Theories of Frege structure equivalent to Feferman's system T_0

Daichi Hayashi

Abstract

Feferman [9] defines an impredicative system T_0 of explicit mathematics, which is proof-theoretically equivalent to the subsystem Δ^1_2-CA + BI of second-order arithmetic. In this paper, we propose several systems of Frege structure with the same proof-theoretic strength as T_0. To be precise, we first consider the Kripke–Feferman theory, which is one of the most famous truth theories, and we extend it by two kinds of induction principles inspired by [21]. In addition, we give similar results for the system based on Aczel's original Frege structure [1]. Finally, we equip Cantini's supervaluation-style theory with the notion of universes, the strength of which was an open problem in [23].

Keywords: Frege structure, explicit mathematics, proof-theoretic strength, cut-elimination, Kripke-style truth

Contents

1. Introduction

Aczel [1] introduced the framework of *Frege structure*, which is essentially a model of lambda calculus augmented with the notions of *truth* and *proposition*, to analyse Russell's paradox in Frege's Grundgesetze. 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 2). 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., [7]). As another applicative framework, Feferman formulated systems of *explicit mathematics* [9], 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 1, where EM is a system of explicit mathematics (see Definition 2.3); EMU is an extension of EM by *universes* (cf. [33]); NEM is an extension of EM by *name induction* (cf. [26]); T_0 is an extension of EM by *inductive generations* (see Definition 2.5); KF and PT are theories of Frege structure based on Kripke-Feferman logic and Aczel-Feferman logic, respectively (see Definition 3.1 and 5.1); KFU and PTU are respectively extensions of KF and PT by *universes* (see Definition 3.3 and 5.4); and VF is a theory of Frege structure based on supervaluation logic (see Definition 7.1). 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; ATR + (Σ^1_1-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-CA + BI is the Δ^1_2 comprehension schema with the bar induction. For details of their proof-theoretic ordinals, see, e.g., [20, 29]).

Table 1: Applicative theories

Ordinal strength	Explicit mathematics	Frege structure	Second-order arithmetic
$\psi_\Omega(\varepsilon_{I+1})$	T_0	?	Δ^1_2-CA + BI
$\psi_\Omega(\varepsilon_{\Omega+1})$	NEM	VF	Π^1_1-CA^-_0
$\varphi 1 \varepsilon_0 0$	EMU	KFU, PTU	ATR + (Σ^1_1-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 (Π^1_1-CA)$^-_0$. Therefore, this paper aims to provide well-motivated theories of Frege structure proof-theoretically equivalent to T_0.

This task also has importance from the foundational viewpoint if Frege structure is seen as a theory of truth. In axiomatic theory of truth, it has been one of the central tasks to obtain stronger truth theories. For one thing, Halbach [15] 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. As far as the author knows, Fujimoto's system $\mathsf{Aut(VF)}$ [14], formulated over Peano arithmetic (PA), is so far the strongest among well-motivated truth theories.[1] Since the strength of $\mathsf{Aut(VF)}$ lies strictly between Δ^1_2-CA and Δ^1_2-CA + BI, the author believes that the theories proposed in this paper break the record, though our base theory is not PA. Moreover, since T_0 is expressively rich enough to interpret various set theories (cf. [27, 34, 35]), we can expect our theories of Frege structure to contribute to Halbach's programme.

The structure of this paper is as follows: In Section 2, 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 3, following Kahle's formulation [23, 24], we introduce Kripke-Feferman-style theories of Frege structure, KF and its extension KFU. In Section 4, we expand KFU by the *proposition induction* schema, to obtain a theory KFUPI that is as strong as T_0. In Section 5, 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.

[1] However, the author also remarks that Cantini gave a recursion-theoretically motivated system of Frege structure which is at least as strong as T_0 [7, p. 253]. Over Peano arithmetic, Schindler [31] formulated a disquotational truth theory of the same consistency strength as the full second-order arithmetic.

In Section 6, 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 7, we introduce a supervaluation-style Frege structure essentially based on Kahle's formulation [23]. Then, following Kahle's suggestion [23, 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.

2. Technical preliminaries

This section defines Feferman's system T_0 of explicit mathematics. We first define the total applicative theory TON, which, in this paper, is used as the common base theory of explicit mathematics and Frege structure.

2.1. Total applicative theory

The first-order language $\mathcal{L}$ for the total operational theory (TON) (see, e.g. [24]) consists of the standard logical symbols, individual variables ($x, y, z, x_0, x_1, \ldots, a, b, c, \ldots, f, g, h$), 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 $\neg A$, $A \land B$, $A \to B$, and $\forall x.\ A$. We assume that $\lor$ 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 paper:

1. $ab := (ab) := \mathsf{App}(a, b)$.
2. $a_1a_2a_3 \ldots a_n :\equiv ((\cdots((a_1a_2)a_3\cdots)a_n)$.
3. $(x, y) := \mathsf{p}xy$.
4. $(x)_i := \mathsf{p}_ix$ for $i \in \{0, 1\}$.
5. $s \neq t := \neg(s = t)$.
6. $\forall x_0, x_1, \ldots x_n.\ A :\equiv \forall x_0.\forall x_1.\ldots\forall x_n.\ A$.

Definition 2.1. *The $\mathcal{L}$-theory* TON *consists of the following axioms:*

- $\mathsf{k}ab = a$,
- $\mathsf{s}abc = ac(bc)$,
- $(a, b)_0 = a \land (a, b)_1 = b$,
- $\mathrm{N}(\mathsf{0}) \land \forall x.\ \mathrm{N}(x) \to \mathrm{N}(\mathsf{s_N}x)$,
- $\forall x.\ \mathrm{N}(x) \to \mathsf{s_N}x \neq 0 \land \mathsf{p_N}(\mathsf{s_N}x) = x$,
- $\forall x.\ \mathrm{N}(x) \land x \neq \mathsf{0} \to \mathrm{N}(\mathsf{p_N}x) \land \mathsf{s_N}(\mathsf{p_N}x) = x$,
- $\mathrm{N}(a) \land \mathrm{N}(b) \land a = b \to \mathsf{d_N}uvab = u$,
- $\mathrm{N}(a) \land \mathrm{N}(b) \land a \neq b \to \mathsf{d_N}uvab = v$,
- $A(\mathsf{0}) \land [\forall x.\ \mathrm{N}(x) \land A(x) \to A(\mathsf{s_N}x)] \to \forall x.\ \mathrm{N}(x) \to A(x)$, *for every $\mathcal{L}$-formula A.*

In this paper, we will repeatedly use the following well-known fact :

Fact 2.2 (cf. [7]).

λ-abstraction. *For each variable x and an $\mathcal{L}$-term t, we can find a term $\lambda x.t$ such that* TON $\vdash (\lambda x.t)x = t$.

Recursion. *There is a term* rec *such that* TON $\vdash \forall f.\ \mathsf{rec}f = f(\mathsf{rec}f)$.

2.2. *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 [17].

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), dom (domain), inv (inversion), and i (inductive generations).

Definition 2.3. *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. $\mathrm{R}(x) \land [\forall y \in x.\ \mathrm{R}(fy)] \to \mathrm{R}(\mathsf{j}(x, f)) \land \Sigma(x, f, \mathsf{j}(x, f))$, *where*

$$\Sigma(x, f, y) :\equiv \forall u.\ u \in y \leftrightarrow \exists v, w.\ u = (v, w) \land v \in x \land w \in fv.$$

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 2.4 (cf. [3]). EM *is proof-theoretically equivalent to* Σ^1_1-AC.

2.3. *Universes and inductive generations*

A *universe* in explicit mathematics is a set that is closed under the name-generating operations in Definition 2.3. More formally, the fact that a is a universe is expressed by the formula:

$$\mathrm{U}(a) :\equiv [\forall b.\ C(a, b) \to b \in a] \to \forall b \in a.\ \mathrm{R}(b),$$

where $C(a, b)$ is the disjunction of the following:

1. $a = \mathsf{nat} \lor a = \mathsf{id}$,
2. $\exists x.\ b = \mathsf{co}(x) \land x \in a$,
3. $\exists x, y.\ b = \mathsf{int}(x, y) \land x \in a \land y \in a$,
4. $\exists x.\ b = \mathsf{dom}(x) \land x \in a$,
5. $\exists f, x.\ b = \mathsf{inv}(x, f) \land x \in a$,
6. $\exists f, x.\ b = \mathsf{join}(x, f) \land x \in a \land \forall y \in x. fy \in a$.

Of course, we require an additional axiom to assure the existence of universes. Here, we introduce the *limit axiom* [33]. For a new constant symbol l, the limit axiom is the following:

$$\mathrm{R}(x) \to \mathrm{R}(\mathsf{l}x) \land x \in \mathsf{l}x.$$

Then, the system EMU is defined as EM equipped with the limit axiom. From Strahm's proof [33], 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 [30]). Therefore, the strength of EMU is called *metapredicative* [33].

Feferman [9] 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 2.5. *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 2.6 ([18]). T_0 *is proof-theoretically equivalent to* Δ^1_2-$\mathsf{CA} + \mathsf{BI}$.

3. 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.

3.1. *System* KF

The theory KF, as a theory of truth, was semantically introduced by Kripke [28]; Feferman [10] then gave its first-order axiomatisation. Cantini [6] later formulated KF as a theory of Frege structure. 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

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

The formal system KF, as an $\mathcal{L}_{FS}$-theory, is defined straightforwardly from the above table.

Definition 3.1 (cf. [7, 8, 23, 24]). *The $\mathcal{L}_{FS}$-theory* KF *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.

($\mathsf{T\text{-}Cons}$) $\neg[\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}x)]$

The system KF is proof-theoretically stronger than TON. In particular, KF can relatively interpret EM. To see this, we define a translation $'$: $\mathcal{L}_{EM} \to \mathcal{L}_{FS}$. 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 [7, 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 7.7). To summarise, the following is obtained:

Fact 3.2 ([7]). *For each $\mathcal{L}_{EM}$-formula A, if* EM $\vdash A$, *then* KF $\vdash A'$.

