THE REVIEW OF SYMBOLIC LOGIC
Volume 4, Number 2, June 2011

THE UNIVERSAL SET AND DIAGONALIZATION IN FREGE STRUCTURES

REINHARD KAHLE
Centria and Departamento de Matemática
Universidade Nova de Lisboa

Abstract. In this paper we summarize some results about sets in Frege structures. The resulting set theory is discussed with respect to its historical and philosophical significance. This includes the treatment of diagonalization in the presence of a universal set.

§1. Introduction. Going back to prior work of Scott (1975), *Frege structures* were introduced by Aczel (1980) as a semantical concept to define a notion of set by means of a partial truth predicate. Beeson (1985) introduced a truth theory for *applicative theories* by adding predicates for propositions and truths. (The underlying applicative theories build the first-order part of Feferman's 1975, 1979 systems of *explicit mathematics*.) He shows that such a truth theory is an axiomatic counterpart of Aczel's *Frege structures*.

Flagg & Myhill (1987a, 1987b) investigate Frege structures from a more general perspective. In particular, they give up the primitive notion of proposition, defining it instead in terms of truth. The most comprehensive study of truth theories on an applicative basis may be found in Cantini's (1996) monograph, see also Cantini (1993).

These truth theories offer two distinguished features, namely "representation of formulae without Gödelization" and "abstraction". The first one refers to the possibility of formalizing the truth theory without Gödelization of the language. Instead, new applicative constants are used to represent formulae on the term level.

In this paper, we are mainly concerned with the second feature, abstraction. By "abstraction" we mean the possibility of introducing a notion of *set* (class, type) in this theory, following Aczel's original motivation. One may say that these theories allow *unrestricted comprehension* and a *partial* element-of relation: by unrestricted comprehension we build *abstraction terms* $\{x \mid \varphi(x)\}$ for arbitrary formulae $\varphi(x)$, but it is not guaranteed that $x \in \{x \mid \varphi(x)\}$ is always equivalent with $\varphi(x)$. An abstraction term is considered as a *set* only, if the element relation for it is total (in a sense defined below).

Since the abstraction terms are abbreviations from the underlying applicative language, the resulting set theory is still first-order. In fact, in its basic formulation it is a proof-theoretically rather weak theory. However, it has already a strong syntactical expressive power: it contains, for instance, the basic theory of *explicit mathematics*. One interesting aspect is that the set-existence assertions in this theory are not given by axioms; they are propositions of the properties of the underlying applicative theories. These theories are of recursion-theoretic nature. Thus, the "philosophical justification" of certain set existence claims is automatically given by the recursion-theoretic construction of the underlying terms.

Received: May 3, 2008

© Association for Symbolic Logic, 2011
doi:10.1017/S1755020310000407

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

As an important consequence it follows that our Frege structures permit the definition of a universal set. From Cantor's well-known paradox shows that the existence of a universal set is incompatible with a power set operation, we can conclude that our set theory excludes the power set operation. Thus, Cantor's (general) diagonalization method is not applicable in this case. However, it is possible to show that Cantor's (first) diagonalization argument for the power set of the natural numbers works in Frege structures.

The structure of the paper is as follows. In the next section we introduce the applicative theory TON which provides the basic term language. In Section 3 we introduce Frege structures as a truth theory over TON and present some basic properties. The set theory of Frege structures is introduced in Section 4. In Section 5 we discuss the consequences of the universal set and the treatment of the diagonalization method. The final section is devoted to some historical and philosophical considerations. In this paper we do not present new technical results, but the main aim is to give an overview of sets in Frege structures and to discuss the significance of this set theory.[1]

Frege structures are discussed by several authors in several variations. Thus, the approach given here might differ from the one the reader is familiar with. First, we work in classical logic, while the original formulation was based on intuitionistic logic. Secondly, we follow Flagg & Myhill (1987a, 1987b), in generalizing Aczel's original approach by defining "proposition" in terms of "truth" (instead of taking it as a primitive predicate). This generalization is essential for several technical results. In general, we follow the formalization of Cantini (1996).[2] In fact, to a large extent we borrow the development of sets in Frege structures from Cantini (1996, secs. 9A and 9B and around).

§2. The applicative theory TON. The theory TON (total theory of operations and numbers), introduced and studied by Jäger & Strahm (1995), is the total version of the theory BON (basic theory of operations and numbers) introduced by Feferman & Jäger (1993) (see also the textbooks Beeson, 1985 and Troelstra & van Dalen, 1988 which contain presentations of first-order applicative theories). It is formulated in $\mathcal{L}_t$, the first-order language of operations and numbers, comprising of individual variables $x, y, z, v, w, f, i, m, n, \ldots$, individual constants k, s (combinators), p, p_0, p_1 (pairing and projection), 0, s_N, p_N (zero, successor, and predecessor), 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. Formulas $(\varphi, \psi, \ldots)$ are constructed from $\neg$, $\wedge$, and $\forall$ in the usual manner, starting from the atomic formulas $t = s$ and $\mathsf{N}(t)$.

In the sequel, we write $s\,t$ for $(s \cdot t)$ with the convention of association to the left. The connectives $\vee$, $\rightarrow$, $\leftrightarrow$, and $\exists$ are defined as usual from the other connectives.

