INSTITUT FÜR INFORMATIK
Lehr- und Forschungseinheit für
Programmier- und Modellierungssprachen
Oettingenstraße 67, D–80538 München

Truth in applicative theories

Reinhard Kahle

appears in *Synthese*
http://www.pms.informatik.uni-muenchen.de/publikationen
Forschungsbericht/Research Report PMS-FB-2001-4, February 2001

Truth in applicative theories

Reinhard Kahle

Abstract

We give a survey on truth theories for applicative theories. It comprises Frege structures, Universes for Frege structures, and a theory of supervaluation. We present the proof-theoretic results for these theories and show their syntactical expressive power. In particular, we present as a novelty a syntactical interpretation of ID_1 in a applicative truth theory based on supervaluation.

1 Introduction

Applicative theories build the first order part of Feferman's systems of explicit mathematics, [Fef75, Fef79]. They comprise *type-free* combinatory logic, natural numbers, pairing and projection. A survey about the proof-theoretic results in the applicative framework can be found in [JKS99].

In [Bee85], Beeson introduced a truth theory for applicative theories by adding a truth predicate T and appropriate new constants. He showed that such a truth theory is an axiomatic counterpart of Aczel's *Frege structures*. These were introduced semantically by Aczel to define a notion of set by means of a partial truth predicate, [Acz80]. In his monograph [Can96] Cantini investigates truth theories on an applicative basis in a comprehensive way.

In this paper we present the main features of truth theories over applicative theories, namely "no Gödelization" and "abstraction". Moreover, we will summarize the proof-theoretic results for these theories. We put special emphasis on the *syntactical expressiveness* which is reflected in embeddings of fixed-point theories and theories of inductive definitions. This will be described here not only for Frege structures, but also for an extension by *universes* which capture the idea of iterated truth predicates, cf. [Kah97a, Kah0xb]. In addition, we refine a result of Cantini [Can96] by giving a *syntactical* embedding of the theory ID_1 in the theory SON, a truth theory based on *supervaluation*.

The structure of the paper is as follows. In the following section, we give a sketch of the background of truth theories in applicative theories. Then, we introduce the applicative theory TON. In section 4, we define Frege structures as a syntactical theory and restate the main results as they can be found in [Can96]. In the presentation, we essentially follow [Kah0xb]. This paper serves also as a reference for the next section about universes over Frege structures. In section 6, we discuss the theory of supervaluation as introduced in [Can96] and publish, for the first time, the embedding of ID_1, [Kah97a]. We finish the paper by giving concluding remarks, including an overview of the proof-theoretic results.

2 Background

The background of truth theories for applicative theories is somehow two-folded: On the one hand, they originate in Aczel's Frege structures as their syntactical counterpart. On the other hand, they come up as a very convenient alternative to ordinary truth theories over Peano arithmetic.

Going back to the prior work of Scott [Sco75], Aczel has introduced Frege structures as a semantical concept to introduce *sets* in terms of a *partial truth predicate*, [Acz80]. Considering Frege's well-known, inconsistent formal system *Grundgesetze der Arithmetik*, usually, the unrestricted comprehension scheme is held responsible for the inconsistency. Thus, in most of the later formal approaches, the comprehension scheme is restricted. In contrast, Aczel still allows unrestricted comprehension, but the element relation, which is defined in terms of the truth predicate, becomes *partial*. In the formal definition, he defines *proposition* and *truth* simultaneously as fixed points of appropriate operator forms.

In his monograph [Bee85], Beeson gave a syntactic characterization of Frege structures based on applicative theories. Adding both predicates for propositions and truth, one can give an axiomatization of Frege structures. He showed that the resulting theory has the proof-theoretic strength of Feferman's theory $\mathsf{EM}_0 + \mathsf{J}$, which is equivalent with the fixed point theory $\widehat{\mathsf{ID}}_1$. Flagg and Myhill have investigated Frege structures in a more general perspective. In particular, they give up the primitive notion of proposition, but define it in terms of truth, [FM87a, FM87b].

Cantini has studied the relation of this concept of truth and abstraction in a very comprehensive way, [Can96]. His book serves as a general reference for results, methods, and extensions in this field. An exposition of the central results for Frege structures as presented here can already be found in his article [Can93]. In particular, he showed that the fixed point theory $\widehat{\mathsf{ID}}_1$ can be embedded into a truth theory over applicative theories. In general, based on a defined notion of sets, Cantini can deal with (total versions of) theories of explicit mathematics, like the theory EET of Feferman and Jäger [FJ96]. But the converse does not hold. In particular, it is not know whether it is possible to give a syntactical interpretation of fixed point theories in explicit mathematics.

As further investigation about Frege structures, we would like to mention the following: Jan Smith has even used a theory of Frege structures to give a semantics for Martin-Löf's type theory, [Smi78, Acz80, Smi84]. Hayashi and Kobayashi have introduced a version of Frege structures which is tuned up, in a way, to match up exactly with a version of Feferman's explicit mathematics [HK95].

Starting with Frege structures, there are two different ways to get stronger theories with respect to the proof-theoretic strength. On the one hand, we can define *levels of truth* as they are known from ordinary truth theories, see below. On the other hand, we can replace the truth axioms for Frege structures by truth axioms for *supervaluation*. It has a non compositional concept of truth, introduced by van Fraassen [vF68, vF70]. Cantini has defined such a theory over Peano Arithmetic with proof-theoretic strength of ID_1 in [Can90]. For applicative theories, he has given a corresponding theory in [Can96]. Here, we present his theory and give a syntactical interpretation of ID_1 in this theory, as was proven in [Kah97a].

Ordinary truth theories are discussed at length, both from a philosophical point of view, as well as from a mathematical one, following the seminal paper of Tarski [Tar35, Tar56]. In particular, Kripke's general analysis of truth in [Kri75] was highly influential. A very good overview of the combination of different properties of formalized proof predicates over Peano Arithmetic and its proof-theoretic implications can be found in Friedman and Sheard [FS82]. Facing the well-known danger of paradoxes when dealing with "negative truth", a standard escape is the *iteration* of proof predicates, which allows us to deal with negative truth of one level as positive truth in the next level, cf. [Tar35, Tar56, Fef91, Hal96].

In the context of applicative theories we take over the idea of iterated truth by introducing *universes*. Universes were first introduced by Martin-Löf in his type theory [Mar84] as types closed under the usual type existence axioms. In a similar way they are investigated in the context of explicit mathematics, [Fef82, Mar93, Mar94, Kah97b, Str99, JKS0x]. Here, we introduce universes as classes or propositional functions closed under the truth conditions. This approach is a modification of Cantini's theory TLR [Can95, Can96] and worked out in [Kah0xb].

We finish this section by defining the reference theories for the proof-theoretic investigations, namely *fixed point theories* and *theories of inductive definition*.

The well-known fixed point theory $\widehat{\mathsf{ID}}_1$ extends Peano arithmetic by (not necessarily least) fixed points $\mathcal{P}^\varphi$ of P-positive arithmetical operator forms $\varphi(P, x)$, i.e., an arithmetical formula in which the subformula $P(t)$ occurs only positively [Acz77]:

Definition 1 The language $\mathcal{L}_{\mathsf{ID}}$ of $\widehat{\mathsf{ID}}_1$ is the language of PA extended by new fixed point constants $\mathcal{P}^{\varphi}$ for each P-positive arithmetical operator form $\varphi(P, x)$.

The axioms of $\widehat{\mathsf{ID}}_1$ are those of PA extended to the new language, plus the following fixed point axiom for each fixed point constants $\mathcal{P}^{\varphi}$:

$$\forall x.\varphi(\mathcal{P}^{\varphi}, x) \leftrightarrow \mathcal{P}^{\varphi}(x).$$

In particular, induction on the natural numbers is available for all formulae of the language $\mathcal{L}_{\mathsf{ID}}$.

In addition, we will consider the theory $\mathsf{ID}_1^{\#}$, which is essentially $\widehat{\mathsf{ID}}_1$ but induction is restricted to formulae in which the new fixed point constants occur only positively. For the proof-theoretic analysis of $\mathsf{ID}_1^{\#}$, we refer to [JS96].

With respect to the proof-theoretic strength, we get, in terms of the binary Veblen function $\boldsymbol{\varphi}$:

Theorem 2
1. $|\mathsf{ID}_1^{\#}| = \boldsymbol{\varphi}\omega 0$,
2. $|\widehat{\mathsf{ID}}_1| = \boldsymbol{\varphi}\varepsilon_0 0$.

When considering *universes* over Frege structures, we need *iterated fixed point theories*. *Finitely* iterated fixed point theories were introduced and studied by Feferman in [Fef82]. The *transfinite* case is analyzed by Jäger et. al. in [JKSS99].