We also remark that EM and KF are proof-theoretically equivalent (for the proof, see, e.g., [7, Section 57]).

3.2. *Universes for* KF

In order to strengthen a given truth theory, iterating truth predicates is effective in many cases (cf. [10, 20, 14]). In the framework of Frege structure, an analogous idea is realisable by using the notion of *universes* (cf. [7, 24]). Following [24], 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 3.3 (cf. [23, 24]). *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}) $\mathrm{U}(u) \to \forall x.\, \mathrm{T}(u(\dot{\mathsf{N}}x)) \leftrightarrow \mathrm{N}(x)$

($\mathsf{U}_{\neg\mathrm{N}}$) $\mathrm{U}(u) \to \forall x.\, \mathrm{T}(u(\dot{\neg}(\dot{\mathsf{N}}x))) \leftrightarrow \neg\mathrm{N}(x)$

($\mathsf{U}_{\neg\neg}$) $\mathrm{U}(u) \to \forall x.\, \mathrm{T}(u(\dot{\neg}(\dot{\neg}x))) \leftrightarrow \mathrm{T}(ux)$

($\mathsf{U}_{\wedge}$) $\mathrm{U}(u) \to \forall x, y.\, \mathrm{T}(u(x \dot{\wedge} y)) \leftrightarrow \mathrm{T}(ux) \wedge \mathrm{T}(uy)$

($\mathsf{U}_{\neg\wedge}$) $\mathrm{U}(u) \to \forall x, y.\, \mathrm{T}(u(\dot{\neg}(x \dot{\wedge} y))) \leftrightarrow \mathrm{T}(u(\dot{\neg}x)) \vee \mathrm{T}(u(\dot{\neg}y))$

($\mathsf{U}_{\rightarrow}$) $\mathrm{U}(u) \to \forall x, y.\, \mathrm{T}(u(x \dot{\rightarrow} y)) \leftrightarrow \mathrm{T}(u(\dot{\neg}x)) \vee \mathrm{T}(uy)$

($\mathsf{U}_{\neg\to}$) $\mathrm{U}(u) \to \forall x, y.\ \mathrm{T}(u(\dot\neg(x\dot\to y))) \leftrightarrow \mathrm{T}(ux) \wedge \mathrm{T}(u(\dot\neg y))$

($\mathsf{U}_{\forall}$) $\mathrm{U}(u) \to \forall f.\ \mathrm{T}(u(\dot\forall f)) \leftrightarrow \forall x.\ \mathrm{T}(u(fx))$

($\mathsf{U}_{\neg\forall}$) $\mathrm{U}(u) \to \forall f.\ \mathrm{T}(u(\dot\neg(\dot\forall f))) \leftrightarrow \exists x.\ \mathrm{T}(u(\dot\neg(fx)))$

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 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 3.3, 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 showed that KFU is a proper extension of KF:

Fact 3.4 ([24, Proposition 14]). KF *is a subtheory of* KFU.

Fact 3.5 ([24, Theorem 33]). KFU *and* EMU *have the same* $\mathcal{L}$*-theorems.*

4. Extension by proposition induction

While adding universes makes KF stronger, KFU falls within metapredicativity in strength, so it is still much weaker than impredicative theories such as Π^1_1-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 [21]. 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 [21, 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.

4.1. 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 4.1. KF *derives the following:*

1. $\mathrm{P}(x\dot=y) \wedge \mathrm{P}(\dot{\mathrm{N}}x)$,
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 4.2. This inductive characterisation of propositions is nearly the same as that of *determinateness* in [16], where Halbach and Fujimoto advocate the definition by appealing to the function of truth as a generalising device [16, p. 5]. 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 4.3. *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 4.4. *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*).

4.2. 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 3.2. For each symbol except i, we can employ the same interpretation as given in the proof of Fact 3.2. 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 2.5:

(IG.1)$'$ $\mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathrm{C}(\mathsf{acc}(a,b)) \wedge \mathsf{Closed}'(a, b, \mathrm{T}(\mathsf{acc}(a,b)(\bullet)))$;

(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 ([21, 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

$$\mathsf{acc}(u,v)x = (\oplus(u,v)(0,x))\dot{\wedge}(\dot{\forall}y(\oplus(u,v)(0,y)\dot{\to}[\oplus(u,v)(1,(y,x))\dot{\to}(\mathsf{l}(\oplus(u,v))(t(u,v,y))])).$$

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 (IG.1)$'$ is derived.

Lemma 4.5. $\mathsf{KFU} \vdash \mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathrm{C}(\oplus(a,b)) \wedge \mathrm{C}(\mathsf{acc}(a,b))$.

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 4.1, 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 4.1. 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 4.1 implies that $\mathsf{acc}(a,b)x$ is a propositon, as required. □

The following is the second half of (IG.1)$'$.

Lemma 4.6. *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:

$$\mathrm{T}((\oplus(a,b)(0,d))\dot{\to}(\oplus(a,b)(1,(d,c))\dot{\to}(l(\oplus(a,b))t(a,b,d)))). \tag{1}$$

Then, the axioms of KF yield the desired conclusion.

Since the term $\oplus(a,b)$ is a class by Lemma 4.5 and is reflected by $l(\oplus(a,b))$, the axioms of KF imply that the formula (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 (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., [4, 8, 21, 26]. 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 4.8 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 4.8.

Lemma 4.7. 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 4.6, 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 4.8. *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 4.6, 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:

   $$\mathrm{T}(\oplus(u, v)(1, (z, x))) \to B_A(u, v, t(u, v, z)). \tag{2}$$

   By the assumption $\mathrm{T}(\mathsf{acc}(u, v)x)$ and Lemma 4.7, we have $\mathrm{T}(\mathsf{acc}(u, v)z)$. Therefore, (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 $B_A(u, v, w)$. $\square$

Using the above lemma, we can derive $(\mathsf{IG.2})'$.

Lemma 4.9. KFUPI *derives the following:*

$$\mathrm{C}(a) \wedge \mathrm{C}(b) \wedge Closed(a, b, A) \to \forall x.\ \mathrm{T}(\mathsf{acc}(a, b)x) \to A(x).$$

Proof. We assume $\mathrm{C}(a), \mathrm{C}(b), Closed(a, b, A)$ and $\mathrm{T}(\mathsf{acc}(a, b)c)$ for an arbitrary c. Then, we have to derive $A(c)$. From the assumptions and Lemma 4.8, 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:

$$\forall y.\ \mathrm{T}(\oplus(a, b)(0, y)) \to \mathrm{T}(\oplus(a, b)(1, (y, c))) \to \mathrm{T}(t(a, b, y)). \tag{3}$$

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 (3) follows from the assumption $\mathrm{T}(\mathsf{acc}(a, b)c)$. $\square$

To summarise, we obtain the interpretation of T_0 into KFUPI, in which $\mathcal{L}$-vocabularies are preserved.

Theorem 4.10. *For each $\mathcal{L}_{EM}$-sentence A, if* $\mathsf{T}_0 \vdash A$*, then* $\mathsf{KFUPI} \vdash A'$.

4.3. Kahle's model for KFU

In this subsection, we introduce Kahle's model construction for KFU [24, pp. 214-215] and observe that this model also satisfies KFUPI. The upper-bound proof of KFUPI in Section 4.4 is essentially based on this model.

First, the base theory TON is interpreted by the closed total term model $\mathcal{CTT}$ (e.g. [7, 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. $\exists b, c.\ \alpha = 0 \wedge a = (b \dot{=} c) \wedge b = c$,
3. $\exists b, c.\ \alpha = 0 \wedge a = \dot{\neg}(b \dot{=} c) \wedge b \neq c$,
4. $\exists b.\ \alpha = 0 \wedge a = \dot{\mathrm{N}}b \wedge \mathrm{N}(b)$,
5. $\exists b.\ \alpha = 0 \wedge a = \dot{\neg}(\dot{\mathrm{N}}b) \wedge \neg\mathrm{N}(b)$,
6. $\exists b.\ a = \dot{\neg}(\dot{\neg}b) \wedge a \notin Y \wedge b \in X$,
7. $\exists b, c.\ a = (b \dot{\wedge} c) \wedge a \notin Y \wedge b \in X \wedge c \in X$,
8. $\exists b, c.\ a = \dot{\neg}(b \dot{\wedge} c) \wedge a \notin Y \wedge [\dot{\neg}b \in X \vee \dot{\neg}c \in X]$,
9. $\exists b, c.\ a = (b \dot{\to} c) \wedge a \notin Y \wedge [\dot{\neg}b \in X \vee c \in X]$,
10. $\exists b, c.\ a = \dot{\neg}(b \dot{\to} c) \wedge a \notin Y \wedge b \in X \wedge \dot{\neg}c \in X$,
11. $\exists f.\ a = (\dot{\forall} f) \wedge a \notin Y \wedge \forall x.\ fx \in X$,
12. $\exists f.\ a = \dot{\neg}(\dot{\forall} f) \wedge a \notin Y \wedge \exists x.\ \dot{\neg}(fx) \in X$,
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_α := 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 4.11. $\mathcal{M}_{\mathsf{KF}} \models \mathsf{KFUPI}$.

Proof. Kahle [24, 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 4.12. As Kahle remarks [24, p. 215], the model $\mathcal{M}_{\mathsf{KF}}$ also satisfies the following principles:

(U-Tran) $\mathrm{U}(u) \wedge \mathrm{U}(v) \to [\mathrm{T}(v(ux)) \to \mathrm{T}(vx)]$.

(U-Dir) $\mathrm{U}(u) \wedge \mathrm{U}(v) \to \exists w.\ \mathrm{U}(w) \wedge u \sqsubset w \wedge v \sqsubset w$.