The logic of TON is classical first-order predicate logic with equality, formulated in a Hilbert-style calculus. The nonlogical axioms of TON include:

[1] In this paper, we do not discuss the relations to other theories which allow a universal set, as presented, for example, in Church (1974) and Oberschelp (1973), nor Quine's original *New Foundations*, cf., Forster (1995). These theories seem to have a rather different setup than then one considered here. Yet another approach, allowing to define sets by use of abstraction for *positive* formulae over a Frege-like system, is given in Libert (2008).

[2] Cantini does not call his truth theories *Frege structures*, but reserves this name for the more restrictive approach of Aczel, cf., Cantini (1996, Remark 8.6.1.(i), p. 53). However, already Flagg and Myhill use the name *Frege structures* in the broader sense as we do here.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

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) $\mathsf{N}(0) \wedge \forall x.\mathsf{N}(x) \to \mathsf{N}(\mathsf{s}_\mathsf{N}\,x)$,
(5) $\forall x.\mathsf{N}(x) \to \mathsf{s}_\mathsf{N}\,x \neq 0 \wedge \mathsf{p}_\mathsf{N}\,(\mathsf{s}_\mathsf{N}\,x) = x$,
(6) $\forall x.\mathsf{N}(x) \wedge x \neq 0 \to \mathsf{N}(\mathsf{p}_\mathsf{N}\,x) \wedge \mathsf{s}_\mathsf{N}\,(\mathsf{p}_\mathsf{N}\,x) = x$.

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

(7) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v = w \to \mathsf{d}_\mathsf{N}\,x\,y\,v\,w = x$,
(8) $\mathsf{N}(v) \wedge \mathsf{N}(w) \wedge v \neq w \to \mathsf{d}_\mathsf{N}\,x\,y\,v\,w = y$.

The definition of applicative theories in Feferman & Jäger (1993) and Jäger & Strahm (1995) also includes primitive recursion. It is needed only for very restricted forms of induction which we do not discuss.

In the standard way, we can introduce the usual λ-notation in TON. Moreover, the type-free approach allows self-application from which we can define a recursion operator rec.

PROPOSITION 2.1 (folklore).

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* $\forall x.\mathsf{rec}\,x = x\,(\mathsf{rec}\,x)$.

The theory TON is proof-theoretically equivalent to Peano arithmetic PA if we add the schema of formula induction on N:

Formula induction on N ($\mathcal{L}_t$-I_N)

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

By interpreting natural numbers as elements of N, we get a straightforward interpretation of PA in TON + ($\mathcal{L}_t$-I_N), cf., Jäger & Strahm (1995).

Models of TON are discussed at length by Beeson (1985), Cantini (1996), Jäger & Strahm (1995), and Strahm (1996). One important example, the *closed total term model* $\mathcal{CTT}$ can roughly be described as follows:

The universe of $\mathcal{CTT}$ consists of the set of all closed terms of the language $\mathcal{L}_t$, that is, we interpret the constants by themselves, and application by juxtaposition. Based on a straightforward reduction relation ϱ for the constants of $\mathcal{L}_t$, we interpret equality of terms by the fact that they have a common reduct with respect to *arbitrary* reductions on the basis of ϱ. For this reason, we need to prove the *Church–Rosser property* to verify transitivity of equality. $\mathsf{N}(t)$ holds if t reduces to a numeral. For more details about $\mathcal{CTT}$ we refer to Jäger & Strahm (1995) and Cantini (1996).

This model can be formalized in PA. Thus, we get the proof-theoretic equivalence of TON + ($\mathcal{L}_t$-I_N) and PA.

§3. Frege structures.

3.1. The theory FON. In this section we introduce *Frege structures* as a truth theory for TON. There are several variant axiomatizations. We follow essentially the theory

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

MF^- of Cantini (1996) (or NMT of Cantini, 1993), cf., also Kahle (2007, chap. 3). The axiomatizations in Beeson (1985) and Hayashi & Kobayashi (1995), for instance, presuppose a notion of proposition which is defined here in terms of the truth predicate. The definition given here, is in some sense the most general one.

The language $\mathcal{L}_F$ of FON is the language of TON extended by the new relation symbol T (truth) and new individual constants $\doteq$, $\dot{\mathsf{N}}$, $\dot{\neg}$, $\dot{\wedge}$, and $\dot{\forall}$.[3]

For readability reasons we use in the following freely an infix notation for some terms containing dotted constants. So $x \doteq y$, for instance, has to be written formally as $\doteq x\ y$.

The axioms of FON (Frege structures over TON) are the axioms of TON extended to the new language plus the following axioms:

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

(1) $x = y \leftrightarrow \mathsf{T}(x \doteq y)$,
(2) $\neg x = y \leftrightarrow \mathsf{T}(\dot{\neg}\,(x \doteq y))$,
(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) $\neg(\mathsf{T}(x) \wedge \mathsf{T}(\dot{\neg}\,x))$.

The axioms I–II express that T is closed under formula construction by $=$, N, T, $\wedge$, $\forall$, and negation where T may be used only positively.

We will use the following abbreviation:

$$\mathsf{P}(t) \quad :\Leftrightarrow \quad \mathsf{T}(t) \vee \mathsf{T}(\dot{\neg}\,t).$$

$\mathsf{T}(t)$ may be read as "t is true" and $\mathsf{P}(t)$ as "t is a proposition."

By use of the recursion operator from Proposition 2.1, we can diagonalize over $\dot{\neg}$ and show that not everything can be a proposition (somehow in contrary to Frege's [inconsistent] conception).