Here, we consider the theories $\widehat{\mathsf{ID}}_\alpha$ up to $\alpha < \varepsilon_0$. They are formulated in a language which expands the language of Peano arithmetic by predicate constants $\mathcal{P}^{\varphi}$ for each inductive operator form $\varphi(P, Q, x, y)$, i.e., a formula of the language of PA, containing $P(t)$ at most positively, while $Q(s)$ is allowed to occur positively and negatively. For the formal definition, we need a primitive recursive pairing operation $\langle\cdot,\cdot\rangle$ with projections $(\cdot)_0$ and $(\cdot)_1$. Then, we write $\mathcal{P}^{\varphi}_s(t)$ for $\mathcal{P}^{\varphi}(\langle t, s\rangle)$ and $\mathcal{P}^{\varphi}_{\prec s}(t)$ for $t = \langle (t)_0, (t)_1\rangle \wedge (t)_1 \prec s \wedge \mathcal{P}^{\varphi}(t)$. Here, $\prec$ denotes a primitive recursive well-ordering of order type ε_0. One can understand the parameter s in $\mathcal{P}^{\varphi}_s(t)$ as the level of the fixed point definition. With $\mathcal{P}^{\varphi}_{\prec s}(t)$, it is expressed that t belongs to the disjoint union of fixed points with levels less than s.

Definition 3 The axiom schemas of $\widehat{\mathsf{ID}}_\alpha$, for α an ordinal less than ε_0, are those of PA, together with induction on the natural numbers for the extended language, plus the following fixed point axioms for each inductive operator form $\varphi(P, Q, x, y)$:

$$\forall \beta \prec \alpha. \forall x. \mathcal{P}^{\varphi}_\beta(x) \leftrightarrow \varphi(\mathcal{P}^{\varphi}_\beta, \mathcal{P}^{\varphi}_{\prec\beta}, x, \beta).$$

$\widehat{\mathsf{ID}}_{<\alpha}$ is the union of the theories $\widehat{\mathsf{ID}}_\beta$, $\beta < \alpha \leq \varepsilon_0$.

The theories $\widehat{\mathsf{ID}}_\alpha$ are defined as the *metapredicative* counterparts to the well-known impredicative theories of inductive definitions ID_α.

The proof-theoretic analysis of $\widehat{\mathsf{ID}}_\alpha$ in [JKSS99] yields the following theorem about the proof-theoretic strength, in terms of the ternary Veblen function $\boldsymbol{\varphi}$. The ternary Veblen function is the generalization of the well-known binary Veblen function, cf. [JKSS99]. In particular, $\boldsymbol{\varphi}100$ is the Feferman-Schütte ordinal Γ_0. The first assertion of the theorem was proven by Feferman in [Fef82].

Theorem 4

1. $|\widehat{\mathsf{ID}}_{<\omega}| = \boldsymbol{\varphi}100 = \Gamma_0$,
2. $|\widehat{\mathsf{ID}}_{<\omega^\omega}| = \boldsymbol{\varphi}1\omega0$,
3. $|\widehat{\mathsf{ID}}_{<\varepsilon_0}| = \boldsymbol{\varphi}1\varepsilon_0 0$.

For the study of *supervaluation*, we need, in addition, the theory ID_1. It is formulated in the same language as $\widehat{\mathsf{ID}}_1$, but now we axiomatize not only fixed points, but *least fixed points*.

Definition 5 The axioms of ID_1 are those of PA extended to the new language, plus the following two axioms for each fixed point constants $\mathcal{P}^\varphi$:

$$\forall x.\varphi(\mathcal{P}^\varphi, x) \to \mathcal{P}^\varphi(x),$$
$$(\forall y.\varphi(\psi, y) \to \psi(y)) \to \forall x.\mathcal{P}^\varphi(x) \to \psi(x).$$

The proof-theoretic ordinal of ID_1 is the Bachmann-Howard ordinal. In terms of the notation system used in [Poh98], we have:

Theorem 6 $|\mathsf{ID}_1| = \Psi_\Omega(\varepsilon_{\Omega+1})$.

3 The theory TON

The theory TON (total theory of operations and numbers) is the total version of BON (basic theory of operations and numbers), which is the first order part of Feferman's theories of explicit mathematics, [JS95]. While BON is based on Beeson's *logic of partial terms*, in TON, we work on the basis of classical predicate logic with equality. In BON, we have a special existence predicate $t\downarrow$ expressing "t is defined" or "t has a value", and the quantifiers range over defined objects only. Additionally, we have strictness, expressing that the subterms of defined terms are defined and that all arguments of predicates are defined, [Bee85, FJ93]. For our truth theory, we have chosen the total version of applicative theories because of a problem with the truth definition for negated existence. This problem, and possibilities to solve it, are discussed at length in [Kah99].

TON is formulated in the language $\mathcal{L}_t$ which comprises:

- individual variables: $x, y, z, u, v, w, f, g, h, \ldots$
- individual constants
  - k, s (combinators),
  - p, p_0, p_1 (pairing and projection),
  - 0, $\mathsf{s_N}$, $\mathsf{p_N}$ (zero, successor and predecessor),
  - $\mathsf{d_N}$ (definition by cases),
- a binary function symbol $\cdot$ for term application, and
- the relation symbols $=$ and N.

Terms $(r, s, t, \ldots)$ are built up from individual variables and individual constants by term application.

Formulae $(\varphi, \psi, \ldots)$ are built from the atomic formulae $t = s$ and $\mathsf{N}(t)$ by closure under negation $(\neg\varphi)$, conjunction $(\varphi \wedge \psi)$ and univeral quantification $(\forall x.\varphi)$.

We use the following conventions: $s\,t$ stands for $(s \cdot t)$ with association to the left. The connectives $\vee$, $\rightarrow$ and $\exists$ are defined as usual. We write $t \neq s$ for $\neg(t = s)$ and quantifiers which are restricted to elements of N are written in the form of $\forall x : \mathsf{N}.\varphi$.

The logic of TON is first-order predicate logic with equality. The non-logical axioms comprise:

I. **Combinatory algebra.**

(1) $\mathsf{k}\,x\,y = x$,

(2) $\mathsf{s}\,x\,y\,z = x\,z\,(y\,z)$,

II. **Pairing and projection.**

(3) $\mathsf{p}_0\,(\mathsf{p}\,x\,y) = x \wedge \mathsf{p}_1\,(\mathsf{p}\,x\,y) = y$,

III. **Natural numbers.**

(4) $\mathsf{N}(0) \wedge \forall x.\mathsf{N}(x) \rightarrow \mathsf{N}(\mathsf{s_N}\,x)$,

(5) $\forall x.\mathsf{N}(x) \rightarrow \mathsf{s_N}\,x \neq 0 \wedge \mathsf{p_N}\,(\mathsf{s_N}\,x) = x$,

(6) $\forall x.\mathsf{N}(x) \wedge x \neq 0 \rightarrow \mathsf{N}(\mathsf{p_N}\,x) \wedge \mathsf{s_N}\,(\mathsf{p_N}\,x) = x$.

IV. **Definition by cases on N.**

(7) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v = w \rightarrow \mathsf{d_N}\,x\,y\,v\,w = x$,

(8) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v \neq w \rightarrow \mathsf{d_N}\,x\,y\,v\,w = y$.

λ abstraction can be introduced in the standard way. Also, because of self-application, we can define a recursion operator rec in TON.

Proposition 7 1. For every variable x and every term t of $\mathcal{L}_t$, there exists a term $\lambda x.t$ of $\mathcal{L}_t$ whose free variables are those of t, excluding x, such that TON proves $(\lambda x.t)\, x = t$.

2. There exists a term rec of $\mathcal{L}_t$ such that TON proves $\forall x.\mathsf{rec}\, x = x\,(\mathsf{rec}\, x)$.

Of course, this proposition also holds for extensions of the language $\mathcal{L}_t$, as we will consider in the following sections.

The theory TON can be shown to be proof-theoretically equivalent with Peano arithmetic PA if we add the schema of formulae induction on N for arbitrary $\mathcal{L}_t$ formulae φ:

Formulae induction on N ($\mathcal{L}_t$-I_{N})

$$\varphi(0) \wedge (\forall x : \mathsf{N}.\varphi(x) \rightarrow \varphi(\mathsf{s_N}\, x)) \rightarrow \forall x : \mathsf{N}.\varphi(x).$$

The lower bound follows from a straightforward interpretation of PA in $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathrm{I}_{\mathsf{N}})$, where the natural numbers are interpreted as elements of N, [JS95]. Models of TON can be found in Beeson [Bee85], Cantini [Can96] and Strahm [JS95, Str96]. To some extent, we can consider the *closed total term model* $\mathcal{CTT}$ as a standard model for TON. Roughly, it can be described as follows: As the universe of $\mathcal{CTT}$, we choose the set of all closed terms of the language $\mathcal{L}_t$, i.e., we interpret the constants by themselves and application by juxtaposition. By use of a straightforward reduction relation ϱ for the constants of $\mathcal{L}_t$, we can interpret equality of terms by the fact that they have a common reduct with respect to *arbitrary* reductions on the basis of ϱ. For the verification of transitivity of equality, we need to prove the *Church-Rosser property*. Finally, $\mathsf{N}(t)$ holds if t reduces to a numeral. A detailed description of $\mathcal{CTT}$, including the proof of the Church-Rosser property, can be found in [JS95, Can96]. Since this model can be formalized in PA, we get the proof-theoretic equivalence (for the notion of proof-theoretic equivalence we refer to [Fef88, Fef00]):

Theorem 8 $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{PA}$.