(U-Nor) $\mathrm{U}(u) \to \exists f.\ \mathrm{C}(f) \wedge u = \mathsf{I}f$.

(U-Lin) $\mathrm{U}(u) \wedge \mathrm{U}(v) \to u \sqsubset v \vee v \sqsubset u \vee \forall x. x \dot{\in} u \leftrightarrow x \dot{\in} v$.

Therefore, Proposition 4.11 also verifies the consistency of KFUPI + (U-Tran) + (U-Dir) + (U-Nor) + (U-Lin).

4.4. 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 4.3. 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 [19].

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:

$$\mathfrak{A}(X, u) := \mathfrak{A}_0(X, u) \vee ([\forall x.\ \mathfrak{A}_0(X, x) \to x \in X] \wedge \mathfrak{A}_1(X, u)).$$

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:

- $\mathrm{P}^\alpha_{\mathfrak{A}}(s) := \mathrm{P}_{\mathfrak{A}}(\alpha, s)$,

[3] In fact, this ordinal ι is identical to the first recursively inaccessible ordinal, for the definition of which, see, e.g. [2].

- $\mathrm{P}^{<\alpha}_{\mathfrak{A}}(s) := \exists \xi < \alpha.\ \mathrm{P}^{\xi}_{\mathfrak{A}}(s)$,
- $\mathrm{P}_{\mathfrak{A}}(s) := \exists \alpha.\ \mathrm{P}^{\alpha}_{\mathfrak{A}}(s)$.

Definition 4.13. *The $\mathcal{L}_{\mathcal{K}}$-theory* $\mathsf{FID}([\mathsf{POS}, \mathsf{QF}])$ *consists of* PA *with the full induction schema for $\mathcal{L}_{\mathcal{K}}$ and the following axioms:*[4]

1. $\alpha \not< \alpha \wedge (\alpha < \beta \wedge \beta < \gamma \to \alpha < \gamma) \wedge (\alpha < \beta \vee \alpha = \beta \vee \alpha > \beta)$.
2. $(\mathsf{OP}.1)$ $\mathrm{P}^{\alpha}_{\mathfrak{A}}(s) \leftrightarrow \mathrm{P}^{<\alpha}_{\mathfrak{A}}(s) \vee \mathfrak{A}(\mathrm{P}^{<\alpha}_{\mathfrak{A}}, s)$,
   $(\mathsf{OP}.2)$ $\mathfrak{A}(\mathrm{P}_{\mathfrak{A}}, s) \to \mathrm{P}_{\mathfrak{A}}(s)$.
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 [24, 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. [7, Theorem 4.13]). Next, to expand $\star$ 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. $\exists b, c.\ a = (b \dot{=} c) \wedge b = c$
2. $\exists b, c.\ a = \dot{\neg}(b \dot{=} c) \wedge b \neq c$
3. $\exists b.\ a = (\dot{\mathsf{N}} b) \wedge \mathrm{N}(b)$
4. $\exists b.\ a = \dot{\neg}(\dot{\mathsf{N}} b) \wedge \neg \mathrm{N}(b)$
5. $\exists b.\ a = \dot{\neg}(\dot{\neg} b) \wedge b \in X$
6. $\exists b, c.\ a = (b \dot{\wedge} c) \wedge b \in X \wedge c \in X$
7. $\exists b, c.\ a = \neg(b \dot{\wedge} c) \wedge [\dot{\neg} b \in X \vee \dot{\neg} c \in X]$
8. $\exists b, c.\ a = (b \dot{\to} c) \wedge [\dot{\neg} b \in X \vee c \in X]$
9. $\exists b, c.\ a = \dot{\neg}(b \dot{\to} c) \wedge b \in X \wedge \dot{\neg} c \in X$
10. $\exists f.\ a = \dot{\forall} f \wedge \forall b.\ fb \in X$
11. $\exists f.\ a = \dot{\neg}(\dot{\forall} f) \wedge \exists b.\ \dot{\neg}(fb) \in X$
12. $\exists f.\ a = \mathsf{I} f(0 \dot{=} 0) \wedge \forall b.\ fb \in X \vee \dot{\neg} fb \in X$

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. $\exists f, b.\ a = \mathsf{I} fb \wedge \mathsf{I} f(0 \dot{=} 0) \in X \wedge \mathsf{I} f(\dot{\mathsf{N}} 0) \notin X \wedge b \in X$.
2. $\exists f, b.\ a = \dot{\neg}(\mathsf{I} fb) \wedge \mathsf{I} f(0 \dot{=} 0) \in X \wedge \mathsf{I} f(\dot{\mathsf{N}} 0) \notin X \wedge b \notin X$.

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 $\mathsf{I} f(\dot{\mathsf{N}} 0) \notin X$ 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 4.14). Finally, the operator form $\mathfrak{A}$ is defined, as explained above:

$$\mathfrak{A}(X, a) := \mathfrak{A}_0(X, a) \vee ([\forall x.\ \mathfrak{A}_0(X, x) \to x \in X] \wedge \mathfrak{A}_1(X, a)).$$

To complete the definition of the translation $\star$, 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 $\star$, 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.

[4] In contrast to Jäger and Studer's formulation of $\mathsf{FID}([\mathsf{POS}, \mathsf{QF}])$ in [19], we now consider only one operator form $\mathfrak{A}$ for simplicity.

Lemma 4.14. *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* $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0)) \wedge \neg\mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$. *Furthermore, for this* α, *we have the following:*

1. $\forall b.\ \mathrm{P}_{\mathfrak{A}}(\mathsf{l}fb) \leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}fb)$,
2. $\forall b.\ \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(\mathsf{l}fb)) \leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\dot{\neg}(\mathsf{l}fb))$.

Proof. We assume that $\forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. Firstly, we prove $\mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$. For a contradiction, we suppose $\neg\mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$. Since $(\mathsf{OP}.2)$ implies that $\mathrm{P}_{\mathfrak{A}}$ is closed under $\mathfrak{A}_0$, we have $\mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(0\dot{=}0))$. 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 $\neg\mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$, we obtain $\mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$, a contradiction. Thus, $\mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$ is proved. By the transfinite induction schema, we also have an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0)) \wedge \neg\mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$.

Next, we show $\mathrm{P}_{\mathfrak{A}}(\mathsf{l}fb) \rightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}fb)$. Note that the right-to-left direction is obvious. Since $\mathrm{P}^{<\alpha}_{\mathfrak{A}}(0\dot{=}0)$, it suffices to consider the case $b \neq (0\dot{=}0)$. Now, we assume $\mathrm{P}_{\mathfrak{A}}(\mathsf{l}fb)$, and hence we can take a least β such that $\mathrm{P}^{\beta}_{\mathfrak{A}}(\mathsf{l}fb)$. Then, β is clearly equal to α, and thus we have $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}fb)$, as required. We can similarly verify the second item $\mathrm{P}_{\mathfrak{A}}(\dot{\neg}(\mathsf{l}fb)) \leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(\dot{\neg}(\mathsf{l}fb))$. □

Lemma 4.15. *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 3.3 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 $u = \mathsf{l}f \wedge \forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. By Lemma 4.14, there exists an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0)) \wedge \neg\mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$. Moreover, for any closed term g, we have:

$$\begin{aligned}
\mathrm{P}_{\mathfrak{A}}(u(\dot{\forall}g)) &\leftrightarrow \mathrm{P}^{\alpha}_{\mathfrak{A}}(u(\dot{\forall}g)) && (\because \text{Lemma 4.14})\\
&\leftrightarrow \mathrm{P}^{<\alpha}_{\mathfrak{A}}(\dot{\forall}g) && (\because \text{def. of } \mathfrak{A})\\
&\leftrightarrow \forall x.\ \mathrm{P}^{<\alpha}_{\mathfrak{A}}(gx) && (\because \mathrm{P}^{<\alpha} \text{ is } \mathfrak{A}_0\text{-closed})\\
&\leftrightarrow \forall x.\ \mathrm{P}^{\alpha}_{\mathfrak{A}}(u(gx)) && (\because \text{def. of } \mathfrak{A})\\
&\leftrightarrow \forall x.\ \mathrm{P}_{\mathfrak{A}}(u(gx)). && (\because \text{Lemma 4.14})
\end{aligned}$$

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 $(\mathsf{U\text{-}Cons})$:

$$\mathrm{U}^{\star}(u) \rightarrow \forall x.\ \neg[\mathrm{T}^{\star}(ux) \wedge \mathrm{T}^{\star}(u(\dot{\neg}x))].$$

As above, from the assumption $\mathrm{U}^{\star}(u)$, we take a closed term f such that $u = \mathsf{l}f \wedge \forall x.\mathrm{P}_{\mathfrak{A}}(fx) \vee \mathrm{P}_{\mathfrak{A}}(\dot{\neg}(fx))$. Also, we get an ordinal α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0)) \wedge \neg\mathrm{P}^{<\alpha}_{\mathfrak{A}}(\mathsf{l}f(\dot{\mathsf{N}}0))$. 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 4.14 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):

$$\mathrm{C}^{\star}(f) \rightarrow \mathrm{U}^{\star}(\mathsf{l}f) \wedge (f \sqsubset \mathsf{l}f)^{\star}.$$

We assume $\mathrm{C}^{\star}(f)$, then $\mathrm{U}^{\star}(\mathsf{l}f)$ is clear from the definition. Thus, we show the translation of $f \sqsubset \mathsf{l}f$:

1. $\forall x.\ \mathrm{P}_{\mathfrak{A}}(fx) \rightarrow \mathrm{P}_{\mathfrak{A}}(\mathsf{l}f(fx))$,

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 $\mathsf{C}^{\star}(f)$, Lemma 4.14 implies that there exists a least α such that $\mathrm{P}^{\alpha}_{\mathfrak{A}}(\mathsf{I}f(\dot{\mathrm{N}}0))$. 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 4.11. $\square$

Theorem 4.16. KFUPI *and* T_0 *are proof-theoretically equivalent.*

Proof. The lower bound of KFUPI is given by Theorem 4.10. As for the upper bound, Lemma 4.15 showed that KFUPI is proof-theoretically reducible to $\mathsf{FID}([\mathsf{POS},\mathsf{QF}])$, which is known to be proof-theoretically equivalent to T_0 [19]. $\square$

5. Frege structure by the Aczel-Feferman schema

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 5.7).

