ (\forall y \in x.\ A(y)) \to A(x)] \to \forall x.\ A(x)$.

Ontological axioms. For all terms a, b and $\vec{c}$ of $\mathcal{L}^*$, all function symbols h and relation symbols R of $\mathcal{L}'$ and all axioms $A(\vec{x})$ of Set-theoretic axioms whose free variables belong to $\vec{x}$,

- $a \in \mathsf{N} \leftrightarrow \neg \mathrm{S}(a)$;
- $\vec{c} \in \mathsf{N} \to h(\vec{c}) \in \mathsf{N}$;
- $R(\vec{c}) \to \vec{c} \in \mathsf{N}$;
- $a \in b \to \mathrm{S}(b)$;
- $\mathrm{Ad}(a) \to [\mathsf{N} \in a \wedge \mathrm{Tran}(a)]$;
- $\mathrm{Ad}(a) \to \forall \vec{x} \in a.\ A^a(\vec{x})$.

Number-theoretic axioms. For all axioms $A(\vec{x})$ of Peano arithmetic whose free variables belong to $\vec{x}$,

- $\forall \vec{x} \in \mathsf{N}.\ A^{\mathsf{N}}(\vec{x})$.

Set-theoretic axioms. For all terms a and b and all Δ_0-formulas $A(x)$ and $B(x, y)$ of $\mathcal{L}^*$,

Pair. $\exists x.\ a \in x \wedge b \in x$;

Transitive Hull. $\exists x.\ a \subset x \wedge \mathrm{Tran}(x)$;

Δ_0-Separation. $\exists y.\ \mathrm{S}(y) \wedge y = \{x \in a : A(x)\}$;

Δ_0-Collection. $[\forall x \in a. \exists y.\ B(x, y)] \to \exists z.\ \forall x \in a.\ \exists y \in z.\ B(x, y)$;

Limit axiom. $\forall x.\ \exists y.\ x \in y \wedge \mathrm{Ad}(y)$.

Now, we describe how $\mathcal{M}_{\mathsf{VFU}}$ is formalised in KPi. Since KPi contains Peano arithmetic, the interpretation of $\mathcal{L}$ and the constant symbols of $\mathcal{L}_{FS}$ can be given by using fixed Gödel numbering. Thus, the remaining task is to formalise the sequent calculus $\vdash$ in the previous subsection for the interpretation of T and U. For that purpose, $\vdash$ in which T-ranks and the derivation lengths are omitted is firstly formalised via the operator $\Phi(X, \ulcorner \Gamma \urcorner)$ defined below, where Γ is (the code of) a sequent in the sense of the previous subsection. Since expressions of $\mathcal{L}_{FS} \cup \{\mathrm{L}_x(y)\}$ are coded as natural numbers, Φ can be given as a formula of $\mathcal{L}^*$:

The Δ_0-formula $\Phi_0(X, \Gamma)$ is defined to be the disjunction of the following:

- $P \in \Gamma$, for an $\mathcal{L}$-literal P true in $\mathcal{CTT}$,
- $A \wedge B \in \Gamma$, for some A, B such that $\Gamma, A \in X$ and $\Gamma, B \in X$;
- $A \vee B \in \Gamma$, for some A, B such that either $\Gamma, A \in X$ or $\Gamma, B \in X$;
- $\forall x A(x) \in \Gamma$, for some $\forall x A(x)$ such that $\Gamma, A(a) \in X$ for all closed terms a;
- $\exists x A(x) \in \Gamma$, for some $\exists x A(x)$ such that $\Gamma, A(a) \in X$ for some closed term a;

Similarly, the Δ_0-formula $\Phi_1(X, \Gamma)$ is defined to be the disjunction of the following:

- $\mathrm{T}(\dot{A}) \in \Gamma$, for some $A \in X$;
- $\neg\mathrm{T}(\dot{A}) \in \Gamma$, for some $\neg A \in X$;

Finally, the Δ_0-formula $\Phi_2(X, \Gamma)$ is defined to be the disjunction of the following:

- $\mathrm{L}_a(b) \in \Gamma$, for some a, b such that $b \in X$ and $(\star)$ holds;
- $\neg\mathrm{L}_a(b) \in \Gamma$, for some a, b such that $b \notin X$ and $(\star)$ holds,

where the condition $(\star)$ consists of the following:

1. $\forall c.\ ac \in X$ or $\dot{\neg}(ac) \in X$;
2. $\mathrm{L}_a(b) \notin X$ and $\neg\mathrm{L}_a(b) \notin X$.