LEMMA 3.1. $\mathsf{FON} \vdash \neg\forall x.\mathsf{P}(x)$.

Proof. Set $l := \mathsf{rec}\,(\lambda x.\dot{\neg}\,x)$, that is, $l = \dot{\neg}\,l$. Now, $\mathsf{T}(l)$ is equivalent with $\mathsf{T}(\dot{\neg}\,l)$. (So l can be seen as a formalization of the liar sentence.) Thus, $\mathsf{P}(l)$ would contradict the axiom of consistency. □

This lemma shows that we cannot have a truth definition for the whole language.[4] In particular, we cannot add full self-reference. Thus, we are not allowed to add a term $\dot{\mathsf{T}}$ such

[3] Cantini introduces for his theory MF^- these constants as abbreviated terms, cf., Cantini (1996, definition 7.1).

[4] As for usual hierarchies of truth, there exists the possibility to define additional *levels* of truth which can treat negative truth of a lower level as a positive statement on the upper level. This is carried out in Cantini's (1996, chap. 8) theory TLR and in our FSU (Kahle, 2003). The latter work also contains a short discussion of the treatment of the liar and the Russell paradox, relativized to levels of truth.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

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:[5]

Self-reference

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

As in TON we will consider formulae induction on N in FON to get mathematical strength:

1. Formulae induction on N ($\mathcal{L}_F$-I_N)

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

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

Our theories FON and FON$+(\mathcal{L}_F$-$\mathsf{I}_\mathsf{N})$ are essentially equivalent with the theories MF$^-$ and MF, respectively, of Cantini (1996).

3.2. Representation of formulae. One of the essential features of Frege structures is the possibility of representing formulae by terms without Gödelizing the language.

We can straightforwardly associate a term with every formula in FON:

DEFINITION 3.2. *By induction of the build up of an $\mathcal{L}_F$ formula, we define:*

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

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 $(\dot{\exists}x.t) := \dot{\neg}\,(\dot{\forall}x.\dot{\neg}\,t)$.

Using the dot-notation $\dot{\varphi}$ we would like to indicate the difference of this representation compared with a Gödelization $\ulcorner\varphi\urcorner$ which is based on an arithmetization of the language.[6] From a technical point of view, the representation is particularly convenient with respect to

[5] In Cantini's theory MF$^-$, these two formulas are added as axioms, with $\dot{\mathsf{T}}$ as an additional abbreviated term.

[6] To be fair, we have to admit that we pay a price here: the introduction of the truth predicate T "costs" something in terms of proof-theoretic strength (see Theorem 3.7 below); this is not the case for an arithmetization of the language.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

variables and substitution: since terms and variables stay essentially untouched in the term representing a formula, $\dot{\varphi}[t/x]$ is equal to $\dot{\overbrace{\varphi[t/x]}}$.[7]

3.3. **T*-positiveness.*** 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 3.3.

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:

PROPOSITION 3.4. (Cantini, 1996, theorem 8.8) If φ is a T-positive formula of $\mathcal{L}_F$, then we have:

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

3.4. ***Fixed points.*** Scott (1975) defines a fixed point theory of sets by use of a Fixed Point Theorem on the ground structure. In the same manner, Cantini (1996, sec. 10A) proves that the truth theories may be used to define *fixed points* of positive operator forms. We include this result here, on the one hand, to illustrate the expressive power of the underlying applicative basis, on the other hand, since it is used to establish the proof-theoretic strength of our theory. First, 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.[8] Now, we can use Proposition 3.4 to define *fixed points* of such operator forms within FON. The proof makes essential use of the possibility of defining fixed points on the term level.

PROPOSITION 3.5. *Let $\varphi(R, x)$ be a positive operator form. Then there exists a term t_φ of $\mathcal{L}_F$ such that*[9]

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

Proof. We define t_φ by use of the Recursion Theorem 2.1:

$$t_\varphi := \mathsf{rec}\,(\lambda y, x.\dot{\overbrace{\varphi(\mathsf{T}(y\,\cdot), x)}}).$$

---

[7] This property depends, in the case of $\dot{\overbrace{\forall x.\varphi}}$, on the definition of the λ-terms. In the total setting this is unproblematic, however, it causes some technical difficulties for Frege structures over partial applicative theories, cf., Kahle (1999).

[8] This notion is analogous to the one used for $\widehat{\mathrm{ID}}_1$, see below.

[9] The dot in $\mathsf{T}(t_\varphi\,\cdot)$ marks the place where the argument of the relation variable R has to be substituted.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

So we have $t_\varphi\, x = \dot{\overbrace{\varphi(\mathsf{T}(t_\varphi\, \cdot), x)}}$ and get

$$\begin{array}{rcl} \mathsf{T}(t_\varphi\, x) & \leftrightarrow & \mathsf{T}(\dot{\overbrace{\varphi(\mathsf{T}(t_\varphi\, \cdot), x)}}) \\ & \leftrightarrow & \varphi(\mathsf{T}(t_\varphi\, \cdot), x). \end{array}$$ □

Note, that the T-positiveness of φ is crucial when applying Proposition 3.4.