5.1. 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:

$\neg$	
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

Note that while $A \vee B$ is definable as $\neg(\neg A \wedge \neg B)$, $\to$ cannot be defined by $\neg$ and $\wedge$.

Based on the truth table, the theory PT (proposition and truth) is defined as follows.

Definition 5.1 (cf. [11]). *The* $\mathcal{L}_{FS}$*-theory* PT *has* TON *and (the universal closure of) the following axioms:*

Compositional axioms.

- $\mathrm{T}(x \dot{=} y) \leftrightarrow x = y$
- $\mathrm{T}(\dot{\neg}(x \dot{=} y)) \leftrightarrow x \neq y$
- $\mathrm{T}(\dot{\mathrm{N}}x) \leftrightarrow \mathrm{N}(x)$
- $\mathrm{T}(\dot{\neg}(\dot{\mathrm{N}}x)) \leftrightarrow \neg\mathrm{N}(x)$
- $\mathrm{T}(\dot{\neg}(\dot{\neg}x)) \leftrightarrow \mathrm{T}(x)$
- $\mathrm{T}(x \dot{\wedge} y) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(y)$
- $\mathrm{T}(\dot{\neg}(x \dot{\wedge} y)) \leftrightarrow [\mathrm{T}(\dot{\neg}x) \wedge \mathrm{T}(\dot{\neg}y)] \vee [\mathrm{T}(\dot{\neg}x) \wedge \mathrm{T}(y)] \vee [\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}y)]$
- $\mathrm{T}(x \dot{\to} y) \leftrightarrow [\mathrm{T}(x) \wedge \mathrm{T}(y)] \vee \mathrm{T}(\dot{\neg}x)$
- $\mathrm{T}(\dot{\neg}(x \dot{\to} y)) \leftrightarrow [\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}y)]$
- $\mathrm{T}(\dot{\forall} f) \leftrightarrow \forall x.\ \mathrm{T}(fx)$
- $\mathrm{T}(\dot{\neg}(\dot{\forall} f)) \leftrightarrow [\forall x.\ \mathrm{T}(fx) \vee \mathrm{T}(\dot{\neg}(fx))] \wedge \exists x.\ \mathrm{T}(\dot{\neg}(fx))$

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 5.2 (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. $\mathrm{P}(x \dot{\rightarrow} y) \leftrightarrow \mathrm{P}(x) \wedge (\mathrm{T}(x) \rightarrow \mathrm{P}(y))$,
5. $\mathrm{P}(\dot{\forall} f) \leftrightarrow \forall x.\, \mathrm{P}(fx)$.

The proof-theoretic analysis of PT is found in, e.g., [3, 8, 17, 13]:

Fact 5.3. PT *and* EM *have the same* $\mathcal{L}$*-theorems.*

Definition 5.4. *The* $\mathcal{L}_{FS}$*-system* PTU *consists of* TON *and the following axioms:*

Compositional axioms in U.

- $\mathrm{U}(u) \rightarrow \forall x, y.\, \mathrm{T}(x \dot{=} y) \leftrightarrow x = y$
- $\mathrm{U}(u) \rightarrow \forall x, y.\, \mathrm{T}(x \dot{\neq} y) \leftrightarrow x \neq y$
- $\mathrm{U}(u) \rightarrow \forall x.\, \mathrm{T}(\dot{\mathsf{N}}x) \leftrightarrow \mathrm{N}(x)$
- $\mathrm{U}(u) \rightarrow \forall x.\, \mathrm{T}(\dot{\neg}(\dot{\mathsf{N}}x)) \leftrightarrow \neg\mathrm{N}(x)$
- $\mathrm{U}(u) \rightarrow \forall x.\, \mathrm{T}(\dot{\neg}(\dot{\neg}x)) \leftrightarrow \mathrm{T}(x)$
- $\mathrm{U}(u) \rightarrow \forall x, y.\, \mathrm{T}(x \dot{\wedge} y) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(y)$
- $\mathrm{U}(u) \rightarrow \forall x, y.\, \mathrm{T}(\dot{\neg}(x \dot{\wedge} y)) \leftrightarrow [\mathrm{T}(\dot{\neg}x) \wedge \mathrm{T}(\dot{\neg}y)] \vee [\mathrm{T}(\dot{\neg}x) \wedge \mathrm{T}(y)] \vee [\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}y)]$
- $\mathrm{U}(u) \rightarrow \forall x, y.\, \mathrm{T}(x \dot{\rightarrow} y) \leftrightarrow [\mathrm{T}(x) \wedge \mathrm{T}(y)] \vee \mathrm{T}(\dot{\neg}x)$
- $\mathrm{U}(u) \rightarrow \forall x, y.\, \mathrm{T}(\dot{\neg}(x \dot{\rightarrow} y)) \leftrightarrow \mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}y)$
- $\mathrm{U}(u) \rightarrow \forall f.\, \mathrm{T}(\dot{\forall} f) \leftrightarrow \forall x.\, \mathrm{T}(fx)$
- $\mathrm{U}(u) \rightarrow \forall f.\, \mathrm{T}(\dot{\neg}(\dot{\forall} f)) \leftrightarrow [\forall x.\, \mathrm{T}(fx) \vee \mathrm{T}(\dot{\neg}(fx))] \wedge \exists x.\, \mathrm{T}(\dot{\neg}(fx))$

Consistency in U.

- $\mathrm{U}(u) \rightarrow \forall x.\, \neg[\mathrm{T}(x) \wedge \mathrm{T}(\dot{\neg}x)]$

Structural properties of U.

- $\mathrm{U}(u) \rightarrow \mathrm{C}(u)$
- $\mathrm{U}(u) \rightarrow \forall x.\, \mathrm{T}(ux) \rightarrow \mathrm{T}(x)$
- $\mathrm{C}(f) \rightarrow \mathrm{U}(\mathsf{l}f) \wedge f \sqsubset \mathsf{l}f$

As far as the author knows, the system PTU has not been explicitly formulated in the literature, though Fujimoto [14, pp. 933-935] formulated and analysed transfinite iterations of the truth theory DT, 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 5.5. PTU *and* EMU *have the same* $\mathcal{L}$*-theorems.*

5.2. 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 5.2. 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. $\exists y.\ x = (\dot{\mathrm{N}} y) \wedge \mathrm{N}(y)$
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. $\exists y, z.\ x = y \dot{\rightarrow} z \wedge B(y) \wedge [\mathrm{T}(y) \rightarrow B(z)]$
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:

$$[\forall x.\ \mathscr{A}^{\mathsf{PT}}(B(\bullet), x) \rightarrow B(x)] \rightarrow \forall x.\ \mathrm{P}(x) \rightarrow B(x),$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

Definition 5.6. *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 4.2 can be given in PTUPI; thus, we have $\mathsf{T}_0 \leq \mathsf{PTUPI}$. Moreover, Kahle's model construction in Section 4.3 and, thus, the upper-bound proof of KFUPI in Section 4.4 are easily modified for PTUPI. As a result, we also obtain the upper bound.

Theorem 5.7. PTUPI *and* T_0 *are proof-theoretically equivalent.*

6. Extension by least universes

In Sections 4 and 5, we extended theories of Frege structure by using the induction principle on propositions, in analogy with name induction in [21]. 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 [21, Conclusion 25], we can expect to obtain a strong theory of Frege structure by requiring some kind of leastness on universes. In fact, Burgess [4] 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 4.3, 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 4.12, similar to [21, 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 $\mathsf{I} g \sqsubset \mathsf{I} f$ holds. Taking these into consideration, we shall define the least universe $\mathsf{I} f$ 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 4.1, rather than *truth*.

Let $\mathrm{C}_{\mathsf{I} f}(x) := \forall y.\ \mathrm{P}_{\mathsf{I} f}(xy) := \forall y.\ \mathrm{T}(\mathsf{I} f(xy)) \vee \mathrm{T}(\mathsf{I} f(\dot{\neg}(xy)))$. Thus, this means that x is a class within $\mathsf{I} f$. 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:

[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.

1. $\exists y, z.\ x = (y \dot{=} z) \land y = z$
2. $\exists y.\ x = (\dot{\mathrm{N}} y) \land \mathrm{N}(y)$
3. $\exists y.\ x = \dot{\neg} y \land B(y)$
4. $\exists y, z.\ x = (y \dot{\land} z) \land \{[B(y) \land B(z)] \lor [B(y) \land \mathrm{T}(\mathsf{l} f(\dot{\neg} y))] \lor [B(z) \land \mathrm{T}(\mathsf{l} f(\dot{\neg} z))]\}$
5. $\exists y, z.\ x = (y \dot{\rightarrow} z) \land \{[B(y) \land B(z)] \lor [B(y) \land \mathrm{T}(\mathsf{l} f(\dot{\neg} y))] \lor [B(z) \land \mathrm{T}(\mathsf{l} f z)]\}$
6. $\exists g.\ x = (\dot{\forall} g) \land \{[\forall y.\ B(gy)] \lor \exists y.\ B(gy) \land \mathrm{T}(\mathsf{l} f(\dot{\neg}(gy))\}$
7. $\exists y.\ x = fy$
8. $\exists g, y.\ x = \mathsf{l} gy \land \mathrm{P}_{\mathsf{l} f}(x)$

The least universe schema (LU) is given by:

$$[\mathrm{C}(f) \land \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, B(\bullet), x) \rightarrow B(x)] \rightarrow \forall x.\ \mathrm{P}_{\mathsf{l} f}(x) \rightarrow B(x), \tag{LU}$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

Definition 6.1. *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 6.2. $\mathsf{KFLU} \vdash \mathrm{C}(f) \rightarrow \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, \mathrm{P}_{\mathsf{l} f}(\bullet), x) \rightarrow \mathrm{P}_{\mathsf{l} f}(x)$.