Thus, the operator $\Phi(X, \Gamma, \alpha)$ roughly means that X contains the premise of one of the rules of $\vdash$. Next, we want to characterise the set of derivable sequents, i.e., the set $\{\Gamma \ : \vdash \Gamma\}$. Let $\mathrm{Fun}(f)$ be a Δ-predicate meaning that f is a function; let a unary Δ-predicate $\mathrm{On}(x)$ express that x is an ordinal number; we use $\alpha, \beta, \gamma, \alpha_0, \beta_0, \gamma_0, \ldots$ as variables ranging over On; a Σ-operation $\mathrm{Dom}(f)$ denotes the domain of f. Then, we define a Δ_0-predicate $\mathscr{H}(s, f)$, which roughly means that for each $\alpha, \beta, \gamma \in s$, the value of f at (α, β, γ) is the set $\{\Gamma \ : \vdash^{\alpha,\beta,\gamma} \Gamma\}$.

$$\mathscr{H}(s, f) := \mathrm{Ad}(s) \wedge \mathrm{Fun}(f) \wedge \forall \alpha, \beta, \gamma \in s.\ f(\alpha, \beta, \gamma) = S,$$

where the set S is the union of the following:

1. $f(<\alpha, \in s, \in s) \cup f(\alpha, \leq \beta, \leq \gamma)$,
2. $\{\Gamma \ : \ \Phi_0(f(\alpha,\beta,<\gamma),\Gamma) \vee \Phi_1(f(\alpha,<\beta,<\gamma),\Gamma)\}$,
3. $\{\Gamma \ : \ \alpha \in \mathrm{Suc} \wedge \Phi_2(f(\alpha-1,\in s,\in s),\Gamma)\}$.

Here, we used the following notations:

- $f(<\alpha,\in s,\in s) := \bigcup_{\alpha_0<\alpha,\beta\in s,\gamma\in s} f(\alpha_0,\beta,\gamma)$;
- $f(\alpha,\leq\beta,\leq\gamma) := \bigcup_{\beta_0\leq\beta,\gamma_0\leq\gamma} f(\alpha,\beta_0,\gamma_0)$;
- $f(\alpha,\beta,<\gamma)$ and $f(\alpha,<\beta,<\gamma)$ are similarly defined;
- $Cl_{\Phi_i}(X) :\leftrightarrow \forall x.\ \Phi_i(X,x) \rightarrow x \in X$, for $i \in \{0,1\}$.

The next lemma shows that $\mathscr{H}(s,f)$ determines the set $f(\alpha,\beta,\gamma)$ for each $\alpha,\beta,\gamma \in s$ regardless of the particular choice of s and f.

Lemma 189. (in KPi) Assume $\mathscr{H}(s,f)$ and $\mathscr{H}(s',g)$. Then the following hold for all $\alpha,\beta,\gamma \in s \cap s'$:

1. $Cl_{\Phi_0}(f(\alpha,\in s,\in s))$ and $Cl_{\Phi_1}(f(\alpha,\in s,\in s))$,
2. $f(\alpha,\in s,\in s) = g(\alpha,\in s',\in s')$.
3. $f(\alpha,\beta,\gamma) = g(\alpha,\beta,\gamma)$,

