Reflections on reflections in explicit mathematics*

Gerhard Jäger Thomas Strahm†

In honor of Professor Wolfram Pohlers' 60th birthday

Abstract

We give a broad discussion of reflection principles in explicit mathematics, thereby addressing various kinds of universe existence principles. The proof-theoretic strength of the relevant systems of explicit mathematics is couched in terms of suitable extensions of Kripke-Platek set theory.

1 Introduction

The chief aim of this paper is to survey various reflection principles in the realm of Feferman's explicit mathematics. We will discuss explicit analogues of the classical notions of inaccessibility, Mahloness, and weak compactness. Proof-theoretically speaking, our explicit formulations will be seen to correspond to their recursive interpretations and their strength can be measured in terms of theories of admissible sets for a recursively inaccessible, recursively Mahlo, and Π_3 reflecting universe of sets, respectively.

Explicit mathematics was introduced by Feferman around 1975. The three landmark papers laying the foundations of the subject are Feferman [4, 5, 6]. With respect to classical and recursive interpretations of an explicit mathematics framework, Feferman [5] is of particular interest.

At the heart of our considerations in this paper are *universes*, which are a frequently studied concept in constructive mathematics at least since the work of Martin-Löf [23]. They were first discussed in the framework of

*Some of the following results have first been presented at the Workshop *Operations, Sets and Types*, Castiglioncello, Italy, October 3–6, 1998.

†Research of both authors partly supported by the Swiss National Science Foundation.

explicit mathematics in Feferman [7] in connection with his proof of Hancock's conjecture. Universes can be considered as types of types (or names), which are closed under previously recognized type formation operations, i.e., a universes *reflects* those operations. Thus, universes are closely related to reflection principles in classical and admissble set theory. For a survey of some of the relevant previous results on universes in explicit mathematics, see Jäger, Kahle, and Studer [18].

The plan of this paper is as follows. In Section 2 we set out the formal framework of explicit mathematics with universes. Section 3 presents Kripke-Platek set theory augmented by various forms of reflection. In Section 4 we give a short description of the ordinal notations we use; in particular, we discuss n-ary φ-functions. Section 5 contains a conceptual discussion on recursive and classical interpretations of explicit mathematics. In Section 6 we introduce the limit axiom into explicit mathematics and address the strength of the so-obtained theories. Section 7 is devoted to the Mahlo axiom. Finally, in Section 8, we elaborate on so-called 2-universes and their relationship to Π_3 reflection.

2 The formal framework of explicit mathematics

The systems of explicit mathematics which we will consider in the following are formulated in the second order language $\mathbb{L}$ for individuals and types. It comprises individual variables $a, b, c, f, g, h, u, v, w, x, y, z, \ldots$ as well as type variables $U, V, W, X, Y, Z, \ldots$ (both possibly with subscripts). $\mathbb{L}$ also includes the individual constants k, s (combinators), $\mathsf{p}, \mathsf{p_0}, \mathsf{p_1}$ (pairing and projections), 0 (zero), $\mathsf{s_N}$ (successor), $\mathsf{p_N}$ (predecessor), $\mathsf{d_N}$ (definition by numerical cases), and additional individual constants, called *generators*, which will be used for the uniform naming of types, namely nat (natural numbers), id (identity), co (complement), is (intersection), dom (domain), inv (inverse image), j (join), and ℓ, m, π (reflectors). There is one binary function symbol $\cdot$ for (partial) application of individuals to individuals. Further, $\mathbb{L}$ has unary relation symbols $\downarrow$ (defined), N (natural numbers), U (universes) as well as three binary relation symbols $\in$ (membership), $=$ (equality), and $\Re$ (naming, representation).

The *individual terms* $(r, s, t, r_1, s_1, t_1, \ldots)$ of $\mathbb{L}$ are built up from individual variables and individual constants by means of our function symbol $\cdot$ for application. In the following we often abbreviate $(s \cdot t)$ simply as (st), st or sometimes also $s(t)$; the context will always ensure that no confusion arises.

We further adopt the convention of association to the left so that $s_1 s_2 \ldots s_n$ stands for $(\ldots(s_1 \cdot s_2)\ldots s_n)$. We also set $t' := \mathsf{s_N} t$. Finally, we define general n tupling by induction on $n \geq 2$ as follows:

$$(s_1, s_2) := \mathsf{p} s_1 s_2 \qquad \text{and} \qquad (s_1, \ldots, s_{n+1}) := ((s_1, \ldots, s_n), s_{n+1}).$$

The *atomic formulas* of $\mathbb{L}$ are the expressions $\mathsf{N}(s)$, $s{\downarrow}$, $(s = t)$, $(U = V)$, $(s \in U)$, $\mathsf{U}(V)$, and $\Re(s, U)$; the *formulas* $(A, B, C, A_1, B_1, C_1, \ldots)$ of $\mathbb{L}$ are generated from the atomic formulas by closing against negation, disjunction, conjunction, as we as existential and universal quantification for individuals and types.