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 4.8 and 4.9 for the derivation of $(\mathsf{IG}.2)'$ are now replaced by the following lemmata:

Lemma 6.3. *For any $\mathcal{L}_{FS}$-formula A, recall that $B_A(u, v, w)$ is the formula $\forall x.\ \mathrm{T}(\mathsf{acc}(u, v)x) \rightarrow F_A(u, v, w, x)$, where $F_A(u, v, w, x)$ is the conjunction of the following formulas:*

1. $w = t(u, v, x) \ \rightarrow\ A(x)$,
2. $\forall z.\ w = \{(\oplus(u, v)(1, (z, x)))\dot{\rightarrow} t(u, v, z)\} \rightarrow \mathrm{T}(\oplus(u, v)(1, (z, x))) \rightarrow A(z)$,
3. $\forall z.\ w = \{\oplus(u, v)(0, z)\dot{\rightarrow}(\oplus(u, v)(1, (z, x))\dot{\rightarrow} t(u, v, z))\} \rightarrow \mathrm{T}(\oplus(u, v)(0, z)) \rightarrow \mathrm{T}(\oplus(u, v)(1, (z, x))) \rightarrow A(z)$.

Then, KFLU *derives the following:*

$$\mathrm{C}(u) \land \mathrm{C}(v) \land \mathsf{Closed}'(u, v, A(\bullet)) \rightarrow \forall w.\ \mathscr{A}^{\mathsf{LU}}(\oplus(u, v), B_A(u, v, \bullet), w) \rightarrow B_A(u, v, w).$$

Lemma 6.4. KFUPI *derives the following:*

$$\mathrm{C}(a) \land \mathrm{C}(b) \land Closed(a, b, A) \rightarrow \forall x.\ \mathrm{T}(\mathsf{acc}(a, b)x) \rightarrow A(x).$$

Proof. We assume $\mathrm{C}(a), \mathrm{C}(b), Closed(a, b, A)$ and $\mathrm{T}(\mathsf{acc}(a, b)c)$ for an arbitrary c. Then, we have to derive $A(c)$. From the assumptions and Lemma 6.3, we obtain $\mathrm{C}(\mathsf{l}(\oplus(a, b)))$ and $\forall w.\ \mathscr{A}^{\mathsf{LU}}(\oplus(u, v), B_A(u, v, \bullet), w) \rightarrow B_A(u, v, w)$. Thus, ($\mathsf{LU}$) yields:

$$\forall x.\ \mathrm{P}_{\mathsf{l}(\oplus(a,b))}(x) \rightarrow B_A(a, b, x).$$

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:

$$\forall y.\ \mathrm{T}(\oplus(a, b)(0, y)) \rightarrow \mathrm{T}(\oplus(a, b)(1, (y, c))) \rightarrow \mathrm{T}(\mathsf{l}(\oplus(a, b))(t(a, b, y))). \tag{4}$$

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 (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 6.5. $\mathcal{M}_{\mathsf{KF}} \models \mathsf{KFLU}$.

Proof. By Proposition 4.11, we concentrate on the schema (LU):

$$[\mathrm{C}(f) \land \forall x.\ \mathscr{A}^{\mathsf{LU}}(f, B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{P}_{\mathsf{l}f}(x) \to B(x).$$

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 4.11). □

In conclusion, we obtain the proof-theoretic strength of KFLU :

Theorem 6.6. 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. $\exists y, z.\ x = (y \dot{=} z) \land y = z$
2. $\exists y, z.\ x = (\dot{\neg}(y \dot{=} z)) \land y \neq z$
3. $\exists y.\ x = (\dot{\mathrm{N}} y) \land \mathrm{N}(y)$
4. $\exists y.\ x = (\dot{\neg}(\dot{\mathrm{N}} y)) \land \neg \mathrm{N}(y)$
5. $\exists y.\ x = \dot{\neg}(\dot{\neg} y) \land B(y)$
6. $\exists y, z.\ x = (y \dot{\land} z) \land B(y) \land B(z)$
7. $\exists y, z.\ x = (\dot{\neg}(y \dot{\land} z)) \land \{[B(\dot{\neg} y) \land B(\dot{\neg} z)] \lor [B(y) \land B(\dot{\neg} z)] \lor [B(\dot{\neg} y) \land B(z)]\}$
8. $\exists y, z.\ x = (y \dot{\to} z) \land [\mathrm{T}(\mathsf{l}fy) \lor \mathrm{T}(\mathsf{l}f(\dot{\neg} y))] \land [\mathrm{T}(\mathsf{l}f(y)) \to B(z)]$
9. $\exists y, z.\ x = (\dot{\neg}(y \dot{\to} z)) \land B(y) \land B(\dot{\neg} z)$
10. $\exists g.\ x = (\dot{\forall} g) \land \forall y. B(gy)$
11. $\exists g.\ x = (\dot{\neg}(\dot{\forall} g)) \land [\forall y. B(gy) \lor B(\dot{\neg}(gy))] \land \exists y. B(\dot{\neg}(gy))$
12. $\exists y.\ [x = fy \lor x = \dot{\neg}(fy)] \land \mathrm{T}(x)$
13. $\exists g, y.\ [x = \mathsf{l}gy \lor x = \dot{\neg}(\mathsf{l}gy)] \land \mathrm{T}(\mathsf{l}fx)$

The least universe schema ($\mathsf{LU}^{\mathsf{PT}}$) is given by:

$$[\mathrm{C}(f) \land \forall x.\ \mathscr{A}^{\mathsf{PTLU}}(f, B(\bullet), x) \to B(x)] \to \forall x.\ \mathrm{T}(\mathsf{l}fx) \to B(x), \tag{$\mathsf{LU}^{\mathsf{PT}}$}$$

for all $\mathcal{L}_{FS}$-formulas $B(x)$.

Thus, the schema ($\mathsf{LU}^{\mathsf{PT}}$) informally says that $\mathrm{T}(\mathsf{l}fx)$ is the least truth set satisfying each clause of $\mathscr{A}^{\mathsf{PTLU}}$.

Definition 6.7. *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 6.8. PTLU *and* T_0 *are proof-theoretically equivalent.*

7. Frege structure by the supervaluation schema

In this section, we consider the supervaluational Frege structure. Kripke initially sketched a semantic theory of truth based on the supervaluation schema [28, p. 711]. Based on Kripke's semantic definition, Cantini [5] defined and studied a formal theory VF (van Fraassen) over Peano arithmetic. In particular, Cantini [5] proved that VF is proof-theoretically equivalent to the theory ID_1 of arithmetical monotone inductive definition. In [7, 23], 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 [23, 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. [23, 25]). 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.

7.1. System VF *and universes*

In this section, we add a constant symbol $\dot{\mathrm{T}}$ to $\mathcal{L}_{FS}$ for technical reasons (see Remark 7.2). Following [23, 24], we inductively define a corresponding term $\dot{A}$ for each formula A.

- $\overset{\cdot}{\overbrace{s = t}} := (s \dot{=} t)$, $\overset{\cdot}{\overbrace{\mathrm{N(s)}}} := \dot{\mathsf{N}}s$, $\overset{\cdot}{\overbrace{\mathrm{T(s)}}} := \dot{\mathsf{T}}s$;
- $\overset{\cdot}{\overbrace{\neg A}} := \dot{\neg}\dot{A}$, $\overset{\cdot}{\overbrace{A \wedge B}} := \dot{A} \dot{\wedge} \dot{B}$, $\overset{\cdot}{\overbrace{A \to B}} := \dot{A} \dot{\to} \dot{B}$;
- $\overset{\cdot}{\overbrace{\forall x.\, A}} := \dot{\forall}(\lambda x.\dot{A})$.

Then, we define the $\mathcal{L}_{FS}$-theory VF, the formulation of which is essentially based on [5, 7, 23].

Definition 7.1. *The $\mathcal{L}_{FS}$-theory* VF *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) $\mathrm{T}(\overset{\cdot}{\overbrace{A \to B}}) \to (\mathrm{T}(\dot{A}) \to \mathrm{T}(\dot{B}))$;

(T-Univ) $(\forall x.\, \mathrm{T}(\dot{A})) \to \mathrm{T}(\overset{\cdot}{\overbrace{\forall x.\, A}})$;[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.

Remark 7.2. In Kahle's formulation of VF (called SON in [23]), 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 [23, pp. 111-112]. While this definition is not problematic in KF (Definition 3.1) and PT (Definition 5.1), 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., [12]). 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 3.3) 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 7.3. *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;*

[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 3.1. The reason the author choses the schematic form is just to simplify the upper-bound proof in Section 7.4. Nevertheless, the author believes that this single axiom does not affect the proof-theoretic strength of both VF and VFU (Definition 7.3).

[7] Note that Stern's supervaluation-style truth [32] is an attempt to overcome this difficulty of VF.

(U-Imp) $\mathrm{U}(u) \to [\mathrm{T}(u(\overbrace{A \to B}^{\cdot})) \to \{\mathrm{T}(u\dot{A}) \to \mathrm{T}(u\dot{B})\}]$;

(U-Univ) $\mathrm{U}(u) \to [\{\forall x.\, \mathrm{T}(u(\dot{A}))\} \to \mathrm{T}(u(\overbrace{\forall x.\, A}^{\cdot}))]$;

(U-Log) $\mathrm{U}(u) \to \mathrm{T}(u(\dot{C}))$ *for any logical theorem* C;

(U-Cons) $\mathrm{U}(u) \to \forall x.\, \neg[\mathrm{T}(ux) \land \mathrm{T}(u(\dot{\neg} x))]$;

(U-Self) $\mathrm{U}(u) \to [\mathrm{T}(ux) \leftrightarrow \mathrm{T}(u(\dot{\mathrm{T}}x))] \ \land \ [\mathrm{T}(u(\dot{\neg}x)) \leftrightarrow \mathrm{T}(u(\dot{\neg}(\dot{\mathrm{T}}x)))]$;

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) $\mathrm{C}(f) \to \mathrm{U}(\mathsf{l}f) \land f \sqsubset \mathsf{l}f$.