Proof. For the item 1, we show $Cl_{\Phi_0}(f(\alpha,\in s,\in s))$. Thus, taking any sequent Γ and assuming $\Phi_0(f(\alpha,\in s,\in s),\Gamma)$, we show $\Gamma \in f(\alpha,\in s,\in s)$. The proof is divided into cases according to the clauses of Φ_0. As the crucial case, suppose that a sentence $\forall x.\ A(x)$ is contained in Γ and $\forall a \in \mathrm{Term}.\ \Gamma, A(a) \in f(\alpha,\in s,\in s)$. Since s is admissible, by Σ-reflection within s, we can take an ordinal $\delta \in s$ such that $\forall a \in \mathrm{Term}.\ \Gamma, A(a) \in f(\alpha,<\delta,<\delta)$. Thus, by applying Φ_0 we obtain $\Gamma \in f(\alpha,<\delta,\delta)$, and hence $\Gamma \in f(\alpha,\in s,\in s)$. The other cases are similar. Moreover, $Cl_{\Phi_1}(f(\alpha,\in s,\in s))$ is similarly proved.

As to item 2: $f(\alpha,\in s,\in s) = g(\alpha,\in s',\in s')$, we show $f(\alpha,\beta,\gamma) \subseteq g(\alpha,\in s',\in s')$ by induction on α, β and γ. Therefore, assuming $\Gamma \in f(\alpha,\beta,\gamma)$, we have to show $\Gamma \in g(\alpha,\in s',\in s')$. By $\mathscr{H}(s,f)$, the proof is divided by cases according to the construction of $f(\alpha,\beta,\gamma)$. If $\Gamma \in f(<\alpha,\in s,\in s)$ or $\Gamma \in f(\alpha,\leq\beta,\leq\gamma)$, then the claim is obvious by the side-induction hypothesis. If $\Phi_0(f(\alpha,\beta,<\gamma),\Gamma)$, then since Φ_0 is positive, we have $\Phi_0(g(\alpha,\in s',\in s'),\Gamma)$ by the side-induction hypothesis. As $g(\alpha,\in s',\in s')$ is Φ_0-closed by the item 1, it follows that $\Gamma \in g(\alpha,\in s',\in s')$, as required. The case $\Phi_1(f(\alpha,\beta,<\gamma),\Gamma)$ is similar. The last case is where $\alpha \in \mathrm{Suc}$ and $\Phi_2(f(\alpha-1,\in s,\in s),\Gamma)$. Then, the main-induction hypothesis yields $\Phi_2(g(\alpha-1,\in s',\in s'),\Gamma)$, thus $\Gamma \in g(\alpha,\in s',\in s')$. In summary, we have $f(\alpha,\in s,\in s) \subseteq g(\alpha,\in s',\in s')$. The converse direction is similar, and thus the proof of item 3 is complete.

Item 3: $f(\alpha,\beta,\gamma) = g(\alpha,\beta,\gamma)$ is easily proved by induction on α, β and γ, with the help of item 2. ■

We further introduce the following notations:

- $I^{\alpha,\beta,\gamma}(x) := \exists s,f.\ \{\alpha,\beta,\gamma\} \subseteq s \wedge \mathscr{H}(s,f) \wedge x \in f(\alpha,\beta,\gamma)$,
- $I^{\alpha,\leq\beta,<\gamma}(x) := \exists \beta_0 \leq \beta.\ \exists \gamma_0 < \gamma.\ I^{\alpha,\beta_0,\gamma_0}(x)$,
- $I^{\alpha}(x) := \exists \beta,\gamma.\ I^{\alpha,\beta,\gamma}(x)$,
- $I^{<\alpha}(x)$, $I^{\alpha,\leq\beta,\leq\gamma}(x)$, $I^{\alpha,\beta,<\gamma}(x)$ and $I^{\alpha,<\beta,<\gamma}(x)$ are similarly defined.

We show that the Σ-predicate $I^{\alpha,\beta,\gamma}(x)$ expresses the desired class $\{\mathit{\Gamma}\ :\ \stackrel{\alpha,\beta,\gamma}{\vdash} \mathit{\Gamma}\}$.

Lemma 190. The following are derivable in KPi.