This proposition can be used to establish an embedding of the fixed point theory $\widehat{\mathrm{ID}}_1$ which postulates not-necessarily-least fixed points of positive arithmetical operator forms, cf., Aczel (1977). A positive operator forms $\varphi(R, x)$ is a formula in the language of Peano Arithmetic having only positive occurrences of the relation symbol R. For the definition of $\widehat{\mathrm{ID}}_1$ we extend the language of Peano Arithmetic by new predicate symbols $\mathcal{P}^\varphi$ for each positive operator form $\varphi(R, x)$. Then, $\widehat{\mathrm{ID}}_1$ extends Peano Arithmetic by the following new axiom for each of new predicate symbol $\mathcal{P}^\varphi$:

$$(\widehat{\mathrm{ID}}_1) \qquad \forall x.\varphi(\mathcal{P}_\varphi, x) \leftrightarrow \mathcal{P}_\varphi(x).$$

Finally, in $\widehat{\mathrm{ID}}_1$ the induction scheme of Peano Arithmetic is extended to the new language.

Now, a standard translation $\cdot^N$ of Peano Arithmetic in $\mathsf{TON} + (\mathcal{L}_t\text{-}\mathsf{I_N})$ can easily be extended to the fixed points constants of $\widehat{\mathrm{ID}}_1$. So we get:

PROPOSITION 3.6. *There exists a translation $\cdot^N$ from the language of $\mathcal{L}_{\mathsf{ID}}$ into the language $\mathcal{L}_F$ such that*

$$\widehat{ID}_1 \vdash \varphi \quad \Rightarrow \quad \mathsf{FON} + (\mathcal{L}_F\text{-}\mathsf{I_N}) \vdash \varphi^N.$$

Cantini (1996, sec. 57) showed that $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathsf{I_N})$ does not exceed the proof-theoretic strength of $\widehat{\mathrm{ID}}_1$ such that we have

THEOREM 3.7. $\mathsf{FON} + (\mathcal{L}_F\text{-}\mathsf{I_N})$ *and $\widehat{ID}_1$ are proof-theoretic equivalent and their proof-theoretic ordinal is $\varphi\, \varepsilon_0\, 0$ (in terms of the Veblen function φ).*

REMARK 3.8. *This Theorem shows that we cannot expect to model* least fixed points *in* FON *as they are given in the proof-theoretically much stronger theory ID_1. But, even for a strengthened theory it is not to see how the Recursion Theorem of the applicative theories could help to define* least *fixed points. Cantini (1996, chap. 12) introduced in a truth theory over applicative theories based on* supervaluation *which allows for an embedding of ID_1, cf., also Kahle (2001). In Kahle & Studer (2001) an applicative theory is introduced which provides a least fixed points for recursive equations on the natural numbers (but not for arbitrary positive operator forms).*

§4. Sets in Frege structures. Now, we introduce *abstraction terms* together with an *element relation* on the basis of the truth predicate, cf., Scott (1975).

DEFINITION 4.1. *Given two $\mathcal{L}_F$ terms t and s and an $\mathcal{L}_F$ formula φ, we define:*

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

Since the term $\{x \mid \varphi\}$ is defined for arbitrary formulae φ, we may say that Frege structures allow unrestricted (or full) comprehension. However, the element relation is only

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

partial, that is, not in every case can we actually prove $t \in \{x \mid \varphi\}$ when $\varphi[t/x]$ holds. But we can do so at least for T-positive formulae. From Proposition 3.4, we get as a corollary:

COROLLARY 4.2. *If φ is a T-positive formula of $\mathcal{L}_F$, then we have:*

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

Note that $\{x \mid \varphi\}$ is, so far, only an abbreviated term, defined for arbitrary formulae φ. *Sets*[10] are now introduced as terms for which the element relation is total.[11]

DEFINITION 4.3 (Sets in FON).

$$\mathsf{Set}(t) :\Leftrightarrow \forall x.\mathsf{P}(t\,x), \textit{ that is, } \forall x.\mathsf{T}(t\,x) \vee \mathsf{T}(\dot{\neg}\,(t\,x)).$$

If for a term t of $\mathcal{L}_F$, $\mathsf{FON} \vdash \mathsf{Set}(t)$, we say that t is a set (in FON).

Sets in FON can also be characterized as T-*characteristic functions*. Since they are *defined* in terms of the applicative ground structure, the set formation principles are induced by this ground structure (rather than given by explicit axioms as in usual set theories). It turns out that the applicative ground structure provides us with a handy collection of such set formation principles. In a very general way, we can build new sets for *elementary* formula, a notion taken over from explicit mathematics, Feferman (1975), and defined here as follows.

DEFINITION 4.4. (Cantini, 1996, definition 9.5.(i)) A formula φ is *elementary* in $x_1, \ldots, x_n$ iff φ is built up from $t = s$, $\mathsf{N}(t)$, $\neg t = s$, $\neg\mathsf{N}(t)$, $t \in x_i$, and $\neg(t \in x_i)$ $(1 \leq i \leq n)$ by means of $\wedge$, $\vee$, $\forall y$, and $\exists y\,(y \notin \{x_1, \ldots, x_n\})$.

Elementary formulas are essentially T-free, except for the element expressions $t \in x_i$ and $\neg(t \in x_i)$, which are abbreviations for $\mathsf{T}(x_i\,t)$ and $\neg\mathsf{T}(x_i\,t)$, respectively. We will use these expressions only in a context where $\mathsf{Set}(x_i)$ is assumed, which implies that the latter is equivalent to the T-positive expression $\mathsf{T}(\dot{\neg}\,(x_i\,t))$.