Since we work with a logic of partial terms, it is not guaranteed that all terms have values, and $s{\downarrow}$ is read as s *is defined* or s *has a value*. Moreover, $\mathsf{N}(s)$ says that s is a natural number, and the formula $\Re(s, U)$ is used to express that the individual s *represents* the type U or is a *name* of U.

In the following we often omit parentheses and brackets whenever there is no danger of confusion. Moreover, we frequently make use of the vector notation $\vec{U}$ and $\vec{s}$ for finite strings of type variables $U_1, \ldots, U_m$ and individual terms $s_1, \ldots, s_n$, respectively, whose length is not important or given by the context. The following table contains a useful list of abbreviations:

$$\begin{array}{rcl}
(s \simeq t) & := & s{\downarrow} \vee t{\downarrow} \,\rightarrow\, s = t, \\
(s \in \mathsf{N}) & := & \mathsf{N}(s), \\
(\exists x \in \mathsf{N}) A(x) & := & (\exists x)(x \in \mathsf{N} \,\wedge\, A(x)), \\
(\forall x \in \mathsf{N}) A(x) & := & (\forall x)(x \in \mathsf{N} \,\rightarrow\, A(x)), \\
(V \subset W) & := & (\forall x)(x \in V \,\rightarrow\, x \in W), \\
(s \,\dot{\in}\, t) & := & (\exists X)(\Re(t, X) \,\wedge\, s \in X), \\
(\exists x \,\dot{\in}\, s) A(x) & := & (\exists x)(x \,\dot{\in}\, s \,\wedge\, A(x)), \\
(\forall x \,\dot{\in}\, s) A(x) & := & (\forall x)(x \,\dot{\in}\, s \,\rightarrow\, A(x)), \\
\Re(s) & := & (\exists X)\Re(s, X), \\
\Re(\vec{r}, \vec{U}) & := & \Re(r_1, U_1) \wedge \ldots \wedge \Re(r_n, U_n),
\end{array}$$

where the vector $\vec{r}$ consists of the individual terms $r_1, \ldots, r_n$ and the vector $\vec{U}$ of the type variables $U_1, \ldots, U_n$.

All our systems of explicit mathematics will be formulated in Beeson's classical *logic of partial terms* (cf. Beeson [2] or Troelstra and Van Dalen [35]) for the individuals and classical logic with equality for the types. Observe that Beeson's formalization includes the usual strictness axioms.

Before turning to various reflection principles in explicit mathematics, we introduce the auxiliary theory $\mathsf{s\text{-}EETJ}$ which provides a framework for explicit elementary types with join. Actually, $\mathsf{s\text{-}EETJ}$ is a variant of the theory EETJ, introduced in Jäger, Kahle, and Studer [18], in which strict versions of the basic type existence axioms are used. For more about strictness in explicit mathematics see Jäger, Kahle, and Studer [18], and Jäger and Studer [21]. The nonlogical axioms of $\mathsf{s\text{-}EETJ}$ can be divided into the following groups:

I. Applicative axioms. These axioms formalize that the individuals form a partial combinatory algebra, that we have paring and projection and the usual closure conditions on the natural numbers plus definition by numerical cases.

(1) $\mathsf{k}ab = a$,

(2) $\mathsf{s}ab\downarrow \wedge \mathsf{s}abc \simeq ac(bc)$,

(3) $\mathsf{p_0}(a,b) = a \wedge \mathsf{p_1}(a,b) = b$,

(4) $0 \in \mathsf{N}$,

(5) $a \in \mathsf{N} \rightarrow a' \in \mathsf{N}$,

(6) $a \in \mathsf{N} \rightarrow a' \neq 0 \wedge \mathsf{p_N}(a') = a$,

(7) $a \in \mathsf{N} \wedge a \neq 0 \rightarrow \mathsf{p_N}a \in \mathsf{N} \wedge (\mathsf{p_N}a)' = a$,

(8) $a \in \mathsf{N} \wedge b \in \mathsf{N} \wedge a = b \rightarrow \mathsf{d_N}xyab = x$,

(9) $a \in \mathsf{N} \wedge b \in \mathsf{N} \wedge a \neq b \rightarrow \mathsf{d_N}xyab = y$.

As usual, from axioms (1) and (2), one derives a theorem about λ abstraction and a form of the recursion theorem.

II. Explicit representation and extensionality. The following axioms state that each type has a name, that there are no homonyms and that equality of types is extensional.