1. $\forall \alpha.\ \exists s,f.\ \alpha \in s \wedge \mathscr{H}(s,f)$.
2. $I^{\alpha,\beta,\gamma}(\mathit{\Gamma})$ if and only if one of the following holds:
   (a) $I^{<\alpha}(\mathit{\Gamma}) \vee I^{\alpha,\leq\beta,\leq\gamma}(\mathit{\Gamma})$,
   (b) $\Phi_0(I^{\alpha,\beta,<\gamma},\mathit{\Gamma}) \vee \Phi_1(I^{\alpha,<\beta,<\gamma},\mathit{\Gamma})$,
   (c) $\alpha \in \mathrm{Suc} \wedge \Phi_2(I^{\alpha-1},\mathit{\Gamma})$.

Proof.

1. By the limit axiom of KPi, let s be an admissible set that contains α. Then, from the definition of the Δ_0-predicate $\mathscr{H}(s,f)$, we can construct a required function f by Δ-recursion, available in KPi (cf. [54, p. 256]).
2. For the left-to-right direction, we assume $I^{\alpha,\beta,\gamma}(\mathit{\Gamma})$; thus, we take sets s and f such that $\{\alpha,\beta,\gamma\} \subseteq s \wedge \mathscr{H}(s,f) \wedge \mathit{\Gamma} \in f(\alpha,\beta,\gamma)$. For instance, we consider the case where $\mathit{\Gamma} \in f(\alpha,\beta,\gamma)$ is obtained from Φ_2, then we have $\alpha \in \mathrm{suc}$ and $\Phi_2(f(\alpha-1,\in s,\in s),\mathit{\Gamma})$. By Lemma 189, we can easily show that $\forall x.\ x \in f(\alpha-1,\in s,\in s) \leftrightarrow I^{\alpha-1}(x)$. Therefore, we get $\Phi_2(I^{\alpha-1},\mathit{\Gamma})$, as required. The other cases are similarly proved by using Lemma 189.
   As for the converse direction, we, for example, assume $\alpha \in \mathrm{Suc} \wedge \Phi_2(I^{\alpha-1},\mathit{\Gamma})$. By item 1, we take sets s and f such that $\max(\alpha,\beta,\gamma) \in s$ and $\mathscr{H}(s,f)$. Then, again by Lemma 189, we have $\forall x.\ I^{\alpha-1}(x) \leftrightarrow x \in f(\alpha-1,\in s,\in s)$. Therefore, we obtain $\Phi_2(f(\alpha-1,\in s,\in s),\mathit{\Gamma})$, and hence it follows that $\mathit{\Gamma} \in f(\alpha,\beta,\gamma)$. Thus, we obtain $I^{\alpha,\beta,\gamma}(\mathit{\Gamma})$. The other cases are similarly proved.

■

Let $I^{\infty}(x) :\leftrightarrow \exists \alpha.\ I^{\alpha}(x)$. We now define an interpretation $^{+}$ of VFU into KPi. The vocabularies of $\mathcal{L}$ are interpreted in exactly the same way as in Lemma 154. Then, we let $\mathrm{T}^{+}(x) :\equiv I^{\infty}(x)$; let $\mathrm{U}^{+}(x) :\leftrightarrow \exists a.\ x = \mathsf{I}a \wedge \forall b.\ I^{\infty}(ab) \vee I^{\infty}(\dot{\neg}(ab))$.

Lemma 191. For each $\mathcal{L}_{FS}$-formula A, if $\mathsf{VFU} \vdash A$, then $\mathsf{KPi} \vdash A^+$.

Proof. The proof is by directly running the proof of Theorem 186 within KPi.