Similarly to Fact 3.4, we can prove the following:

Lemma 7.4. VF *is a subtheory of* VFU.

7.2. 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 4.10. 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 [5, 7].

The following two lemmata are essentially by way of [7, Lemma 59.2].

Lemma 7.5. *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 7.6. *In* VFU, *the following are derivable:*

1. $\mathrm{T}(\dot{A}) \land \mathrm{T}(\dot{B}) \leftrightarrow \mathrm{T}(\overbrace{A \land B}^{\cdot})$;
2. $\mathrm{T}(\dot{A}) \lor \mathrm{T}(\dot{B}) \to \mathrm{T}(\overbrace{A \lor B}^{\cdot})$;
3. $(\mathrm{T}(\dot{A}) \lor \mathrm{T}(\overbrace{\ \neg A\ }^{\cdot})) \to [\{\mathrm{T}(\dot{A}) \to \mathrm{T}(\dot{B})\} \leftrightarrow \mathrm{T}(\overbrace{A \to B}^{\cdot})]$;
4. $[\forall x.\, \mathrm{T}(\dot{A})] \leftrightarrow \mathrm{T}(\overbrace{\ \forall x A\ }^{\cdot})$;
5. $[\exists x.\, \mathrm{T}(\dot{A})] \to \mathrm{T}(\overbrace{\ \exists x A\ }^{\cdot})$;
6. $[\mathrm{T}(a) \lor \mathrm{T}(\dot{\neg} a)] \to [\neg \mathrm{T}(\dot{\mathrm{T}} a) \leftrightarrow \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}} a))]$.

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 7.7. *Let a term* $\underline{\mathsf{j}}$ *be such that* $\underline{\mathsf{j}}(x, f) = \lambda z.\overbrace{\exists v, w.\, z = (v, w) \land \mathrm{T}(xv) \land \mathrm{T}(fvw)}^{\cdot}$. *Then,* $\mathsf{VFU} \vdash (\mathrm{join})'$, *that is,*

$$\mathsf{VFU} \vdash \mathrm{C}(x) \land [\forall y.\, \mathrm{T}(xy) \to \mathrm{C}(fy)] \to \mathrm{C}(\underline{\mathsf{j}}(x, f)) \land \Sigma'(x, f, \underline{\mathsf{j}}(x, f)),$$

where, $\Sigma'(x, f, \underline{\mathsf{j}}(x, f))$ *is the following:*

$$\forall z.\, \mathrm{T}(\underline{\mathsf{j}}(x, f)z) \leftrightarrow \exists v, w.\, z = (v, w) \land \mathrm{T}(xv) \land \mathrm{T}(fvw).$$

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 $\forall y(\mathrm{T}(xy) \to \mathrm{C}(fy))$. As the second conjunct is obvious from Lemma 7.5, 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 7.6:

$$
\begin{aligned}
&\neg\mathrm{T}(\overbrace{\exists v, w.\, z = (v, w) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw)}^{\cdot}) \\
&\Longrightarrow\ \neg\exists v, w.\, \mathrm{T}(\overbrace{z = (v, w) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw)}^{\cdot}) \\
&\Longleftrightarrow \neg\exists v, w.\, \mathrm{T}(z = (v, w)) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw) \\
&\Longleftrightarrow \forall v, w.\, \neg\mathrm{T}(z = (v, w)) \vee \neg\mathrm{T}(xv) \vee \neg\mathrm{T}(fvw) \\
&\Longleftrightarrow \forall v, w.\, \mathrm{T}(\dot{\neg}(z \dot{=} (v, w))) \vee \mathrm{T}(\dot{\neg}(xv)) \vee \mathrm{T}(\dot{\neg}(fvw)) \\
&\Longrightarrow \forall v, w.\, \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}}(z \dot{=} (v, w)))) \vee \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}}(xv))) \vee \mathrm{T}(\dot{\neg}(\dot{\mathrm{T}}(fvw))) \\
&\Longrightarrow \forall v, w.\, \mathrm{T}(\overbrace{\neg(\mathrm{T}(z = (v, w)) \wedge \mathrm{T}(xv) \wedge \mathrm{T}(fvw))}^{\cdot}) \\
&\Longleftrightarrow \mathrm{T}(\dot{\neg}(\underline{\mathsf{j}}(x, f)z)).
\end{aligned}
$$

□

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
$WF[a, b, f] := \forall x.\, \mathrm{T}(ax) \to [\forall y.\, \mathrm{T}(ay) \to \mathrm{T}(b(y, x)) \to \mathrm{T}(\mathsf{l}_{a,b}(fy))] \to \mathrm{T}(\mathsf{l}_{a,b}(fx))$;
$TI[a, b, z] := \mathrm{T}(az) \wedge \forall f.\, WF[a, b, f] \to \mathrm{T}(\mathsf{l}_{a,b}(fz))$.

The term $\mathsf{l}_{a,b}$ 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 7.8. *For each class u and v, we can take a universe $\mathsf{l}_{u,v}$ that reflects on both u and v. That is,*

$$\mathsf{VFU} \vdash \forall u, v.\, \mathrm{C}(u) \wedge \mathrm{C}(v) \to [\mathrm{U}(\mathsf{l}_{u,v}) \wedge u \sqsubset \mathsf{l}_{u,v} \wedge v \sqsubset \mathsf{l}_{u,v}].$$

Proof. Take any classes u and v. We define a term $\oplus$ to be such that $u \oplus v = \lambda x.\overbrace{\mathrm{N}((x)_0) \to \mathrm{T}(\mathsf{d_N}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0})}^{\cdot}$ and let $\mathsf{l}_{u,v} := \mathsf{l}(u \oplus v)$. First, to show that $\mathsf{l}_{u,v}$ 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 $\neg\mathrm{N}((x)_0)$, we have $\mathrm{T}(\overbrace{\mathrm{N}((x)_0) \to \mathrm{T}(\mathsf{d_N}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0})}^{\cdot})$, hence $\mathrm{P}((u \oplus v)x)$. Thus, we can assume $\mathrm{N}((x)_0)$. If $(x)_0 = \mathsf{0}$, then we have $\mathsf{d_N}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0} = u(x)_1$. As u is a class, $(u \oplus v)x$ is a proposition, as required. Similarly, if $(x)_0 \neq \mathsf{0}$, then $\mathsf{d_N}(u(x)_1)(v(x)_1)((x)_0)\mathsf{0} = v(x)_1$, 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 $\mathsf{l}_{u,v}$ is a class. Second, we show that $u \sqsubset \mathsf{l}_{u,v}$. 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 $\mathrm{T}(\mathsf{l}_{u,v}((u \oplus v)(\mathsf{0}, x)))$, which, by ($\mathsf{U\text{-}Log}$), ($\mathsf{U\text{-}Imp}$), and ($\mathsf{U\text{-}Self}$), implies $\mathrm{T}(\mathsf{l}_{u,v}(ux))$. Similarly, $\mathrm{T}(\dot{\neg}(ux))$ implies $\mathrm{T}(\mathsf{l}_{u,v}(\dot{\neg}(ux)))$. Thus, the conclusion $u \sqsubset \mathsf{l}_{u,v}$ is obtained. In the same way, we also have $v \sqsubset \mathsf{l}_{u,v}$. □

Lemma 7.9.
1. $\mathsf{VFU} \vdash \mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathrm{C}(\mathsf{acc}(a, b))$;
2. $\mathsf{VFU} \vdash \mathrm{C}(a) \wedge \mathrm{C}(b) \to \mathsf{Closed}'(a, b, \mathsf{acc}(a, b))$;
3. $\mathsf{VFU} \vdash \mathrm{C}(a) \wedge \mathrm{C}(b) \wedge \mathsf{Closed}'(a, b, A(\bullet)) \to \forall x.\, \mathrm{T}(\mathsf{acc}(a, b)x) \to A(x)$, *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 7.6, we observe that $\overbrace{\forall y.\, \mathrm{T}(ay) \to \mathrm{T}(b(y, x)) \to \mathrm{T}(\mathsf{l}_{a,b}(fy))}^{\cdot}$ 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 $\mathsf{Closed}'(a, b, \mathrm{T}(\mathsf{l}_{a,b}(A'(\bullet))))$ in the following way:

$$
\begin{aligned}
&\mathrm{T}(\mathsf{l}_{a,b}(\overbrace{\mathsf{Closed}'(a, b, A'(\bullet))}^{\cdot}))\\
&\Longrightarrow \forall x.\ \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{\mathrm{T}(ax)}^{\cdot})) \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to A'(y)}^{\cdot})) \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(x)}^{\cdot}))\\
&\Longleftrightarrow \forall x.\ \mathrm{T}(ax) \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to A'(y)}^{\cdot})) \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(x)}^{\cdot}))\\
&\Longleftrightarrow \forall x.\ \mathrm{T}(ax) \to [\forall y.\ \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{\mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to A'(y)}^{\cdot}))] \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(x)}^{\cdot}))\\
&\Longleftrightarrow \forall x.\ \mathrm{T}(ax) \to [\forall y.\ \mathrm{T}(ay) \to \mathrm{T}(b(y,x)) \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(y)}^{\cdot}))] \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(x)}^{\cdot})).
\end{aligned}
$$

Letting $f := \lambda x.\overbrace{A'(x)}^{\cdot}$, the assumption $\mathrm{T}(\mathsf{acc}(a, b)x)$ implies in VFU the formula $\mathsf{Closed}'(a, b, \mathrm{T}(\mathsf{l}_{a,b}(A'(\bullet)))) \to \mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(x)}^{\cdot}))$. Therefore, we obtain $\mathrm{T}(\mathsf{l}_{a,b}(\overbrace{A'(x)}^{\cdot}))$, which yields $A'(x)$ by (U-Out). Finally, combining this with the assumption $\mathsf{Closed}'(a, b, A(\bullet))$, the conclusion $A(x)$ follows. □

Theorem 7.10. *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.