Now, the following proposition is easy to verify.

PROPOSITION 4.5. (Cantini, 1996, corollary 9.7.(ii)) If $\varphi(y, x_1, \ldots, x_n)$ is elementary in $x_1, \ldots, x_n$, then

$$\mathsf{FON} \vdash \mathsf{Set}(x_1) \wedge \cdots \wedge \mathsf{Set}(x_n) \rightarrow \mathsf{Set}(\{y \mid \varphi(y, x_1, \ldots, x_n)\}).$$

This proposition corresponds to the *elementary comprehension* of explicit mathematics (Feferman, 1975, 1979). In addition, the *join principle* of explicit mathematics can also be defined in FON, cf., Cantini (1996, proposition 9.9). As a consequence we can embed Feferman's theory of explicit mathematics with join, $\mathsf{EM}_0 + \mathsf{J}$ (Feferman, 1975, 1979;

[10] What we call here *set* is also called *propositional function* (see the previous note) or *class*. In particular, Cantini (1996) uses the name *class* (with reference to Feferman) and we followed him in other papers (e.g., in Kahle, 2003). But, because the abstraction terms can occur without problem on the left side of the element relation, it seems to be more appropriate to call them indeed *sets*.

[11] This definition is *literally* the one Aczel (1980, definition 3.4) used to define *sets* as *propositional functions*. Note, however, that in Aczel's approach the truth axioms are relativized to (axiomatized) propositions. Therefore, our theory might admit "more" sets, at least as terms on the syntactic level.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

Beeson, 1985) or $\mathsf{EETJ}+(F\text{-}I_{\mathsf{N}})$ (Feferman & Jäger, 1996), in $\mathsf{FON}+(\mathcal{L}_F\text{-}\mathrm{I}_{\mathsf{N}})$, cf., Cantini (1996, remark 9.9.1).

REMARK 4.6. *Since explicit mathematics is formulated over* partial *applicative theories, there is no direct interpretation of Frege structures in explicit mathematics. In Kahle (1999) we introduced Frege structures for partial applicative theories. However, one faces some technical problems which make the theories less handy. Let us mention that Hayashi & Kobayashi* (1995) gave a version of Frege structures which is exactly tuned to be equivalent to a version of explicit mathematics. This formalization uses propositions (P) as a primitive predicate.

§5. The universal set. A distinguished feature of our set theory (as well as of explicit mathematics) is the possibility to define a universal set:

PROPOSITION 5.1. *Defining* $V := \{x \mid x = x\}$, *we have* $\mathsf{FON} \vdash \mathsf{Set}(V)$.

The proof is trivial, since $x = x$ is a T-free formula.

REMARK 5.2. *Since our sets are given by terms, we can immediately conclude* $V \in V$, *which is an abbreviation for* $\mathsf{T}((\lambda x.x \doteq x)\,(\lambda x.x \doteq x))$.

In our context, the universal sets has to be distinguished from the "set" of all sets, given by $\mathsf{SET} := \{x : \mathsf{Set}(x)\}$.

PROPOSITION 5.3 (cf., Cantini, 1996, propositoin 9.3). $\mathsf{FON} \vdash \neg\mathsf{Set}(\mathit{SET})$.

Proof. The proof uses a Russell-style argument: let $r = \{x \mid x \in \mathsf{SET} \wedge \neg(x \in x)\}$. We have $r \in \mathsf{SET}$ and $r \in r \leftrightarrow (r \in \mathsf{SET} \wedge \neg(r \in r)) \leftrightarrow \neg(r \in r)$, contradiction. □

5.1. Cantor's paradox. Nowadays, a universal *set* is widely discredited by Cantor's paradox. This paradox makes essential use of the power set axiom, and it is performed by the famous diagonalization method introduced by Cantor. It is, in fact, an immediate corollary from Cantor's theorem:

THEOREM 5.4 (of ordinary set theory). *The power set* $\mathcal{P}(A)$ *of a given set* A *has higher cardinality than* A.

Proof. Assume that f is a surjective function from A to $\mathcal{P}(A)$. Let $C := \{x \in A \mid x \notin f(x)\}$. C is by definition a subset of A, that is, $C \in \mathcal{P}(A)$. However, C is not in the range of f. Let us assume the contrary, that is, it exists an $c \in A$ such that $C = f(c)$. But this results in a contradiction when we ask whether c belongs to C:

$$\begin{aligned} c \in C &\Leftrightarrow c \notin f(c) \\ &\Leftrightarrow c \notin C \end{aligned}$$

□

Since the power set of the universal set would have higher cardinality than the universal set itself, it would contain at least one element which could not be in the universal set, thus we get:

COROLLARY 5.5 (Cantor's paradox). *There is no universal set in ordinary set theory.*

5.2. Power sets. Cantor's paradox is blocked in our theory by the fact that we cannot define a proper power set operation. Defining in FON the subset relation $x \subseteq y$ by the formula $\forall z.z \in x \to z \in y$, the straightforward idea to define a power *set* operation $\mathsf{Pow}(y)$

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

would be as $\{x \mid \mathsf{Set}(x) \wedge x \subseteq y\}$. But this is not an elementary formula, since neither $\mathsf{Set}(x)$ nor $x \subseteq y$, as shorthand notation for $\forall z.\neg\mathsf{T}(x\,z) \vee \mathsf{T}(y\,z)$, is elementary. There is, in fact, no possibility to define $\mathsf{Pow}(y)$ as an elementary formula in FON cf., Cantini (1996, proposition 9.15(iii)).