4 Frege structures

For the definition of Frege structures, we extend the language $\mathcal{L}_t$ to the new language $\mathcal{L}_F$ by adding a new predicate T for truth and new individual constants $\dot{=}$, $\dot{\mathsf{N}}$, $\dot{\neg}$, $\dot{\wedge}$, and $\dot{\forall}$ for the representation of formulae by terms.

For the sake of readability, we will freely use an infix notation for terms containing the dotted constants. For instance, $x \mathbin{\dot{=}} y$ has to be written formally as $\dot{=}\, x\, y$.

The theory FON (Frege structures over TON) consists of the axioms of TON extended to the expanded language plus the following ones:

I. **Closure under prime formulae of TON**

(1) $x = y \leftrightarrow \mathsf{T}(x \mathbin{\dot=} y)$,

(2) $\neg x = y \leftrightarrow \mathsf{T}(\dot\neg\,(x \mathbin{\dot=} y))$,

(3) $\mathsf{N}(x) \leftrightarrow \mathsf{T}(\dot{\mathsf{N}}\,x)$,

(4) $\neg\mathsf{N}(x) \leftrightarrow \mathsf{T}(\dot\neg\,(\dot{\mathsf{N}}\,x))$.

II. **Closure under composed formulae**

(5) $\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot\neg\,(\dot\neg\,x))$,

(6) $\mathsf{T}(x) \wedge \mathsf{T}(y) \leftrightarrow \mathsf{T}(x \mathbin{\dot\wedge} y)$,

(7) $\mathsf{T}(\dot\neg\,x) \vee \mathsf{T}(\dot\neg\,y) \leftrightarrow \mathsf{T}(\dot\neg\,(x \mathbin{\dot\wedge} y))$,

(8) $(\forall x.\mathsf{T}(f\,x)) \leftrightarrow \mathsf{T}(\dot\forall\,f)$,

(9) $(\exists x.\mathsf{T}(\dot\neg\,(f\,x))) \leftrightarrow \mathsf{T}(\dot\neg\,(\dot\forall\,f))$.

III. **Consistency**

(10) $\forall x.\neg(\mathsf{T}(x) \wedge \mathsf{T}(\dot\neg\,x))$.

We will use the following abbreviations:

$$\begin{aligned} \mathsf{F}(t) &:\Leftrightarrow \mathsf{T}(\dot\neg\,t), \\ \mathsf{P}(t) &:\Leftrightarrow \mathsf{T}(t) \vee \mathsf{F}(t), \\ \mathsf{C}(t) &:\Leftrightarrow \forall x.\mathsf{P}(t\,x). \end{aligned}$$

$\mathsf{T}(t)$ can be read as "t is true", $\mathsf{F}(t)$ as "t is false", $\mathsf{P}(t)$ as "t is a proposition", and $\mathsf{C}(t)$ as "t is a class". In the literature, $\mathsf{C}(t)$ is often characterized as a *propositional function*.

By diagonalizing $\dot\neg$, we get that the truth predicate is partial:

Lemma 9 $\mathsf{FON} \vdash \neg\forall x.\mathsf{P}(x)$.

Proof: If we set $r := \mathsf{rec}\,(\lambda x.\dot\neg\,x)$, we have that $\mathsf{T}(r)$ is equivalent to $\mathsf{T}(\dot\neg\,r)$, i.e., $\mathsf{F}(r)$, which contradicts the axiom of consistency III.(10). □

This lemma shows that we cannot have a truth definition for the whole language. In particular, we cannot add full self-reference. That means that we are not allowed to add a term $\dot{\mathsf{T}}$ such that $\neg\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot\neg\,(\dot{\mathsf{T}}\,x))$. But we can introduce a rather trivial version of self-reference just by defining $\dot{\mathsf{T}}$ as the identity function $\lambda x.x$. With this, we get the following clauses:

Self-reference

$\mathsf{T}(x) \leftrightarrow \mathsf{T}(\dot{\mathsf{T}}\,x)$,

$\mathsf{T}(\dot\neg\,x) \leftrightarrow \mathsf{T}(\dot\neg\,(\dot{\mathsf{T}}\,x))$, In the context of Frege structures, we can consider at least three natural forms of induction over N.

1. Class induction on N (C-I_{N})

$$\mathsf{C}(f) \wedge \mathsf{T}(f\,0) \wedge (\forall x : \mathsf{N}.\mathsf{T}(f\,x) \to \mathsf{T}(f\,(\mathsf{s_N}\,x))) \to \forall x : \mathsf{N}.\mathsf{T}(f\,x).$$

2. Truth induction on N (T-I_{N})

$$\mathsf{T}(f\,0) \wedge (\forall x : \mathsf{N}.\mathsf{T}(f\,x) \to \mathsf{T}(f\,(\mathsf{s_N}\,x))) \to \forall x : \mathsf{N}.\mathsf{T}(f\,x).$$

3. Formulae induction on N ($\mathcal{L}_F$-I_{N})

$$\varphi(0) \wedge (\forall x : \mathsf{N}.\varphi(x) \to \varphi(\mathsf{s_N}\,x)) \to \forall x : \mathsf{N}.\varphi(x)$$

for arbitrary $\mathcal{L}_F$ formulae φ.

Our theories FON, $\mathsf{FON}+(\mathsf{C}$-$\mathrm{I}_{\mathsf{N}})$, $\mathsf{FON}+(\mathsf{T}$-$\mathrm{I}_{\mathsf{N}})$, and $\mathsf{FON}+(\mathcal{L}_F$-$\mathrm{I}_{\mathsf{N}})$ are essentially equivalent with the theories MF^-, MF_c, MF_p, and MF, respectively, of Cantini in [Can96].

For the proof-theoretic analysis given in [Can96], we have:

Theorem 10 1. $\mathsf{FON} + (\mathsf{C}$-$\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{PA}$,

2. $\mathsf{FON} + (\mathsf{T}$-$\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{ID}_1^{\#}$,

3. $\mathsf{FON} + (\mathcal{L}_F$-$\mathrm{I}_{\mathsf{N}}) \equiv \widehat{\mathsf{ID}}_1$.

Flagg and Myhill [FM87b] give a general model construction for Frege structures by defining truth as a fixed point satisfying the corresponding closure conditions (cf. also the original definition in [Acz80]). These models can be directly formalized in ID_1, but the leastness condition is essential to verify the axiom of consistency. In contrast, Cantini proves the upper bounds of FON by pure proof-theoretic methods.

In the following, we will give the essential features of Frege structures, namely, representation of formulae by terms without Gödelization, abstraction and the syntactical embedding of $\widehat{\mathsf{ID}}_1$.

In a straightforward way, we can associate a term with every formula in FON.

Definition 11 By induction of the build up of an $\mathcal{L}_F$ formula, we define:

$$\begin{aligned} \overbrace{t = s}^{\cdot} &\equiv t \doteq s \\ \overbrace{\mathsf{N}(t)}^{\cdot} &\equiv \dot{\mathsf{N}}\,t \end{aligned}$$

$$
\begin{aligned}
\overbrace{\mathsf{T}(t)}^{\cdot} &\equiv \dot{\mathsf{T}}\,t \equiv t \\
\overbrace{\neg\varphi}^{\cdot} &\equiv \dot{\neg}\,\dot{\varphi} \\
\overbrace{\varphi \wedge \psi}^{\cdot} &\equiv \dot{\varphi}\,\dot{\wedge}\,\dot{\psi} \\
\overbrace{\forall x.\varphi}^{\cdot} &\equiv \dot{\forall}\,(\lambda x.\dot{\varphi})
\end{aligned}
$$

Since we work in classical logic, we can introduce abbreviations for the other connectives on the term level in the usual way:

$(t\,\dot{\vee}\,s) := \dot{\neg}\,(\dot{\neg}\,t\,\dot{\wedge}\,\dot{\neg}\,s)$, $(t\,\dot{\rightarrow}\,s) := \dot{\neg}\,t\,\dot{\vee}\,s$, and $(\dot{\exists}x.t) := \dot{\neg}\,(\dot{\forall}x.\dot{\neg}\,t)$.