7.3. 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 [5, 7], which is formalisable in a suitable set theory, and thus the upper-bound of VFU is obtained (see subsection 7.4). 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 $\overset{\alpha,\beta,\gamma}{\vdash} \Gamma$ means that Γ is derived in the system with the U-rank α, the T-rank, and the derivation length γ. We introduce several notations:

- $\overset{\alpha}{\vdash} \Gamma$ means $\overset{\alpha,\beta,\gamma}{\vdash} \Gamma$ for some α, β, γ;
- $\overset{<\alpha}{\vdash} \Gamma$ means $\overset{\alpha_0}{\vdash} \Gamma$ for some $\alpha_0 < \alpha$;
- $\overset{\alpha,\beta}{\vdash} \Gamma$ means $\overset{\alpha,\beta,\gamma}{\vdash} \Gamma$ for some γ;
- $\overset{\alpha,\beta,<\gamma}{\vdash} \Gamma$ means $\overset{\alpha,\beta,\gamma_0}{\vdash} \Gamma$ for some $\gamma_0 < \gamma$;
- $\overset{\alpha,<\beta,<\gamma}{\vdash} \Gamma$ means $\overset{\alpha,\beta_0,\gamma_0}{\vdash} \Gamma$ for some $\beta_0 < \beta$ and $\gamma_0 < \gamma$;
- $\overset{\leq\alpha,\leq\beta,\leq\gamma}{\vdash} \Gamma$ means $\overset{\alpha_0,\beta_0,\gamma_0}{\vdash} \Gamma$ for some $\alpha_0 \leq \alpha$, $\beta_0 \leq \beta$ and $\gamma_0 \leq \gamma$;
- $\vdash \Gamma$ means $\overset{\alpha}{\vdash} \Gamma$ 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 $(\neg\mathrm{T})$ respectively introduce T and $\neg$T, with an increase in the T-rank. Note that the context of the premise of (T) and $(\neg\mathrm{T})$ 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 $(\neg\mathrm{U})$ have the side condition $\dagger$, 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 $\vdash^{<\alpha} a : \mathrm{Class}$. 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 $\neg \vdash^{<\alpha} \mathrm{L}_a(b) : \mathrm{Prop}$.

Table 2: Sequent system $\vdash^{\alpha,\beta,\gamma}$

$$\dfrac{\mathcal{CTT} \models P}{\vdash^{\alpha,\beta,\gamma} \Gamma, P}\ (\mathrm{Lit}) \qquad \dfrac{}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{T}(a), \neg\mathrm{T}(a)}\ (\mathrm{Log})$$

$$\dfrac{\vdash^{\alpha,\beta,<\gamma} \Gamma, A_0 \wedge A_1, A_0 \quad \vdash^{\alpha,\beta,<\gamma} \Gamma, A_0 \wedge A_1, A_1}{\vdash^{\alpha,\beta,\gamma} \Gamma, A_0 \wedge A_1}\ (\wedge) \qquad \dfrac{\vdash^{\alpha,\beta,<\gamma} \Gamma, A_0 \vee A_1, A_i\ (i \leq 1)}{\vdash^{\alpha,\beta,\gamma} \Gamma, A_0 \vee A_1}\ (\vee)_i$$

$$\dfrac{\vdash^{\alpha,\beta,<\gamma} \Gamma, \exists x.\, A(x), A(a)}{\vdash^{\alpha,\beta,\gamma} \Gamma, \exists x.\, A(x)}\ (\exists) \qquad \dfrac{\vdash^{\alpha,\beta,<\gamma} \Gamma, \forall x.\, A(x), A(a), \text{ for all } a}{\vdash^{\alpha,\beta,\gamma} \Gamma, \forall x.\, A(x)}\ (\forall)$$

$$\dfrac{\vdash^{\alpha,<\beta,<\gamma} A}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{T}(\dot{A})}\ (\mathrm{T}) \qquad \dfrac{\vdash^{\alpha,<\beta,<\gamma} \neg A}{\vdash^{\alpha,\beta,\gamma} \Gamma, \neg\mathrm{T}(\dot{A})}\ (\neg\mathrm{T})$$

$$\dfrac{\vdash^{<\alpha} \Gamma}{\vdash^{\alpha,\beta,\gamma} \Gamma}\ (\mathrm{Weak})$$

$$\dfrac{\vdash^{<\alpha} b}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{L}_a(b)}\ (\mathrm{U})^{\dagger} \qquad \dfrac{\neg \vdash^{<\alpha} b}{\vdash^{\alpha,\beta,\gamma} \Gamma, \neg\mathrm{L}_a(b)}\ (\neg\mathrm{U})^{\dagger}$$

By transfinite induction, we easily obtain the following:

Lemma 7.11. (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 7.12 (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 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. [7, Theorem 62.1]). Thus, we confine ourselves to the base cases $A \equiv \mathrm{T}(\dot{B})$ and $A \equiv \mathrm{L}_a(\dot{B})$.

$A \equiv \mathrm{T}(\dot{B})$ Firstly, 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 7.11 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 $(\neg\mathrm{T})$:

$$\frac{\vdash^{\alpha,\beta',\gamma'} B}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{T}(\dot{B})}\ (\mathrm{T}) \qquad \frac{\vdash^{\delta,\varepsilon',\zeta'} \neg B}{\vdash^{\delta,\varepsilon,\zeta} \Delta, \neg\mathrm{T}(\dot{B})}\ (\neg\mathrm{T}),$$

where $\beta' < \beta, \gamma' < \gamma, \varepsilon' < \varepsilon$ and $\zeta' < \zeta$. Since $\beta' < \beta$, the induction hypothesis for the premises yields that $\vdash^{\max(\alpha,\delta)} \emptyset$, which contradicts Lemma 7.11. 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 $(\neg\mathrm{U})$, respectively:

$$\frac{\vdash^{<\alpha} b}{\vdash^{\alpha,\beta,\gamma} \Gamma, \mathrm{L}_a(b)}\ (\mathrm{U})^{\dagger} \qquad \frac{\neg \vdash^{<\delta} b}{\vdash^{\delta,\varepsilon,\zeta} \Delta, \neg\mathrm{L}_a(b)}\ (\neg\mathrm{U})^{\dagger}.$$

Here, we have the following side conditions:

$$\neg \vdash^{<\alpha} \mathrm{L}_a(b) : \mathrm{Prop} \tag{5}$$

$$\neg \vdash^{<\delta} \mathrm{L}_a(b) : \mathrm{Prop} \tag{6}$$

By (5) and (6), we clearly have $\alpha = \delta$. Therefore, the premise $\vdash^{<\alpha} b$ is identical with $\vdash^{<\delta} b$, which contradicts the other premise $\neg \vdash^{<\delta} b$. 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 4.3. 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 7.3.

Lemma 7.13. *Let a and b be any closed terms and suppose* $\vdash a : \mathrm{Class}$. *Then, the following hold in* $\mathcal{M}_{\mathsf{VFU}}$*:*

(U-Elem) *If* $\mathcal{CTT} \models P$, *then* $\vdash \mathrm{L}_a(\dot{P})$, *for each $\mathcal{L}$-literal sentence P.*

(U-Imp) *If* $\vdash \mathrm{L}_a(\overbrace{A \rightarrow B}^{\cdot})$ *and* $\vdash \mathrm{L}_a(\dot{A})$, *then* $\vdash \mathrm{L}_a(\dot{B})$.

(U-Univ) *If* $\vdash \mathrm{L}_a(\dot{\overbrace{A(c)}})$ *for every closed term c, then* $\vdash \mathrm{L}_a(\dot{\overbrace{\forall x.\, A(x)}})$.

(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* $\vdash \mathrm{L}_a(\dot{\overbrace{\mathrm{T}(b)}})$. *Similarly,* $\vdash \mathrm{L}_a(\dot{\neg} b)$ *if and only if* $\vdash \mathrm{L}_a(\dot{\overbrace{\neg \mathrm{T}(b)}})$.

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 $\vdash \mathrm{L}_a(\dot{\overbrace{A \to B}})$ 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 7.12, which contradicts Lemma 7.11. 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 7.12, it follows that $\vdash^{<\alpha} B$, and thus we obtain $\vdash^{\alpha} \mathrm{L}_a(\dot{B})$ by (U).

(U-Univ) Assuming $\vdash \mathrm{L}_a(\dot{\overbrace{A(c)}})$ for any closed term c, we show $\vdash \mathrm{L}_a(\dot{\overbrace{\forall x.\, A(x)}})$. Similarly to the above, we take the least successor ordinal α such that $\vdash^{<\alpha} a$: Class. Then, for all c, we clearly have $\vdash^{\alpha} \mathrm{L}_a(\dot{\overbrace{A(c)}})$, 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 $\vdash \mathrm{L}_a(\dot{\overbrace{\forall x.\, A(x)}})$.

The other cases are similarly proved. □

Similarly to Lemma 7.13, the structural properties of VFU are satisfied in $\mathcal{M}_{\mathsf{VFU}}$:

Lemma 7.14. *Let a and b be any closed terms and suppose* $\vdash a$: Class. *Then, the following hold in* $\mathcal{M}_{\mathsf{VFU}}$*:*

(U-Class) $\vdash \mathrm{L}_a(b)$ *or* $\vdash \neg\mathrm{L}_a(b)$.