PROPOSITION 5.6. $\mathsf{FON} \vdash \exists x.\mathsf{Set}(x) \wedge \neg\mathsf{Set}(\mathsf{Pow}(x))$.

Proof. We have $\mathsf{Set}(V)$. Since $\mathsf{Pow}(V) = \{x \mid \mathsf{Set}(x) \wedge x \subseteq V\}$ is equivalent to SET, we get from $\mathsf{Set}(\mathsf{Pow}(V))$, $\mathsf{Set}(\mathsf{SET})$ in contradiction to Proposition 5.3. □

REMARK 5.7 (Cantini, 1996, proposition 9.15(i)). *As an alternative for a power set operation, Cantini also considers the operation* $\mathsf{Pow}^*(y)$ *defined as* $\{x \mid x \subseteq y\}$*, where the set condition on* x *is dropped. But it does not work properly, since we have* $\mathsf{FON} \vdash \neg\forall x.\forall y.x \subseteq y \to x \in \mathsf{Pow}^*(y)$.

REMARK 5.8. *Although there is no power* set*, we can speak about all subsets of a given set in the sense that we can, in principle, prove* formulas *of the form* $\forall x.x \subseteq y \to \psi(x)$. *Leaving open what we can actually prove for certain power "sets" in this form, we can at least speak about them. The fact that we cannot build the* set *just means that we cannot apply the set-theoretic operations available in Frege structures. This is exactly the situation we have for (proper) classes in certain set theories. In this sense one may say that Frege structures allow the introduction of a concept of (proper) classes on the level of formulas.*

REMARK 5.9. *In* explicit mathematics*, the twin framework of* Frege structures*, Feferman* (1979) proposed two different forms of formalizations of power set (class/type), a weak and a strong one. He shows by a simple diagonalization argument that the weak one (and *a fortiori* also the strong one) is inconsistent *in the presence of join*. However, the weak one is consistent with many systems of explicit mathematics without join, cf., Feferman (1979) and Glass (1996). (Jäger, 1997, showed that the strong one is inconsistent even in the absence of join as long as one has *uniform* elementary comprehension.) As a consequence of the applicative "power" of our ground structure, Frege structures have join automatically built in. Therefore, there is no correspondence to the weak power set axiom in our context.

5.3. Cantor's diagonalization method. Let us turn now to Cantor's diagonalization method as used in the proof of the uncountability of the real numbers. This proof, praised by Hilbert (1905, p. 48) as "einer *der schönsten Beweise* der Mengenlehre" ("one of the most beautiful proofs in set theory", also cited in Moore, 2002, p. 48), is more concrete than the abstract argument (derived from it) given in the proof of Theorem 5.4.

The essence of the argument can be given by the following proposition:[12]

PROPOSITION 5.10 (Cantor). *The set of infinite, binary sequences*

$$\Delta = \{(a_0, a_1, \ldots) \mid \forall i.a_i = 0 \vee a_i = 1\}$$

is uncountable.

Proof: The proof is by reductio ad absurdum and it uses the standard diagonalization argument. Assuming the countability of $\Delta = \{s_n | s_n = (a_{n0}, a_{n1}, \ldots)\}$ there is a

[12] Taken from Moschovakis (1994, p. 10ff) who adds the remaining steps to get the uncountability of the real numbers in subsequent propositions.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

surjective function f with $f(n) = s_n = (a_{n0}, a_{n1}, \dots)$. Now, we construct a sequence $s' = (s'_0, s'_1, \dots)$ with $s'_i = \begin{cases} 1, & \text{if } a_{ii} = 0, \\ 0, & \text{if } a_{ii} = 1. \end{cases}$ Since s' is not in the range of f, we have a contradiction. □

Although this proof is constructive (the negated statement "¬ (Δ is countable)" is proven by reductio ad absurdum, that is, "(Δ is countable) → ⊥)," the constructive character is better highlighted if it is given in its positive form:[13]

PROPOSITION 5.11. *For every countable set of infinite, binary sequences*

$$\Gamma = \{(a_0, a_1, \dots) \mid \forall i.a_i = 0 \vee a_i = 1\}$$

there exists a sequence b such that $b \notin \Gamma$.

As we said, the *proof* of Proposition 5.10 is constructive, but, looking at the matter from the concrete constructive point of view, it might depend on whether the set Δ of *all* infinite, binary sequences can actually be constructed. In the presence of *exponentiation*, a constructive set formation principle "replacing" the power set axiom (Beeson, 1985, secs. 9.3 and 11.2), this is indeed possible. However, due to the constructive set-up, it collects, of course, "only" those sequences which are constructively definable.

This is analog to the situation we have in Frege structures. We can define the set of all infinite, binary sequences in the following form:

$$\Lambda := \{f \mid \forall i.\mathsf{N}(i) \to f\,i = 0 \vee f\,i = 1\}$$

Since $\forall i.\mathsf{N}(i) \to f\,i = 0 \vee f\,i = 1$ is an elementary formula we have:

PROPOSITION 5.12. $\mathsf{FON} \vdash \mathsf{Set}(\Lambda)$.