In the following, we have to pay special attention to formulae in which the truth predicate occurs only positively. The formal definition of T-positiveness is given simultaneously with T-negativeness:

Definition 12

1. $t = s$, $\mathsf{N}(t)$, $\neg t = s$ and $\neg\mathsf{N}(t)$ are T-positive as well as T-negative.
2. $\mathsf{T}(t)$ is T-positive; $\neg\mathsf{T}(t)$ is T-negative.
3. If φ is T-positive (T-negative), then $\neg\varphi$ is T-negative (T-positive).
4. If φ and ψ are T-positive (T-negative), then so is $\varphi \wedge \psi$.
5. If φ is T-positive (T-negative), then so is $\forall x.\varphi$.

Now, we can prove by straightforward induction that FON provides a truth definition for T-positive formulae [Can96, Th. 8.8.]:

Proposition 13 If φ is a T-positive formulae of $\mathcal{L}_F$, then we have:

$$\mathsf{FON} \vdash \mathsf{T}(\dot{\varphi}) \leftrightarrow \varphi.$$

This proposition is the essential tool to introduce a notion of *set* and an *element relation* on the basis of the truth predicate, cf. [Sco75].

Definition 14 Given two $\mathcal{L}_F$ terms t and s and an $\mathcal{L}_F$ formula φ, we define:

$$
\begin{aligned}
\{x|\varphi\} &:= \lambda x.\dot{\varphi}, \\
t \in s &:\Leftrightarrow \mathsf{T}(s\,t).
\end{aligned}
$$

Since the term $\{x|\varphi\}$ is defined for arbitrary formulae φ, we can say that Frege structures allow full or unrestricted comprehension. But it is clear that the element relation has its intended meaning for T-positive formulae only. From proposition 13, we get as a corollary:

Corollary 15 If φ is a T-positive formulae of $\mathcal{L}_F$, then we have:

$$\mathsf{FON} \vdash x \in \{x|\varphi\} \leftrightarrow \varphi.$$

In analogy to $\widehat{\mathsf{ID}}_1$, we define the notion of positive operator form in FON as a T-positive formulae $\varphi(R,x)$ which contains the (unary) relation variable R only positively. Now, we can use proposition 13 to define *fixed points* of such operator forms within FON, cf. [Can96]. The proof makes essential use of the possibility to define fixed points on the term level.

Proposition 16 Let $\varphi(R,x)$ be a positive operator form. Then there exists a term t_φ of $\mathcal{L}_F$ such that

$$\mathsf{FON} \vdash \forall x.\mathsf{T}(t_\varphi\, x) \leftrightarrow \varphi(\mathsf{T}(t_\varphi\, \cdot), x).$$

Proof: We define $t_\varphi := \mathsf{rec}\,(\lambda y, x.\overbrace{\varphi(\mathsf{T}(y\,\cdot), x)}^{\cdot})$. Thus, we get with the recursion theorem and the T-positiveness of φ:

$$\begin{aligned} \mathsf{T}(t_\varphi\, x) &\leftrightarrow \mathsf{T}(\overbrace{\varphi(\mathsf{T}(t_\varphi\,\cdot), x)}^{\cdot}) \\ &\leftrightarrow \varphi(\mathsf{T}(t_\varphi\,\cdot), x). \end{aligned}$$

□

By proposition 13, we can reduce induction for T free formulae in FON to (C-I_N). Thus, the standard translation $\cdot^N$ of Peano Arithmetic in $\mathsf{TON} +$ ($\mathcal{L}_t$-I_N) [JS95] carries over to $\mathsf{FON} + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N})$. Using the previous proposition, we can easily extend this interpretation to the fixed points constants of $\mathsf{ID}_1^\#$ and $\widehat{\mathsf{ID}}_1$. So we get:

Proposition 17 There exists a translation $\cdot^N$ from the language of PA or $\mathcal{L}_{\mathsf{ID}}$, respectively, into the language $\mathcal{L}_F$ such that

1. $\mathsf{PA} \vdash \varphi \quad \Rightarrow \quad \mathsf{FON} + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
2. $\mathsf{ID}_1^\# \vdash \varphi \quad \Rightarrow \quad \mathsf{FON} + (\mathsf{T}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
3. $\widehat{\mathsf{ID}}_1 \vdash \varphi \quad \Rightarrow \quad \mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$.

Please note that the recursion theorem in applicative theories does not help to define *least fixed points* (for a discussion of possibilities of least fixed point operators in the applicative framework, we refer to [KS0x]). For this reason, there is no possibility to define a truth theory based on applicative theories which allows an embedding of ID_1 in the same manner. However, later, we will define a truth theory based on supervaluation which allows a syntactical embedding of ID_1.

5 Universes over Frege structures

The concept of *universes* goes back to Martin-Löf who introduced it in his type theory, cf. [Mar84]. In the framework of explicit mathematics, it was first studied by Feferman, [Fef82]. The introduction of universes in explicit mathematics by use of a (non-uniform) *limit axiom* goes back to Marzetta [Mar93, Mar94]. The uniform version is studied in [Kah97b, Str99, JKS0x]. Here, we discuss a notion of universes over Frege structures. On the one hand, it is motivated by the concept in explicit mathematics, and on the other hand, it captures the idea of iterated truth predicates. From the latter point of view, it can be seen as an adaptation of Cantini's theory TLR of *reflective truth with levels* [Can96] which allows a *uniform* definition of truth levels *within* the applicative framework.

As for FON, the syntactical expressiveness of theories of universes over Frege structures can be shown by a syntactical embedding of *(transfinitely) iterated fixed point theories* $\widehat{\mathsf{ID}}_\alpha$.

For the definition of universes we need an *ordering relation* which allows us to reflect the usual order relation of truth predicates on the term level.

Definition 18

$$t \sqsubset s \quad :\Leftrightarrow \quad \forall x.(\mathsf{T}(t\,x) \to \mathsf{T}(s\,(t\,x))) \land (\mathsf{T}(\dot{\neg}\,(t\,x)) \to \mathsf{T}(s\,(\dot{\neg}\,(t\,x)))).$$

For $t \sqsubset s$, we can say that s reflects the *truth-course-of-value* of t, i.e., in short: *s reflects t*.

As intended, $t \sqsubset s$ allows us to handle *negative* statements with respect to truth at level t as *positive* statements at level s.

We formulate the theory FSU (Frege structures with universes) in the language $\mathcal{L}_U$, which expands $\mathcal{L}_F$ by the additional relation symbol U and the additional individual constant ℓ.

As axioms of FSU, we have those of TON extended to the expanded language and the following:

I. **Basic axioms**

(1) $\mathsf{U}(u) \to \mathsf{C}(u)$,

(2) $\mathsf{U}(u) \to \forall x.\mathsf{T}(u\,x) \to \mathsf{T}(x)$.

II. **Closure under prime formulae of TON**

(3) $\mathsf{U}(u) \to \forall x, y.x = y \leftrightarrow \mathsf{T}(u\,(x \doteq y))$,

(4) $\mathsf{U}(u) \to \forall x, y.x \neq y \leftrightarrow \mathsf{T}(u\,(\dot{\neg}\,(x \doteq y)))$,

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

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

III. **Closure under composed formulae**

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

(8) $\mathsf{U}(u) \to \forall x, y.\mathsf{T}(u\,x) \wedge \mathsf{T}(u\,y) \leftrightarrow \mathsf{T}(u\,(x\,\dot{\wedge}\,y))$,

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

(10) $\mathsf{U}(u) \to ((\forall x.\mathsf{T}(u\,(f\,x))) \leftrightarrow \mathsf{T}(u\,(\dot{\forall}\,f)))$,

(11) $\mathsf{U}(u) \to ((\exists x.\mathsf{T}(u\,(\dot{\neg}\,(f\,x)))) \leftrightarrow \mathsf{T}(u\,(\dot{\neg}\,(\dot{\forall}\,f))))$.

IV. **Order structure**

(12) $\mathsf{U}(u) \wedge \mathsf{U}(v) \wedge \mathsf{T}(v\,(u\,t)) \to \mathsf{T}(v\,t)$,

V. **Local consistency**

(13) $\mathsf{U}(u) \to \forall x.\neg(x \in u \wedge \dot{\neg}\,x \in u)$,

VI. **Limit axiom**

(14) $\forall f.\mathsf{C}(f) \to \mathsf{U}(\ell\,f) \wedge f \sqsubset \ell\,f$.

Remark 19 With respect to our notion of set, it is maybe more convenient to read the closure conditions in terms of the element relation. For instance, we have

I.(2) $\mathsf{U}(u) \to \forall x.x \in u \to \mathsf{T}(x)$,

III.(7) $\mathsf{U}(u) \to \forall x.x \in u \leftrightarrow \dot{\neg}\,(\dot{\neg}\,x) \in u$,

IV.(12) $\mathsf{U}(u) \wedge \mathsf{U}(v) \wedge \overbrace{(t \in u)}^{\cdot} \in v \to t \in v$.

This reading shows the "Janus face" of the truth predicates which provides us with an element relation, too. From this point of view, the closure conditions can be considered as set theoretical ones.