(U-Class) We want to show $\mathsf{KPi} \vdash (\mathrm{U}(u) \to \mathrm{C}(u))^+$. Thus, taking any term f such that $u = \mathsf{I}f$ and $\forall a.\ I^\infty(fa) \vee I^\infty(\dot\neg(fa))$, we show $\forall a.\ I^\infty(\mathsf{I}fa) \vee I^\infty(\dot\neg(\mathsf{I}fa))$. Then, since I^∞ is a Σ-predicate, we can, by Σ-reflection, take the least successor ordinal α such that $\forall a.\ I^{<\alpha}(fa) \vee I^{<\alpha}(\dot\neg(fa))$. Therefore, Lemma 190 implies that $\forall a.\ I^\alpha(\mathsf{I}fa) \vee I^\alpha(\dot\neg(\mathsf{I}fa))$, and thus we obtain $\forall a.\ I^\infty(\mathsf{I}fa) \vee I^\infty(\dot\neg(\mathsf{I}fa))$.

(U-True) Similar to the above.

(U-Out) Since $\mathsf{KPi} \vdash (\mathsf{U\text{-}True})^+$, it suffices to verify $(\mathsf{T\text{-}Out})^+$, that is, $\mathsf{KPi} \vdash I^\infty(\dot A) \to A^+$ for each $\mathcal{L}_{FS}$-formula A. For that purpose, we formalise Lemma 185 within KPi, similarly to [12, Lemma 5.8.2]. In particular, we can show the following for each natural number k:

$$\mathsf{KPi} \vdash \forall \Gamma \in \mathrm{Seq}^k.\ I^\infty(\Gamma) \to \exists x \in \Gamma.\ \mathrm{T}_k(x),$$

where $\Gamma \in \mathrm{Seq}^k$ means every sentence in Γ has the logical complexity $\leq k$; the predicate $\mathrm{T}_k(x)$ is a partial truth predicate such that $\mathsf{KPi} \vdash \mathrm{T}_k(\dot A) \leftrightarrow A^+$ for each $\mathcal{L}_{FS}$-formula A with the logical complexity $\leq k$. Then, we have $\mathsf{KPi} \vdash I^\infty(\dot A) \to A^+$ for each $\mathcal{L}_{FS}$-formula A, as required.

The other cases are similarly proved by using Lemma 190. ∎

Combining Theorem 180 with Lemma 191, we obtain the proof-theoretic strength of VFU:

Theorem 192. VFU and T_0 are proof-theoretically equivalent.

Chapter 6
Conclusion

In Chapter 3, we modify Leigh and Rathjen's ordinal analysis for the Friedman–Sheard systems $\mathsf{E}, \mathsf{F}, \mathsf{G}$, and I. As a result, we show that truth-as-recursive-provability interpretation can be given for those systems, and we verify Leigh and Rathjen's results. In addition, we observe that a similar technique can be applied to the analysis for FS.

Theorem 34 (cf. [50, Corollary 3.27]) $\mathsf{G} \leq \mathsf{PA} + \mathsf{TI}(< \varphi 20)$.

Theorem 48 (cf. [50, Corollary 3.35]) $\mathsf{I} \leq \mathsf{PA} + \mathsf{TI}(< \varphi 20)$.

Theorem 63 (cf. [50, Theorem 3.15]) $\mathsf{F} \leq \mathsf{PA} + \mathsf{TI}(< \varphi \omega 0)$.

Theorem 73 (cf. [50, Theorem 3.45]) $\mathsf{E} \leq \mathsf{PA} + \mathsf{TI}(< \varphi \omega 0)$.

Theorem 83 ([28]) $\mathsf{FS} \leq \mathsf{PA} + \mathsf{TI}(< \varphi 20)$.

It is not clear whether this truth-as-recursive-provability can be used to analyse other theories, such as KF.

In Chapter 4, we formulate a formal theory of Krivine's classical realisability. We first define a first-order theory CR of compositional realisability (Definition 90). Next, we observe that our classical number realisability can realise every theorem of PA, which is formalisable in CR (Theorem 96). As for proof-theoretic strength, CR is shown to be conservative over PA (Proposition 97). On the other hand, we formulate a kind of reflection principle under which CR essentially amounts to CT (Proposition 100). We also consider a weaker reflection principle that is sound with respect to any pole, and we prove that the principle can make CR as strong as CT (Theorem 103). We further generalise CR to a Friedman–Sheard-style system FSR for realisability (Definition 115). Then, we prove that FSR has a nice property, *self-realisability* (Theorem 125). We also observe that FSR is conservative (Proposition 126). Finally, the proof-theoretic strength of GCR^+ and GCR^{++} remains open (Conjecture 128).