PROPOSITION 5.13. Λ *is uncountable within Frege structures, that is, there is* no *term t such that*

$$\mathsf{FON} \vdash (\forall n.\mathsf{N}(n) \to t\,n \in \Lambda) \wedge (\forall x.x \in \Lambda \to \exists n.\mathsf{N}(n) \wedge t\,n = x).$$

Proof. Assuming that there is a surjective function from N to Λ given as a term t we can construct the "sequence" $b := \mathsf{d}_{\mathsf{N}}\,1\,0\,(t\,x\,x)\,0$. Since b would be an element of Λ, but for n with $\mathsf{N}(n)$, $b\,n \neq t\,n\,n$, t cannot be surjective. Thus, Λ is not countable—*within Frege structures*. □

This result raises the question, why we can apply Cantor's first diagonalization method to get an "uncountable" set over N, but why we cannot apply Cantor's second diagonalization method to get sets of higher cardinality in general. The difference is easy to observe on the technical level: In the former case, we diagonalize over (the formalized) natural numbers; in particular, negative statements about natural numbers are elementary (and therefore innocent with respect to set formation). For the latter case, we would need to have, first, the possibility to define a power set operation, which is not available, as it is not expressable by an elementary relation. But it is worth noting that, if we restrict ourselves to sets only, the

[13] In fact, this is the same situation as for Euclid's proof of the "existence of infinitary many primes"—stressed by constructivists: Euclid only states the version: "Prime numbers are more than any assigned multitude of prime numbers." (Euclid, Elements, Book IX, Proposition 20). The proof—which is in fact only a schematic proof by example—does not use reductio ad absurdum.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

elementhood relation becomes elementary, and (Russell-style) diagonalization is possible as performed in the proof of Proposition 5.3, the nonexistence of the set of all set.

REMARK 5.14. *The result that* Λ *is not countable* within *Frege structures might be useful as a didactic example to illustrate Skolem's paradox. Based on the models of* TON *one can easily construct models of Frege structures which have a countable universe; thus the interpretation of* Λ *will be countable, too. However, the surjective function associated with it is not definable within* FON. *This is the usual way to explain Skolem's paradox for a first-order theory of the real numbers, with the difference that in* FON *we can actually present the countable model quite concretely, what is not possible for a theory of real numbers.*

§6. Conclusion. In its presented form, FON can be considered as the result of "repairing" Frege's original system in the *Grundgesetze* by restricting comprehension to elementary formulae. (More exactly, by allowing unrestricted comprehension, but restricting the (meaningful) element relation to sets formed by formulae equivalent to an elementary one.) The resulting theory is not very strong, but one can get stronger theories in the same spirit, for instance, by adding *levels of truth* (Cantini, 1996; Kahle, 2003)—as already remarked in Note 4—or by incorporating *supervaluation* (Cantini, 1996; Kahle, 2001)—as noted in Remark 3.8.

Despite its weakness, the usefulness of FON as mathematical theory is already stressed at other places. One can cite the embedding results for $\widehat{\mathsf{ID}}_1$ and EETJ, theories which allow to develop an interesting part of mathematics.

In this paper, we focused on the set theory which can be introduced, in a very elegant way, in Frege structures. Even if it is mathematically weak, it has the interesting property of admitting a universal set. Historically, the question which set construction principles are legitimate, and which not, is as old as set theory itself. It was most prominently articulated in Poincar's critique of Zermelo's axiomatization. In fact, more than Russell's paradox, it was Cantor's paradox which questioned the traditional logical formation of concepts: If a concept is defined, as soon as one has a procedure to determine whether an element belongs to it or not, the universal set exists trivially.[14] With this background, Hilbert (2005, p. 215) once posed the following questions as a marginal in one of his lecture notes: "Why is the totality of all sets not permissible? Why is the set of all real numbers a permissible collection?" (translated in Moore, 1948, p. 48.) It is an essential feature of our theory, that the question *why a set is permissible* can be answered: Based on the natural truth-theoretic axiomatization, the existence of a set follows from the underlying applicative (that is, essentially recursion-theoretic) structure.

§7. Acknowledgement. This paper is based on a plenary talk given at the *XIII Simposio Latinoamericano de Logica Matematica* held in Oaxaca, México. I thank the organizers for inviting me. The work was partially supported by the ESF research project *Dialogical Foundations of Semantics* within the ESF Eurocores program *LogICCC*, LogICCC/0001/2007 (funded by the Portuguese Science Foundation, FCT) and by the FCT-project "Hilbert's Legacy in the Philosophy of Mathematics," PTDC/FIL-FCI/109991/2009.

[14] See, for instance, the note Kahle (2009).

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

BIBLIOGRAPHY

Aczel, P. (1977). The strength of Martin Löf's intuitionistic type theory with one universe. In Miettinen, S., and Väänänen, J., editors. *Proceedings of the Symposiums on Mathematical Logic in Oulo 1974 and in Helsinki 1975*. Report No. 2 from the Dept. of Philosophy, Helsinki: University of Helsinki.

Aczel, P. (1980). Frege structures and the notion of proposition, truth and set. In Barwise, J., Keisler, H., and Kunen, K., editors. *The Kleene Symposium*, Amsterdam: North-Holland, pp. 31–59.

Beeson, M. (1985). *Foundations of Constructive Mathematics*. Ergebnisse der Mathematik und ihrer Grenzgebiete; 3.Folge, Bd. 6. Berlin: Springer.