The axioms express that universes are classes, collecting true elements only, and that they are closed under the usual truth conditions. The theory is formulated over TON and not FON, but the axioms of FON are derivable in FSU. Thus, the closure conditions also hold for the "global" truth predicate. First, we show that global truth is equivalent with truth in a universe:

Lemma 20 $\mathsf{FSU} \vdash \mathsf{T}(x) \leftrightarrow \exists u.\mathsf{U}(u) \wedge \mathsf{T}(u\,x)$.

Proof: The direction from the right to the left is axiom I.(2). For the other direction, it follows from $\mathsf{T}(x)$ that the constant function $\lambda y.x$ is a class. By applying the limit axiom, we get $\exists u.\mathsf{U}(u) \wedge \lambda y.x \sqsubset u$. The definition of $\sqsubset$ yields $\forall z.\mathsf{T}((\lambda y.x)z) \to \mathsf{T}(u\,((\lambda y.x)z))$. So we have $\mathsf{T}(x) \to \mathsf{T}(u\,x)$ and, from the assumption, we get $\mathsf{T}(u\,x)$. □

From the order relation axiom, it follows that universes are inclusive and linearly ordered with respect to the defined order relation, cf. [Kah0xb].

Lemma 21

1. $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge u \sqsubset v \wedge x \in u \to x \in v$,
2. $\mathsf{FSU} \vdash \mathsf{U}(u) \to \neg u \sqsubset u$,
3. $\mathsf{FSU} \vdash \mathsf{U}(u) \wedge \mathsf{U}(v) \wedge \mathsf{U}(w) \wedge u \sqsubset v \wedge v \sqsubset w \to u \sqsubset w$.

As a corollary, we can prove the axioms of FON in FSU. The proof is straightforward by induction of the length of the derivation of φ in FON and by use of lemma 20. Only the verification of closure for universal quantification needs an extra application of the limit axiom in the direction from left to right. The axiom of consistency follows from local consistency, cf. [Can96, lemma 37.7.(i)].

Corollary 22 $\mathsf{FON} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} \vdash \varphi$.

The existence of universes is guaranteed by the *limit axiom*: Each class is reflected by a universe. In particular, these universes are built *uniformly*. This uniformity allows us to build a hierarchy of universes into the transfinite case, depending on the induction principle. As for FON, we consider class induction (C-I_{N}), truth induction (T-I_{N}) and formulae induction ($\mathcal{L}_U$-I_{N}) which is now, of course, a scheme for arbitrary $\mathcal{L}_U$ formulae.

For the proof theory of FSU, we get the following equivalences, cf. [Kah0xb]:

Theorem 23

1. $\mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \widehat{\mathsf{ID}}_{<\omega}$,
2. $\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \widehat{\mathsf{ID}}_{<\omega^\omega}$,
3. $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \widehat{\mathsf{ID}}_{<\varepsilon_0}$.

For the syntactical embedding of the fixed point theories in FSU, we first introduce the notion of a *hierarchy of universes*. Therefore, we assume that we have a standard wellordering $\prec$ of ordertype ε_0 in FSU. We say that a theory $\mathcal{T}$ proves the existence of a hierarchy of universes of length α if there is term t such that:

$$\mathcal{T} \vdash \forall \beta \prec \alpha . \mathsf{U}(t\,\beta) \wedge \forall \gamma \prec \beta . t\,\gamma \sqsubset t\,\beta.$$

We say that $\mathcal{T}$ proves the existence of hierarchies of universes of length $< \alpha$ if it proves the existence of hierarchies of universes of length β for every β less than α.

In analogy with the usual proofs of transfinite induction in arithmetic [Sch77, Can89, JS96], we can show the following proposition in FSU. Here, $\mathrm{TI}(\alpha, \varphi)$ expresses transfinite induction up to α for the formula φ.

Proposition 24

1. For ordinals α less than ω^ω, we have:

$$\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_\mathsf{N}) \vdash \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot)).$$

2. For ordinals α less than ε_0 and each $\mathcal{L}_U$ formula φ, it holds that:

$$\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_\mathsf{N}) \vdash \mathrm{TI}(\alpha, \varphi).$$

By use of these induction principles, we can prove [Kah0xb]:

Proposition 25

1. $\mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N})$ proves the existence of hierarchies of universes of length $< \omega$.
2. $\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_\mathsf{N})$ proves the existence of hierarchies of universes of length $< \omega^\omega$.
3. $\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_\mathsf{N})$ proves the existence of hierarchies of universes of length $< \varepsilon_0$.

The proof uses the following universe operation, which is defined by recursion on the ordinals less then ε_0.

1. $\mathsf{univ}\, 0 = \ell\,(\lambda y.0 \dot{=} 0)$,
2. $\mathsf{univ}\,(\alpha + 1) = \ell\,(\mathsf{univ}\,\alpha)$,
3. $\mathsf{univ}\,\Lambda = \ell\,(\lambda y.\overbrace{\exists x \prec \Lambda.y \in \mathsf{univ}\,x}^{\cdot})$, Λ is a limit ordinal.

Its definition is similiar to that used for a corresponding theory in explicit mathematics, cf. [Str99].

Now, hierarchies of universes can be used to define iterated fixed points. As for $\widehat{\mathsf{ID}}_\alpha$, we call an $\mathcal{L}_U$ formula $\varphi(P, Q, x, y)$ an *inductive operator form* if it is T-positive and contains the relation variable P only positively (while Q is allowed to occur positively and negatively). Also, we will use the usual short hand notation for pairing and projection, i.e., (t, s) for $(\mathsf{p}\, t\, s)$, $(t)_0$ for $(\mathsf{p}_0\, t)$, and $(t)_1$ for $(\mathsf{p}_1\, t)$. Let $t_\beta\, s$ be an abbreviation for $t\,(s, \beta)$ and $t_{\prec\beta}\, r$ for $((t\, r) \dot{\in} \mathsf{univ}\,(r)_1) \dot{\wedge} ((r)_1 \dot{\prec} \beta)$.

Proposition 26 Let $\varphi(P, Q, x, y)$ be an inductive operator form. Then there exists a term t^φ of $\mathcal{L}_U$ such that

$$\mathsf{FSU} + \mathrm{TI}(\alpha, \mathsf{T}(f\,\cdot)) \vdash \forall\beta \prec \alpha.\forall x.\mathsf{T}(t^\varphi_\beta\, x) \leftrightarrow \varphi(\mathsf{T}(t^\varphi_\beta\,\cdot), \mathsf{T}(t^\varphi_{\prec\beta}\,\cdot), x, \beta).$$

The (rather technical) proof of this proposition, cf. [Kah0xb], makes essential use of the fact that the universe which plays the role for P can reflect the universes which play the role for Q.

By the use of propostions 24 – 26, we can easily extend the standard interpretation $\cdot^N$ of PA in TON to get the embedding of the iterated fixed point theories:

Proposition 27 There exists a translation $\cdot^N$ from the language of $\widehat{\mathsf{ID}}_\alpha$ into the language $\mathcal{L}_U$ such that

1. $\widehat{\mathsf{ID}}_{<\omega} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
2. $\widehat{\mathsf{ID}}_{<\omega^\omega} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$,
3. $\widehat{\mathsf{ID}}_{<\varepsilon_0} \vdash \varphi \quad \Rightarrow \quad \mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_\mathsf{N}) \vdash \varphi^N$.

All three embeddings preserve arithmetical sentences.

6 Supervaluation

In [Can96, Ch. 12], Cantini has introduced an *impredicative* truth theory for applicative theories. Here, the truth axioms of Frege structures are replaced by axioms for *supervaluation*. The idea of supervaluation is due to van Fraassen [vF68, vF70]. It expresses that formulae which follow by pure logic are true independently of the logical complexity and syntactical structure of their subformulae, cf. [Can90]. For example, we have $\mathsf{T}(\overbrace{\varphi \to \varphi}^{\cdot})$ for arbitrary formulae φ. It is easy to observe that this is, in general, not provable for T negative formulae in Frege structures.

In [Can90], Cantini has already presented a truth theory over Peano arithmetic which is proof-theoretically equivalent to ID_1. Here, we present his axiomatization of such a theory over applicative theories in [Can96], called VF_p. In both cases, Cantini shows the proof-theoretic lower bound by an embedding of $\mathsf{ID}_1(\mathrm{acc})$, the theory of *accessibility elementary inductive definitions*, [BFPS81]. The upper bound is shown by a provability interpretation in the theory KPu, a version of Kripke Platek set theory equivalent with ID_1 introduced and studied by Jäger [Jäg82]. After presenting the theory SON, we will refine Cantini's result by giving an syntactical interpretation of ID_1 instead of $\mathsf{ID}_1(\mathrm{acc})$, [Kah97a]. As for the embedding of $\widehat{\mathsf{ID}}_1$ in Frege structures, this result shows the special syntactical expressiveness of truth theories over applicative theories.