(U-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 (U-Out) :

Lemma 7.15. *Let a be any closed term and suppose* $\vdash a$: Class. *Then, the following is satisfied in* $\mathcal{M}_{\mathsf{VFU}}$:

(U-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 (U-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 [7, 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 7.16. $\mathcal{M}_{\mathsf{VFU}} \models$ VFU.

Remark 7.17. Similar to Remark 4.12, we can also verify the axioms (U-Tran), (U-Dir), (U-Nor) and (U-Lin) in $\mathcal{M}_{\mathsf{VFU}}$.

7.4. *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 4.4 is not expressive enough to formalise the cut-admissibility argument of Lemma 7.12. 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 [19]. 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 7.18 (cf. [19]). *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;
- $Cl_{\Phi_i}(X) :\leftrightarrow \forall x.\ \Phi_i(X, x) \to x \in X$, 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 7.19 (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:

- $I^{\alpha,\beta,\gamma}(x) := \exists s, f.\ \{\alpha, \beta, \gamma\} \subseteq s \wedge \mathscr{H}(s, f) \wedge x \in f(\alpha, \beta, \gamma)$,
- $I^{\alpha,\leq\beta,<\gamma}(x) := \exists \beta_0 \leq \beta.\ \exists \gamma_0 < \gamma.\ I^{\alpha,\beta_0,\gamma_0}(x)$,
- $I^{\alpha}(x) := \exists \beta, \gamma.\ I^{\alpha,\beta,\gamma}$,
- $I^{<\alpha}(x)$, $I^{\alpha,\leq\beta,\leq\gamma}(x)$, $I^{\alpha,\beta,<\gamma}(x)$ and $I^{\alpha,<\beta,<\gamma}(x)$ are similarly defined.

We now show that the Σ-predicate $I^{\alpha,\beta,\gamma}(x)$ expresses the required class $\{\Gamma : \vdash^{\alpha,\beta,\gamma} \Gamma\}$.

Lemma 7.20. *The following are derivable in* KPi.

1. $\forall \alpha.\ \exists s, f.\ \alpha \in s \wedge \mathscr{H}(s, f)$.
2. $I^{\alpha,\beta,\gamma}(\Gamma)$ *if and only if one of the following holds:*
   (a) $I^{<\alpha}(\Gamma) \vee I^{\alpha,\leq\beta,\leq\gamma}(\Gamma)$,
   (b) $\Phi_0(I^{\alpha,\beta,<\gamma}, \Gamma) \vee \Phi_1(I^{\alpha,<\beta,<\gamma}, \Gamma)$,
   (c) $\alpha \in \mathrm{Suc} \wedge \Phi_2(I^{\alpha-1}, \Gamma)$.

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. [30, p. 256]).
2. For the left-to-right direction, we assume $I^{\alpha,\beta,\gamma}(\Gamma)$; thus, we take sets s and f such that $\{\alpha, \beta, \gamma\} \subseteq s \wedge \mathscr{H}(s, f) \wedge \Gamma \in f(\alpha, \beta, \gamma)$. For instance, we consider the case where $\Gamma \in f(\alpha, \beta, \gamma)$ is obtained from Φ_2, then we have $\alpha \in$ suc and $\Phi_2(f(\alpha - 1, \in s, \in s), \Gamma)$. By Lemma 7.19, we can easily show that $\forall x.\ x \in f(\alpha - 1, \in s, \in s) \leftrightarrow I^{\alpha-1}(x)$. Therefore, we get $\Phi_2(I^{\alpha-1}, \Gamma)$, as required. The other cases are similarly proved by using Lemma 7.19.
   As for the converse direction, we, for example, assume $\alpha \in \mathrm{Suc} \wedge \Phi_2(I^{\alpha-1}, \Gamma)$. 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 7.19, we have $\forall x.\ I^{\alpha-1}(x) \leftrightarrow x \in f(\alpha - 1, \in s, \in s)$. Therefore, we obtain $\Phi_2(f(\alpha - 1, \in s, \in s), \Gamma)$, and hence it follows that $\Gamma \in f(\alpha, \beta, \gamma)$. Thus, we obtain $I^{\alpha,\beta,\gamma}(\Gamma)$. 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 4.15. Then, we let $\mathrm{T}^+(x) :\equiv I^{\infty}(x)$; let $\mathrm{U}^+(x) :\leftrightarrow \exists a.\ x = \mathsf{l}a \wedge \forall b.\ I^{\infty}(ab) \vee I^{\infty}(\dot{\neg}(ab))$.

Lemma 7.21. *For each $\mathcal{L}_{FS}$-formula A, if* VFU $\vdash A$, *then* KPi $\vdash A^+$.

Proof. The proof is by directly running the proof of Theorem 7.16 within KPi.

(U-Class) We want to show KPi $\vdash (\mathrm{U}(u) \to \mathrm{C}(u))^+$. Thus, taking any term f such that $u = \mathsf{l}f$ and $\forall a.\ I^\infty(fa) \vee I^\infty(\dot\neg(fa))$, we show $\forall a.\ I^\infty(\mathsf{l}fa) \vee I^\infty(\dot\neg(\mathsf{l}fa))$. 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 7.20 implies that $\forall a.\ I^\alpha(\mathsf{l}fa) \vee I^\alpha(\dot\neg(\mathsf{l}fa))$, and thus we obtain $\forall a.\ I^\infty(\mathsf{l}fa) \vee I^\infty(\dot\neg(\mathsf{l}fa))$.

(U-True) Similar to the above.

(U-Out) Since KPi $\vdash$ (U-True)$^+$, it suffices to verify (T-Out)$^+$, that is, KPi $\vdash I^\infty(\dot A) \to A^+$ for each $\mathcal{L}_{FS}$-formula A. For that purpose, we formalise Lemma 7.15 within KPi, similarly to [5, 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 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 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 7.20. □

Combining Theorem 7.10 with Lemma 7.21, we obtain the proof-theoretic strength of VFU:

Theorem 7.22. VFU *and* T_0 *are proof-theoretically equivalent.*

8. Conclusion

The results of this paper are summarised as follows.

Conclusion 8.1. All the following theories are proof-theoretically equivalent to T_0:

- KFUPI and PTUPI,
- KFLU and PTLU,
- VFU.

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 [7] 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 by the addition of such a principle. Furthermore, Jäger and Strahm [22] 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.

References

[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] Jon Barwise. *Admissible sets and structures, perspectives in mathematical logic*. Springer Berlin, 1975.

[3] Michael J Beeson. *Foundations of constructive mathematics: Metamathematical studies*. Springer-Verlag, Berlin, 1985.

[4] John P Burgess. Friedman and the axiomatization of Kripke's theory of truth. *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, pages 125–148, 2009.

[5] Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *The Journal of Symbolic Logic*, 55(1):244–259, 1990.

[6] Andrea Cantini. Extending the first-order theory of combinators with self-referential truth. *The Journal of symbolic logic*, 58(2):477–513, 1993.

[7] 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.

[8] Andrea Cantini. About Truth and Types. In Reinhard Kahle, Thomas Strahm, and Thomas Studer, editors, *Advances in proof theory*, pages 31–64. Springer, 2016.

[9] Solomon Feferman. A language and axioms for explicit mathematics. In John Newsome Crossley, editor, *Algebra and Logic*, pages 87–139, Berlin, Heidelberg, 1975. Springer Berlin Heidelberg.

[10] Solomon Feferman. Reflecting on incompleteness. *The Journal of Symbolic Logic*, 56(1):1–49, 1991.

[11] Solomon Feferman. Axioms for determinateness and truth. *The Review of Symbolic Logic*, 1(2):204–217, 2008.

[12] Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.

[13] Kentaro Fujimoto. Relative truth definability of axiomatic truth theories. *Bulletin of symbolic logic*, 16(3):305–344, 2010.

[14] Kentaro Fujimoto. Autonomous progression and transfinite iteration of self-applicable truth. *The Journal of Symbolic Logic*, 76(3):914–945, 2011.

[15] Volker Halbach. Truth and reduction. *Erkenntnis*, 53(1-2):97–126, 2000.

[16] Volker Halbach and Kentaro Fujimoto. Classical determinate truth I. *Journal of Symbolic Logic*, pages 1–44, 2023.

[17] 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.

[18] Gerhard Jäger. A well-ordering proof for Feferman's theory T_0. *Archiv für mathematische Logik und Grundlagenforschung*, 23(1):65–77, 1983.

[19] Gerhard Jäger. First order theories for nonmonotone inductive definitions: recursively inaccessible and Mahlo. *The Journal of Symbolic Logic*, 66(3):1073–1089, 2001.

[20] 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.

[21] Gerhard Jäger, Reinhard Kahle, and Thomas Studer. Universes in explicit mathematics. *Annals of Pure and Applied Logic*, 109(3):141–162, 2001.

[22] Gerhard Jäger and Thomas Strahm. Reflections on reflections in explicit mathematics. *Annals of Pure and Applied Logic*, 136(1-2):116–133, 2005.

[23] Reinhard Kahle. Truth in applicative theories. *Studia Logica*, 68(1):103–128, 2001.

[24] Reinhard Kahle. Universes over Frege structures. *Annals of Pure and Applied Logic*, 119(1-3):191–223, 2003.

[25] 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.

[26] 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.

[27] 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.

[28] Saul Kripke. Outline of a theory of truth. *The journal of philosophy*, 72(19):690–716, 1976.

[29] 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.

[30] Wolfram Pohlers. *Proof theory: The first step into impredicativity*. Springer Science & Business Media, 2008.

[31] Thomas Schindler. A Disquotational Theory of Truth as Strong as Z_2^-. *Journal of Philosophical Logic*, 44:395–410, 2015.

[32] Johannes Stern. Supervaluation-style truth without supervaluations. *Journal of Philosophical Logic*, 47:817–850, 2018.

[33] 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.

[34] Sergei Tupailo. Realization of analysis into Explicit Mathematics. *The journal of symbolic logic*, 66(4):1848–1864, 2001.

[35] 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.