Cantini, A. (1993). Extending the first-order theory of combinators with self-referential truth. *Journal of Symbolic Logic*, **58**(2), 477–513.

Cantini, A. (1996). *Logical Frameworks for Truth and Abstraction*, Vol. 135 of *Studies in Logic and the Foundations of Mathematics*. Amsterdam: North-Holland.

Church, A. (1974). Set theory with a universal set. In Henkin, L., editor. *Proceedings of the Tarski Symposium*, Vol. XXV of *Proceedings of Symposia in Pure Mathematics*. Providence, R.1.: American Mathematical Society, pp. 297–308.

Feferman, S. (1975). A language and axioms for explicit mathematics. In Crossley, J., editor. *Algebra and Logic*, Vol. 450 of *Lecture Notes in Mathematics*, Berlin: Springer, pp. 87–139.

Feferman, S. (1979). Constructive theories of functions and classes. In Boffa, M., van Dalen, D., and McAloon, K., editors. *Logic Colloquium 78*, Amsterdam: North–Holland, pp. 159–224.

Feferman, S., & Jäger, G. (1993). Systems of explicit mathematics with non-constructive μ-operator. Part I. *Annals of Pure and Applied Logic*, **65**(3), 243–263.

Feferman, S., & Jäger, G. (1996). Systems of explicit mathematics with non-constructive μ-operator. Part II. *Annals of Pure and Applied Logic*, **79**, 37–52.

Flagg, R., & Myhill, J. (1987a). An extension of Frege structures. In Kueker, D., Lopez-Escobar, E., and Smith, C., editors. *Mathematical Logic and theoretical computer science*, New York: Dekker, pp. 197–217.

Flagg, R., & Myhill, J. (1987b). Implication and analysis in classical Frege structures. *Annals of Pure and Applied Logic*, **34**, 33–85.

Forster, T. E. (1995). *Set Theory with a Universal Set: Exploring an Untyped Universe*. Number 31 in Oxford Logic Guides. Oxford, UK: Oxford University Press.

Glass, T. (1996). On power set in explicit mathematics. *Journal of Symbolic Logic*, **61**(2), 468–489.

Hayashi, S. & Kobayashi,S. (1995). 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.

Hilbert, D. (1905). *Logische Principien des mathematischen Denkens*. Vorlesung Sommersemester 1905, Ausarbeitung von Ernst Hellinger (Bibliothek des Mathematischen Instituts der Universität Göttingen).

Jäger, G. (1997). Power types in explicit mathematics? *Journal of Symbolic Logic*, **62**(4), 1142–1146.

Jäger, G. & Strahm, T. (1995). Second order theories with ordinals and elementary comprehension. *Archive for Mathematical Logic*, **34**(6), 345–375.

Kahle, R. (1999). Frege structures for partial applicative theories. *Journal of Logic and Computation*, **9**(5), 683–700.

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press

Kahle, R. (2001). Truth in applicative theories. *Studia Logica*, **68**(1), 103–128.
Kahle, R. (2003). Universes over Frege structures. *Annals of Pure and Applied Logic*, **119**(1–3), 191–223.
Kahle, R. (2007). *The Applicative Realm*, Vol. 40 of *Textos de Matemática*. Departamento de Matemática, Universidade de Coimbra. *Habilitationsschrift* at the *Fakultät für Informations- und Kommunikationswissenschaften*, Coimbra: *Universität Tübingen*.
Kahle, R. (2009). The universal set—a (never fought) battle between philosophy and mathematics. In Pombo, O., & Nepomuceno, Á., editors. *Lógica e Filosofia da Ciência*, Vol. 2 of *Colecção Documenta*. Lisbon: Centro de Filosofia das Ciências da Universidade de Lisboa, pp. 53–65.
Kahle, R., & Studer, T. (2001). Formalizing non-termination of recursive programs. *Journal of Logic and Algebraic Programming*, **49**(1–2), 1–14.
Libert, T. (2008). Positive Frege and its Scott-style semantics. *Mathematical Logic Quarterly*, **54**(4), 410–434.
Moore, G. H. (2002). Hilbert on the infinite: The role of set theory in the evolution of Hilbert's thought. *Historia Mathematica*, **29**, 40–64.
Moschovakis, Y. (1994). *Notes on Set Theory*. Undergraduate Texts in Mathematics. Berlin: Springer.
Oberschelp, A. (1973). *Set Theory over Classes*, Vol. 106 of *Dissertationes Mathematicae*. Warszawa: Instytut Matematyczny PAN
Scott, D. (1975). Combinators and classes. In Böhm, C., editor. *λ-Calculus and Computer Science Theory*, Vol. 37 of *Lecture Notes in Computer Science*. Berlin: Springer, pp. 1–26.
Strahm, T. (1996). On the proof theory of applicative theories. PhD Thesis, Institut für Informatik und angewandte Mathematik, Universität Bern.
Troelstra, A., & van Dalen, D. (1988). *Constructivism in Mathematics*, Vol. 2. Amsterdam: North-Holland.

CENTRIA AND DEPARTAMENTO DE MATEMÁTICA
UNIVERSIDADE NOVA DE LISBOA
P–2829-516 CAPARICA, PORTUGAL
E-mail: kahle@mat.uc.pt

https://doi.org/10.1017/S1755020310000407 Published online by Cambridge University Press