We define the theory SON (supervaluation over theories of operation and numbers), which corresponds to the theory VF^- of Cantini in [Can96, § 59]. The language $\mathcal{L}_\mathcal{S}$ of SON is the same as $\mathcal{L}_F$. As in definition 11, we can associate with every formula φ a term $\dot{\varphi}$.

The axioms of SON are those of TON extended to the new language, plus the following ones:

I. T-out

(1) $\mathsf{T}(\dot{\varphi}) \to \varphi$.

II. T-elem

(2) $x = y \to \mathsf{T}(x \dot{=} y)$,
(3) $\neg x = y \to \mathsf{T}(\dot{\neg}\,(x \dot{=} y))$,
(4) $\mathsf{N}(x) \to \mathsf{T}(\dot{\mathsf{N}}\,x)$,
(5) $\neg\mathsf{N}(x) \to \mathsf{T}(\dot{\neg}\,(\dot{\mathsf{N}}\,x))$.

III. T-imp

(6) $\mathsf{T}(\dot{\varphi} \dot{\to} \dot{\psi}) \to (\mathsf{T}(\dot{\varphi}) \to \mathsf{T}(\dot{\psi}))$.

IV. T-univ

(7) $(\forall x.\mathsf{T}(\dot{\varphi})) \to \mathsf{T}(\dot{\forall}\,(\lambda x.\dot{\varphi}))$.

V. T-log.

(8) $\mathsf{T}(\dot{\varphi})$, if φ is a *logical axiom*.

As for FON, we leave out (trivial) self-reference, but define $\dot{\mathsf{T}}$ as $\lambda x.x$.

By T-log, we have $\mathsf{T}(\dot{\varphi})$ as an axiom of SON whenever φ is a $\mathcal{L}_{\mathcal{S}}$ formula which is an instantiation of an axiom of predicate logic with equality. Moreover, in the following, we study $\mathcal{L}_{\mathcal{S}}$ formulae φ which are provable by use of pure logic only, i.e., by use of axioms and rules of predicate logic with equality. In this case, we write

$$\mathsf{PL} \vdash \varphi.$$

The following lemma collects the elementary facts which we need about SON, [Can96]. In particular, the first one, provable by induction on the length of the derivation, is crucial.

Lemma 28 1. For $\mathcal{L}_{\mathcal{S}}$ formulae φ for which $\mathsf{PL} \vdash \varphi$ holds, we have: $\mathsf{SON} \vdash \mathsf{T}(\dot{\varphi})$.

2. $\mathsf{SON} \vdash (\forall x.\mathsf{T}(\dot{\varphi})) \leftrightarrow \mathsf{T}(\dot{\forall}\,(\lambda x.\dot{\varphi}))$,

3. $\mathsf{SON} \vdash \mathsf{T}(\dot{\varphi}) \wedge \mathsf{T}(\dot{\psi}) \leftrightarrow \mathsf{T}(\dot{\varphi} \dot{\wedge} \dot{\psi})$,

4. $\mathsf{SON} \vdash (\exists x.\mathsf{T}(\varphi)) \to \mathsf{T}(\dot{\exists}\,(\lambda x.\dot{\varphi}))$,

5. $\mathsf{SON} \vdash \mathsf{T}(\dot{\varphi}) \lor \mathsf{T}(\dot{\psi}) \to \mathsf{T}(\dot{\varphi} \mathbin{\dot{\lor}} \dot{\psi})$.

Note that we have equivalences in the case of conjunction and universal quantification, while for disjunction and existential quantification only one direction holds.

Concerning induction principles, we are mainly interested in truth induction (T-I_{N}) because in SON, it is equivalent with the scheme of formulae induction ($\mathcal{L}_{\mathcal{S}}$-$\mathrm{I}_{\mathsf{N}}$), i.e., induction over N for arbitrary $\mathcal{L}_{\mathcal{S}}$ formulae.

Proposition 29 (T-I_{N}) and ($\mathcal{L}_{\mathcal{S}}$-$\mathrm{I}_{\mathsf{N}}$) are equivalent over SON.

Proof: Of course, we have to show the derivation of ($\mathcal{L}_{\mathcal{S}}$-$\mathrm{I}_{\mathsf{N}}$) from ($\mathsf{T}$-$\mathrm{I}_{\mathsf{N}}$) only. The proof is a part of the proof of theorem 59.6 in [Can96], and runs as follows: We set

$$\begin{aligned} \mathrm{Clos}_{\mathsf{N}}(\psi) \quad &:\Leftrightarrow \quad \psi(0) \land \forall x : \mathsf{N}.\psi(x) \to \psi(\mathsf{s}_{\mathsf{N}}\, x), \\ \psi'(x) \quad &:\Leftrightarrow \quad \mathrm{Clos}_{\mathsf{N}}(\psi) \to \psi(x). \end{aligned}$$

It follows that $\mathsf{PL} \vdash \mathrm{Clos}_{\mathsf{N}}(\psi')$. From this, we get by lemma 28.1 that SON proves $\mathsf{T}(\overset{\cdot}{\overbrace{\mathrm{Clos}_{\mathsf{N}}(\psi')}})$. Writing $\mathrm{Clos}_{\mathsf{N}}(\psi')$ out in full and applying lemma 28.3 yields the premise of (T-I_{N}) for ψ'. Thus, we get by this induction principle $\forall x : \mathsf{N}.\mathsf{T}(\overset{\cdot}{\overbrace{\psi'(x)}})$. The axiom T-out yields $\forall x : \mathsf{N}.\psi'(x)$ which is, writing $\varphi'(x)$ out in full, ($\mathcal{L}_{\mathcal{S}}$-$\mathrm{I}_{\mathsf{N}}$) for the formula ψ. Since ψ was arbitrary, the proof is finished. □

From [Can96], we get the proof-theoretic equivalence of ID_1 and $\mathsf{SON}+$ (T-I_{N}), which is essentially his theory VF_p. Together with the previous proposition, we have:

Theorem 30 $\mathsf{SON} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{SON} + (\mathcal{L}_{\mathcal{S}}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{ID}_1$.

For completeness, we will mention the maybe surprising fact that the restriction to class induction (C-I_{N}) will not exceed the proof-theoretic strength of Peano arithmetic, cf. [Can96].

Theorem 31 $\mathsf{SON} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}}) \equiv \mathsf{PA}$.

In addition to Cantini's result, we will show in the following that $\mathsf{SON}+$ (T-I_{N}) (and ($\mathcal{L}_{\mathcal{S}}$-$\mathrm{I}_{\mathsf{N}}$)) also admits a *syntactical embedding* of ID_1 itself. In contrast to $\mathsf{ID}_1(\mathrm{acc})$, the syntactic power of ID_1 is much bigger by allowing the definition of fixed points of arbitrary positive operator forms. The steps of the proof follow essentially Cantini's proof of the embedding of $\mathsf{ID}_1(\mathrm{acc})$ [Can96, theorem 59.4], which has to be generalized from binary relations to arbitrary positive operator forms.

Let the positive operator form $\varphi(R, x)$ be defined, as for FON, as a T-positive formulae which contains the relation variable R only positively. Then, we have:

Lemma 32 If $\varphi(R, x)$ is a positive operator form and $\psi(x)$ and $\chi(x)$ are arbitrary $\mathcal{L}_\mathcal{S}$-formula with the one free variable x, we have:

1. $\mathsf{PL} \vdash \varphi(\psi, x) \wedge (\forall x.\psi(x) \to \chi(x)) \to \varphi(\chi, x)$.
2. $\mathsf{SON} \vdash \varphi(\mathsf{T}(\overbrace{\psi(\cdot)}^{\cdot}), x) \to \mathsf{T}(\overbrace{\varphi(\psi, x)}^{\cdot})$.

The proof is straightforward by induction of the build up of φ and, for the second claim, by use of lemma 28.

In the following, we will use as abbreviations:

$$\begin{aligned} \mathrm{Clos}_\varphi(\psi) \quad &:\Leftrightarrow \quad \forall x.\varphi(\psi, x) \to \psi(x), \\ \psi'(x) \quad &:\Leftrightarrow \quad \mathrm{Clos}_\varphi(\psi) \to \psi(x). \end{aligned}$$

In close analogy to the proof of proposition 29, we can show the φ-closure of ψ' by use of pure logic only. From that, we get the φ-closure of $\mathsf{T}(\dot{\psi}')$ in SON.

Lemma 33 If $\varphi(R, x)$ is a positive operator form and $\psi(x)$ an arbitrary $\mathcal{L}_\mathcal{S}$-formula with the one free variable x, we have:

1. $\mathsf{PL} \vdash \mathrm{Clos}_\varphi(\psi'(\cdot))$.
2. $\mathsf{SON} \vdash \mathrm{Clos}_\varphi(\mathsf{T}(\overbrace{\psi'(\cdot)}^{\cdot}))$.

Proof: We have to show $\forall x.\varphi(\psi', x) \to \psi'(x)$ by use of pure logic only, i.e., $\mathsf{PL} \vdash \forall x.\varphi(\psi', x) \to (\mathrm{Clos}_\varphi(\psi') \to \psi(x))$.