Interestingly, our theories CR and FSR are both conservative over PA. Perhaps, this fact may be useful for arguing that although CT and FS are not conservative, they still have the spirit of deflationism. Also, our realisability theories might help to construct verificationism of classical logic.

In Chapter 5, we formulate five systems of Frege structures. The results are summarised as follows: all the following theories are proof-theoretically equivalent to T_0:

- KFUPI (Theorem 155) and PTUPI (Theorem 162),
- KFLU (Theorem 168) and PTLU (Theorem 170),
- VFU (Theorem 192).

The author suggests two directions for future studies. First, given that most of the truth theories have been studied over Peano arithmetic, it would be desirable to find systems over Peano arithmetic that corrrespond to our theories. Second, we can consider extending our systems further by stronger universe-generating axioms. One example by Cantini [14] is the Mahlo principle, which is an analogue of the recursively Mahlo axiom in Kripke–Platek set theory. Therefore, the question is how strong the systems of Frege structure will be with the addition of such a principle. Furthermore, Jäger and Strahm [39] formulated an even stronger principle in explicit mathematics; thus, it may be possible to give its counterpart in the framework of Frege structure. Of course, philosophical discussions would also be required on how well these are motivated as truth-theoretic principles.

Bibliography

[1] Peter Aczel. Frege structures and the notions of proposition, truth and set. In Jon Barwise, H. Jerome Keisler, and Kenneth Kunen, editors, *The Kleene Symposium*, volume 101 of *Studies in Logic and the Foundations of Mathematics*, pages 31–59. Elsevier, 1980.

[2] Federico Aschieri. *Learning, realizability and games in classical arithmetic*. PhD thesis, Queen Mary University of London, 2011.

[3] Federico Aschieri. Interactive realizability for classical Peano arithmetic with Skolem axioms. In Patrick Cégielski and Arnaud Durand, editors, *Computer Science Logic 2012*, volume 16, pages 31–45. 2012.

[4] Jeremy Avigad. A realizability interpretation for classical arithmetic. In *Logic Colloquium*, volume 98, pages 57–90, 2000.

[5] Jon Barwise. *Admissible sets and structures, perspectives in mathematical logic*. Springer Berlin, 1975.

[6] Michael J Beeson. *Foundations of constructive mathematics: Metamathematical studies*. Springer-Verlag, Berlin, 1985.

[7] Ulrich Berger and Paulo Oliva. Modified bar recursion and classical dependent choice. In *Logic Colloquium*, volume 1, pages 89–107, 2005.

[8] Valentin Blot. Typed realizability for first-order classical analysis. *Logical Methods in Computer Science*, 11:1–43, 2015.

[9] Valentin Blot. Realizability for Peano arithmetic with winning conditions in HON games. *Annals of Pure and Applied Logic*, 168(2):254–277, 2017.

[10] John P Burgess. Friedman and the axiomatization of Kripke's theory of truth. In Neil Tennant, editor, *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, pages 125–148. College Publications, 2009.

[11] Andrea Cantini. Notes on formal theories of truth. *Mathematical Logic Quarterly*, 35(2):97–130, 1989.

[12] Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic*, 55(1):244–259, 1990.

[13] Andrea Cantini. Extending the first-order theory of combinators with self-referential truth. *The Journal of Symbolic Logic*, 58(2):477–513, 1993.

[14] Andrea Cantini. *Logical frameworks for truth and abstraction: An axiomatic study*, volume 135 of *Studies in logic and the foundations of mathematics*. Elsevier, Amsterdam, 1996.

[15] Andrea Cantini. About truth and types. In Reinhard Kahle, Thomas Strahm, and Thomas Studer, editors, *Advances in proof theory*, pages 31–64. Springer, 2016.