Let us assume $\varphi(\psi', x)$ and $\mathrm{Clos}_\varphi(\psi)$. By definition of ψ', we get $\forall y.\psi'(y) \to \psi(y)$ from the latter. Lemma 32.1 yields $\varphi(\psi, x)$, which is the premise of $\mathrm{Clos}_\varphi(\psi)$. Thus, we get $\psi(x)$. Since this proof uses pure logic only, we have shown the first assertion.

By lemma 28.1, we get from the first assertion that $\mathsf{T}(\overbrace{\mathrm{Clos}_\varphi(\psi')}^{\cdot})$ holds in SON. By lemma 28.2, axiom III.(6), and lemma 32.2, we can move the truth predicate inside the formula and get $\mathrm{Clos}_\varphi(\mathsf{T}(\dot{\psi}'))$. □

Now, we can define *least fixed points* for positive operator forms in SON.

Proposition 34 Let $\varphi(R, x)$ be a positive operator form. Then there exists a $\mathcal{L}_\mathcal{S}$ formula $\mathcal{P}^\varphi(x)$ such that for all $\mathcal{L}_\mathcal{S}$ formulae ψ, the following holds:

$$\begin{aligned} \mathsf{SON} \quad &\vdash \quad \forall x.\varphi(\mathcal{P}^\varphi, x) \to \mathcal{P}^\varphi(x), \\ \mathsf{SON} \quad &\vdash \quad (\forall y.\varphi(\psi, y) \to \psi(y)) \to \forall x.\mathcal{P}^\varphi(x) \to \psi(x). \end{aligned}$$

Proof: We set

$$\mathcal{P}^{\varphi}(x) :\Leftrightarrow \forall z.\mathrm{Clos}_{\varphi}(\mathsf{T}(z\,\cdot)) \to \mathsf{T}(z\,x).$$

For the first assertion, let us assume $\varphi(\mathcal{P}^{\varphi}, x)$. We have to show $\mathcal{P}^{\varphi}(x)$, i.e. $\forall z.\mathrm{Clos}_{\varphi}(\mathsf{T}(z\,\cdot)) \to \mathsf{T}(z\,x)$. So we assume, additionally, $\mathrm{Clos}_{\varphi}(\mathsf{T}(z\,\cdot))$. From the definition of $\mathcal{P}^{\varphi}(x)$, we get from this that $\forall y.\mathcal{P}^{\varphi}(y) \to \mathsf{T}(z\,y)$ holds. Now, we apply lemma 32.1 and get from the first assumption $\varphi(\mathsf{T}(z\,\cdot), x)$. Using the second assumption again, we finally have $\mathsf{T}(z\,x)$.

For the second assertion, we assume $\mathcal{P}^{\varphi}(x)$, i.e., $\forall z.\mathrm{Clos}_{\varphi}(\mathsf{T}(z\,\cdot)) \to \mathsf{T}(z\,x)$. By substituting $\lambda x.\overbrace{\psi'(x)}^{\cdot}$ for z, we get $\mathrm{Clos}_{\varphi}(\mathsf{T}(\overbrace{\psi'(\cdot)}^{\cdot})) \to \mathsf{T}(\overbrace{\psi'(x)}^{\cdot})$. Lemma 33.2 yields $\mathsf{T}(\overbrace{\psi'(x)}^{\cdot})$ and with axiom T-out, we get $\psi'(x)$, i.e., $\mathrm{Clos}_{\varphi}(\psi) \to \psi(x)$, or, in other words,

$$(\forall y.\varphi(\psi, y) \to \psi(y)) \to \psi(x).$$

Thus, the second assertion is shown. □

By use of this proposition, we can easily extend the translation $\cdot^N$ of the language of Peano Arithmetic in $\mathcal{L}_p$ to a translation of $\mathcal{L}_{\mathsf{ID}}$ in $\mathcal{L}_{\mathcal{S}}$ which verifies the fixed point axioms. Induction on the natural number of ID_1 is obviously included in ($\mathcal{L}_{\mathcal{S}}$-$\mathrm{I}_{\mathsf{N}}$) and, by proposition 29, also in (T-I_{N}). Thus, we have:

Theorem 35 There exists a translation $\cdot^N$ of $\mathcal{L}_{\mathsf{ID}}$ in $\mathcal{L}_{\mathcal{S}}$ such that

$$\mathsf{ID}_1 \vdash \varphi \quad \Rightarrow \quad \mathsf{SON} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}}) \vdash \varphi^N.$$

7 Concluding remarks

Truth theories for applicative theories distinguish themselves by the possibility of representing formulae by terms *without any form of Gödelization*. In contrast, we can extend the language. The possibility to extend the language in a systematic and controlled, or *generic*, way is an essential feature of applicative theories. In particular, for truth theories, this seems to be much more natural than the use of a technical coding machinery. For this reason, Frege structures are also used in linguistics, e.g., to model *nominalisations*, cf. Hamm [Ham99].

From a mathematical point of view, the possibility to define a concept of sets based on the truth predicate is crucial. Using this "Janus face" of the truth predicate, our theories can serve as alternatives for theories which formalize sets directly, like Feferman's theories of explicit mathematics. In addition, the structure provided by the truth predicate is more flexible and, as we have shown, extends the syntactical expressive power.

Thus, truth theories over applicative theories perfectly fit into the landscape of proof theory, cf. [Kah0xa]. Here we give a snapshot of this landscape, placing the truth theories next to theories of explicit mathematics and fixed point theories or theories of inductive definitions together with the proof-theoretic ordinals.

$\mathsf{PRA}+$ $\mathrm{TI}(\alpha, X)$	Applicative theories Truth	Expl. math.	$\widehat{\mathsf{ID}}$ ID
ε_0	$\mathsf{FON} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}})$	$\mathsf{BON} + (\mathcal{L}_p\text{-}\mathrm{I}_{\mathsf{N}})$	PA
$\varphi\omega 0$	$\mathsf{FON} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}})$		$\mathsf{ID}_1^{\#}$
$\varphi\varepsilon_0 0$	$\mathsf{FON} + (\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}})$	$\mathsf{EM}_0 + \mathsf{J}$	$\widehat{\mathsf{ID}}_1$
$\Gamma_0 = \varphi 100$	$\mathsf{FSU} + (\mathsf{C}\text{-}\mathrm{I}_{\mathsf{N}})$	$\mathsf{EMU}\!\upharpoonright$	$\widehat{\mathsf{ID}}_{<\omega}$
$\varphi 1\omega 0$	$\mathsf{FSU} + (\mathsf{T}\text{-}\mathrm{I}_{\mathsf{N}})$		$\widehat{\mathsf{ID}}_{<\omega^\omega}$
$\varphi 1\varepsilon_0 0$	$\mathsf{FSU} + (\mathcal{L}_U\text{-}\mathrm{I}_{\mathsf{N}})$	EMU	$\widehat{\mathsf{ID}}_{<\varepsilon_0}$
$\Psi_\Omega(\varepsilon_{\Omega+1})$	$\mathsf{SON} + (\mathcal{L}_{\mathcal{S}}\text{-}\mathrm{I}_{\mathsf{N}})$	NEM	ID_1

As theories of explicit mathematics, we have in addition to BON with the induction principle $(\mathcal{L}_p\text{-}\mathrm{I}_{\mathsf{N}})$ for arbitrary formulae of the language $\mathcal{L}_p$ of BON, the following ones: $\mathsf{EM}_0 + \mathsf{J}$: Explicit mathematics plus join, [Bee85]. EMU: Explicit mathematics with universes, [Str99], where $\mathsf{EMU}\!\upharpoonright$ denotes the theory with induction restricted to *types*. NEM: Name induction in explicit mathematics, [KS00].

As references for the proof-theoretic investigations, we refer to the following papers: For truth in applicative theories: [Can96, Kah0xb], for explicit mathematics, [Bee85, Str99, KS00], for theories of fixed points and inductive definitions, [JS96, Acz77, Fef82, JKSS99, BFPS81].

We will finish with two suggestion to extend the approach of truth in applicative theories further on in the impredicative world.

First, one can consider *universes for supervaluation* in analogy to Frege structures, hoping to get theories which allow syntactical embeddings of iterated inductive definition ID_α. Such theories have not been studied yet but, again, Cantini has presented a framework which is very closely related which admits embeddings of finite constructive number classes $\mathsf{ID}_{n+1}(\mathcal{O})$, cf. [Can91].

Secondly, one can think of a strengthening of Frege structures by adding an induction principle for the truth predicate. Maybe such a principle yields a uniform step to impredicative theories for the basic theory, as well as for the theories with universes. A similiar step is possible in the context of explicit mathematics by adding a induction principle for the naming relation $\Re$, [KS00, JKS0x].

References

[Acz77] Peter Aczel. The strength of Martin Löf's intuitionistic type theory with one universe. In Seppo Miettinen and Jonko Väänänen, editors, *Proceedings of the Symposiums on Mathematical Logic in Oulo 1974 and in Helsinki 1975*. Report No. 2 from the Dept. of Philosophy, University of Helsinki, 1977.

[Acz80] Peter Aczel. Frege structures and the notion of proposition, truth and set. In J. Barwise, H. Keisler, and K. Kunen, editors, *The Kleene Symposium*, pages 31–59. North-Holland, 1980.

[Bee85] Michael Beeson. *Foundations of Constructive Mathematics*. Ergebnisse der Mathematik und ihrer Grenzgebiete; 3.Folge, Bd. 6. Springer, Berlin, 1985.

[BFPS81] Wilfried Buchholz, Solomon Feferman, Wolfram Pohlers, and Wilfried Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical studies*, volume 897 of *Lecture Notes in Mathematics*. Springer, 1981.

[Can89] Andrea Cantini. Notes on formal theories of truth. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 35:97–130, 1989.

[Can90] Andrea Cantini. A theory of formal truth arithmetically equivalent to ID_1. *Journal of Symbolic Logic*, 55(1):244–259, 1990.

[Can91] Andrea Cantini. A logic of abstraction related to finite constructive number classes. *Archive for Mathematical Logic*, 31:69–83, 1991.

[Can93] Andrea Cantini. Extending the first-order theory of combinators with self-referential truth. *Journal of Symbolic Logic*, 58(2), 1993.

[Can95] Andrea Cantini. Levels of truth. *Notre Dame Journal of Formal Logic*, 36:185–213, 1995.

[Can96] Andrea Cantini. *Logical Frameworks for Truth and Abstraction*, volume 135 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1996.

[Fef75] Solomon Feferman. A language and axioms for explicit mathematics. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer, 1975.

[Fef79] Solomon Feferman. Constructive theories of functions and classes. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium 78*, pages 159–224. North–Holland, Amsterdam, 1979.

[Fef82] Solomon Feferman. Iterated inductive fixed-point theories: Application to Hancock's conjecture. In G. Metakides, editor, *Patras Logic Symposion*, pages 171–196. North–Holland, Amsterdam, 1982.

[Fef88] Solomon Feferman. Hilbert's program relativized: Proof-theoretical and foundational reductions. *Journal of Symbolic Logic*, 53(2):364–384, 1988.

[Fef91] Solomon Feferman. Reflection on imcompleteness. *Journal of Symbolic Logic*, 57(1):1–49, 1991.

[Fef00] Solomon Feferman. Does reductive proof theory have a viable rationale? *Erkenntnis*, 53(1–2):63–96, 2000.

[FJ93] Solomon Feferman and Gerhard Jäger. Systems of explicit mathematics with non-constructive μ-operator. Part I. *Annals of Pure and Applied Logic*, 65(3):243–263, December 1993.

[FJ96] Solomon Feferman and Gerhard Jäger. Systems of explicit mathematics with non-constructive μ-operator. Part II. *Annals of Pure and Applied Logic*, 79:37–52, 1996.

[FM87a] Robert Flagg and John Myhill. An extension of Frege structures. In D. Kueker, E. Lopez-Escobar, and C. Smith, editors, *Mathematical Logic and theoretical computer science*, pages 197–217. Dekker, 1987.

[FM87b] Robert Flagg and John Myhill. Implication and analysis in classical Frege structures. *Annals of Pure and Applied Logic*, 34:33–85, 1987.

[FS82] Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1982.

[Hal96] Volker Halbach. *Axiomatische Wahrheitstheorien.* Akademie-Verlag, Berlin, 1996.

[Ham99] Fritz Hamm. Modelltheoretische Untersuchungen zur Semantik von Nominalisierungen. Habilitionsschrift, Seminar für Sprachwissenschaften, Universität Tübingen, 1999.

[HK95] Susumu Hayashi and Satoshi Kobayashi. A new formulation of Feferman's system of functions and classes and its relation to Frege structures. *International Journal of Foundations of Computer Science*, 6(3):187–202, 1995.

[Jäg82] Gerhard Jäger. Zur Beweistheorie der Kripke-Platek-Mengenlehre über den natürlichen Zahlen. *Archiv für mathematische Logik und Grundlagenforschung*, 22:121–139, 1982.

[JKS99] Gerhard Jäger, Reinhard Kahle, and Thomas Strahm. On applicative theories. In A. Cantini, E. Casari, and P. Minari, editors, *Logic and Foundation of Mathematics*, pages 88–92. Kluwer, 1999.

[JKS0x] Gerhard Jäger, Reinhard Kahle, and Thomas Studer. Universes in explicit mathematics. *Annals of pure and applied logic*, 200x. To appear.

[JKSS99] Gerhard Jäger, Reinhard Kahle, Anton Setzer, and Thomas Strahm. The proof-theoretic analysis of transfinitely iterated fixed point theories. *Journal of Symbolic Logic*, 64(1):53–67, March 1999.

[JS95] Gerhard Jäger and Thomas Strahm. Totality in applicative theories. *Annals of Pure and Applied Logic*, 74:105–120, 1995.

[JS96] Gerhard Jäger and Thomas Strahm. Some theories with positive induction of ordinal strength $\varphi\omega 0$. *Journal of Symbolic Logic*, 61(3):818–842, September 1996.

[Kah97a] Reinhard Kahle. *Applikative Theorien und Frege-Strukturen.* PhD thesis, Institut für Informatik und angewandte Mathematik, Universität Bern, 1997.

[Kah97b] Reinhard Kahle. Uniform limit in explicit mathematics with universes. Technical Report IAM-97-002, IAM, Universität Bern, März 1997.

[Kah99] Reinhard Kahle. Frege structures for partial applicative theories. *Journal of Logic and Computation*, 9(5):683–700, October 1999.

[Kah0xa] Reinhard Kahle. Mathematical proof theory in the light of ordinal analysis. *Synthese*, 200x. To appear.

[Kah0xb] Reinhard Kahle. Universes over Frege structures. 200x. Submitted.

[Kri75] Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[KS00] Reinhard Kahle and Thomas Studer. A theory of explicit mathematics equivalent with ID_1. In P. Clote and H. Schwichtenberg, editors, *CSL 2000*, volume 1862 of *LNCS*, pages 356–370. Springer, 2000.

[KS0x] Reinhard Kahle and Thomas Studer. Formalizing non-termination of recursive programs. 200x. Submitted.

[Mar84] Per Martin-Löf. *Intuitionistic Type Theory.* Bibliopolis, 1984.

[Mar93] Markus Marzetta. Universes in the theory of types and names. In E. Börger et al., editor, *Computer Science Logic '92*, volume 702 of *Lecture Notes in Computer Science*, pages 340–351. Springer, 1993.

[Mar94] Markus Marzetta. *Predicative Theories of Types and Names.* PhD thesis, Universität Bern, Institut für Informatik und angewandte Mathematik, 1994.

[Poh98] Wolfram Pohlers. Subsystems of set theory and second order number theory. In S. Buss, editor, *Handbook of Proof Theory*, chapter IV, pages 209–335. North-Holland, 1998.

[Sch77] Kurt Schütte. *Proof Theory*, volume 225 of *Grundlehren der mathematischen Wissenschaften.* Springer, Berlin, 1977.

[Sco75] Dana Scott. Combinators and classes. In C. Böhm, editor, *λ-Calculus and Computer Science Theory*, volume 37 of *Lecture Notes in Computer Science*, pages 1–26. Springer, 1975.

[Smi78] Jan Smith. *On the relation between a type theoretic and a logical formulation of the theory of constructions.* PhD thesis, Dept. of Mathematics, University of Göteborg, 1978.

[Smi84] Jan Smith. An interpretation of Martin-Löf's type theory in a type-free theory of proposition. *Journal of Symbolic Logic*, 49(3):730–753, 1984.

[Str96] Thomas Strahm. *On the proof theory of applicative theories.* PhD thesis, Universität Bern, Institut für Informatik und angewandte Mathematik, 1996.

[Str99] Thomas Strahm. First steps into metapredicativity in explicit mathematics. In S. Cooper and J. Truss, editors, *Sets and Proofs*, volume 258 of *London Mathematical Society Lecture Note Series*, pages 383–402. Cambridge University Press, 1999.

[Tar35] Alfred Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1:261–405, 1935. Reprinted in: Alfred Tarski. Collected Papers. Volume 2. pp. 51–198. Birkhäuser, Basel. 1986. Original version of [Tar56].

[Tar56] Alfred Tarski. The concept of truth in formalized languages. In *Logic, Semantics, Metamathematics*, pages 152–278. Clarendon Press, Oxford, 1956. Revised and supplemented English translation of [Tar35].

[vF68] Bas van Fraassen. Presupposition, implication and self-reference. *Journal of Philosophy*, 65:135–152, 1968.

[vF70] Bas van Fraassen. Inference and self-reference. *Synthese*, 21:425–438, 1970.