[16] Michael Dummett. *Truth and other enigmas*. Harvard University Press, 1978.

[17] Solomon Feferman. A language and axioms for explicit mathematics. In John Newsome Crossley, editor, *Algebra and Logic*, pages 87–139, Berlin, Heidelberg, 1975. Springer.

[18] Solomon Feferman. Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56(1):1–49, 1991.

[19] Solomon Feferman. Axioms for determinateness and truth. *The Review of Symbolic Logic*, 1(2):204–217, 2008.

[20] Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.

[21] Harvey Friedman and Michael Sheard. The disjunction and existence properties for axiomatic systems of truth. *Annals of Pure and Applied Logic*, 40(1):1–10, 1988.

[22] Kentaro Fujimoto. Relative truth definability of axiomatic truth theories. *Bulletin of Symbolic Logic*, 16(3):305–344, 2010.

[23] Kentaro Fujimoto. Autonomous progression and transfinite iteration of self-applicable truth. *The Journal of Symbolic Logic*, 76(3):914–945, 2011.

[24] Kentaro Fujimoto and Volker Halbach. Classical determinate truth I. *The Journal of Symbolic Logic*, 89(1):218–261, 2024.

[25] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493–565, 1936.

[26] Timothy G Griffin. A formulae-as-type notion of control. In *the 17th ACM SIGPLAN-SIGACT symposium on Principles of programming languages*, pages 47–58, 1989.

[27] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic*, volume 3. Springer, Berlin, 1993.

[28] Volker Halbach. A system of complete and consistent truth. *Notre Dame Journal of Formal Logic*, 35(3):311–327, 1994.

[29] Volker Halbach. Truth and reduction. *Erkenntnis*, 53(1-2):97–126, 2000.

[30] Volker Halbach. *Axiomatic theories of truth.* Cambridge University Press, 2014.

[31] Daichi Hayashi. On cut-elimination arguments for axiomatic theories of truth. *Studia Logica*, 110(3):785–818, 2022.

[32] Daichi Hayashi. Theories of Frege structure equivalent to Feferman's system T_0. *Annals of Pure and Applied Logic*, 2024. https://doi.org/10.1016/j.apal.2024.103510.

[33] Daichi Hayashi and Graham E. Leigh. A compositional theory of classical realisability. In Metcalfe George, Studer Thomas, and de Queiroz Ruy, editors, *Logic, Language, Information, and Computation - 30th International Workshop, WoLLIC 2024*, volume 14672 of *Lecture Notes in Computer Science*, pages 64–79, 2024.

[34] Susumu Hayashi and Satoshi Kobayashi. A new formalization of Feferman's system of functions and classes and its relation to Frege structure. *International Journal of Foundations of Computer Science*, 6(03):187–202, 1995.

[35] Gerhard Jäger. A well-ordering proof for Feferman's theory T_0. *Archiv für mathematische Logik und Grundlagenforschung*, 23(1):65–77, 1983.

[36] Gerhard Jäger. First order theories for nonmonotone inductive definitions: recursively inaccessible and Mahlo. *The Journal of Symbolic Logic*, 66(3):1073–1089, 2001.

[37] Gerhard Jäger, Reinhard Kahle, Anton Setzer, and Thomas Strahm. The proof-theoretic analysis of transfinitely iterated fixed point theories. *The Journal of Symbolic Logic*, 64(1):53–67, 1999.

[38] Gerhard Jäger, Reinhard Kahle, and Thomas Studer. Universes in explicit mathematics. *Annals of Pure and Applied Logic*, 109(3):141–162, 2001.

[39] Gerhard Jäger and Thomas Strahm. Reflections on reflections in explicit mathematics. *Annals of Pure and Applied Logic*, 136(1-2):116–133, 2005.

[40] Reinhard Kahle. Truth in applicative theories. *Studia Logica*, 68(1):103–128, 2001.

[41] Reinhard Kahle. Universes over Frege structures. *Annals of Pure and Applied Logic*, 119(1-3):191–223, 2003.

[42] Reinhard Kahle. Sets, truth, and recursion. In Theodora Achourioti, Henri Galinon, José Martínez Fernández, and Kentaro Fujimoto, editors, *Unifying the Philosophy of Truth*, pages 143–152. Springer, 2015.

[43] Reinhard Kahle and Thomas Studer. A theory of explicit mathematics equivalent to ID_1. In *International Workshop on Computer Science Logic*, pages 356–370. Springer, 2000.

[44] Sato Kentaro. A new model construction by making a detour via intuitionistic theories II: Interpretability lower bound of Feferman's explicit mathematics T_0. *Annals of Pure and Applied Logic*, 166(7-8):800–835, 2015.

[45] Stephen Cole Kleene. On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic*, 10(4):109–124, 1945.

[46] Saul Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72(19):690–716, 1976.

[47] Jean-Louis Krivine. Typed lambda-calculus in classical Zermelo-Fraenkel set theory. *Archive for Mathematical Logic*, 40(3):189–205, 2001.

[48] Jean-Louis Krivine. Dependent choice, 'quote' and the clock. *Theoretical Computer Science*, 308(1-3):259–276, 2003.

[49] Jean-Louis Krivine. Realizability in classical logic. *Panoramas et synthèses*, 27:197–229, 2009.

[50] Graham Emil Leigh and Michael Rathjen. An ordinal analysis for theories of self-referential truth. *Archive for Mathematical Logic*, 49(2):213–247, 2010.

[51] Graham E. Leigh. Reflecting on truth. *IfCoLog Journal of Logics and their Applications*, 3(4):557–594, 2016.

[52] Paulo Oliva and Thomas Streicher. On Krivine's realizability interpretation of classical second-order arithmetic. *Fundamenta Informaticae*, 84(2):207–220, 2008.

[53] Wolfram Pohlers. Subsystems of set theory and second order number theory. In Samuel R. Buss, editor, *Handbook of proof theory*, volume 137, pages 209–335. 1998.

[54] Wolfram Pohlers. *Proof theory: The first step into impredicativity*. Springer Science & Business Media, 2008.

[55] Bertrand Russell. Mathematical logic as based on the theory of types. *American Journal of Mathematics*, 30(3):222–262, 1908.

[56] Thomas Schindler. A Disquotational theory of truth as strong as Z_2^-. *Journal of Philosophical Logic*, 44:395–410, 2015.

[57] Helmut Schwichtenberg. Proof theory: Some applications of cut-elimination. In *Studies in Logic and the Foundations of Mathematics*, volume 90, pages 867–895. Elsevier, 1977.

[58] Stewart Shapiro. Proof and truth: Through thick and thin. *The Journal of Philosophy*, 95(10):493–521, 1998.

[59] Johannes Stern. Supervaluation-style truth without supervaluations. *Journal of Philosophical Logic*, 47:817–850, 2018.

[60] Thomas Strahm. First steps into metapredicativity in explicit mathematics. In S. Barry Cooper and John K. Truss, editors, *Sets and Proofs*, London Mathematical Society Lecture Note Series, pages 383–402. Cambridge University Press, 1999.

[61] Alfred Tarski. The concept of truth in formalized languages. In A. Tarski, editor, *Logic, Semantics, Metamathematics*, pages 152–278. Hackett, Indianapolis, 2nd edition, 1983.

[62] A. Troelstra and D. van Dalen. *Constructivism in mathematics: an introduction*, volume I. Elsevier, Amsterdam, 1988.

[63] Sergei Tupailo. Realization of analysis into explicit mathematics. *The Journal of Symbolic Logic*, 66(4):1848–1864, 2001.

[64] Sergei Tupailo. Realization of constructive set theory into explicit mathematics: a lower bound for impredicative Mahlo universe. *Annals of Pure and Applied Logic*, 120(1-3):165–196, 2003.

[65] Bas C van Fraassen. Singular terms, truth-value gaps, and free logic. *The Journal of Philosophy*, 63(17):481–495, 1966.