(1) $(\exists x)\Re(x,U)$,

(2) $\Re(a,U) \wedge \Re(a,V) \rightarrow U = V$,

(3) $(\forall x)(x \in U \leftrightarrow x \in V) \;\rightarrow\; U = V$.

III. Basic type existence axioms. In the following we provide a finite axiomatization of uniform elementary comprehension plus join.

Natural numbers

(1) $\Re(\mathsf{nat})$,

(2) $(\forall x)(x \,\dot{\in}\, \mathsf{nat} \;\leftrightarrow\; \mathsf{N}(x))$.

Identity

(3) $\Re(\mathsf{id})$,

(4) $(\forall x)(x \,\dot{\in}\, \mathsf{id} \;\leftrightarrow\; (\exists y)(x = (y, y)))$.

Complements

(5) $\Re(a) \;\leftrightarrow\; \Re(\mathsf{co}(a))$,

(6) $\Re(a) \;\rightarrow\; (\forall x)(x \,\dot{\in}\, \mathsf{co}(a) \;\leftrightarrow\; x \,\dot{\notin}\, a)$.

Intersections

(7) $\Re(a) \wedge \Re(b) \;\leftrightarrow\; \Re(\mathsf{is}(a, b))$,

(8) $\Re(a) \wedge \Re(b) \;\rightarrow\; (\forall x)(x \,\dot{\in}\, \mathsf{is}(a, b) \;\leftrightarrow\; x \,\dot{\in}\, a \wedge x \,\dot{\in}\, b)$.

Domains

(9) $\Re(a) \;\leftrightarrow\; \Re(\mathsf{dom}(a))$,

(10) $\Re(a) \;\rightarrow\; (\forall x)(x \,\dot{\in}\, \mathsf{dom}(a) \;\leftrightarrow\; (\exists y)((x, y) \,\dot{\in}\, a))$.

Inverse images

(11) $\Re(a) \;\leftrightarrow\; \Re(\mathsf{inv}(a, f))$,

(12) $\Re(a) \;\rightarrow\; (\forall x)(x \,\dot{\in}\, \mathsf{inv}(a, f) \;\leftrightarrow\; fx \,\dot{\in}\, a)$.

Joins

(13) $\Re(a) \wedge (\forall x \,\dot{\in}\, a)\Re(fx) \;\leftrightarrow\; \Re(\mathsf{j}(a, f))$,

(14) $\Re(a) \wedge (\forall x \,\dot{\in}\, a)\Re(fx) \;\rightarrow\; (\forall x)(x \,\dot{\in}\, \mathsf{j}(a, f) \;\leftrightarrow\; \Sigma(a, f, \mathsf{j}(a, f)))$.

In this last axiom, the formula $\Sigma(a, f, b)$ expresses that b names the disjoint union of f over a, i.e.

$$\Sigma(a, f, b) \ := \ (\forall x)(x \mathrel{\dot{\in}} b \leftrightarrow (\exists y, z)(x = (y, z) \wedge y \mathrel{\dot{\in}} a \wedge z \mathrel{\dot{\in}} fy)).$$

IV. Uniqueness of generators. These axioms essentially guarantee that different generators create different names (see the beginning of Section 2 for the definition of generators). To achieve this we have, for syntactically different generators r_0 and r_1 and arbitrary generators s and t:

(1) $r_0 \neq r_1$,

(2) $(\forall x)(sx \neq t)$,

(3) $(\forall x, y)(sx = ty \ \rightarrow \ s = t \wedge x = y)$.

In the original formulation of explicit mathematics, elementary comprehension is not dealt with by a finite axiomatization but directly as an infinite axiom scheme. If an $\mathbb{L}$ formula A is called *elementary* provided that it contains neither the relation symbol $\Re$ nor bound type variables, then we have the following result.

Theorem 1 *For every elementary formula $A(u, \vec{v}, \vec{W})$ with at most the indicated free variables there exists a closed term t so that one can prove in* $\mathsf{s\text{-}EETJ}$*:*

1. $\Re(\vec{w}, \vec{W}) \ \rightarrow \ \Re(t(\vec{v}, \vec{w}))$,

2. $\Re(\vec{w}, \vec{W}) \ \rightarrow \ (\forall x)(x \mathrel{\dot{\in}} t(\vec{v}, \vec{w}) \leftrightarrow A(x, \vec{v}, \vec{W}))$.

This theorem is first stated in Feferman and Jäger [9]; its proof is standard and left to the reader as an exercise. Join and uniqueness of generators are not needed for this argument.

In the following we employ two forms of induction on the natural numbers, type induction and formula induction. Type induction is the axiom

$$(\mathsf{T\text{-}I_N}) \quad (\forall X)(0 \in X \wedge (\forall x \in \mathsf{N})(x \in X \rightarrow x' \in X) \ \rightarrow \ (\forall x \in \mathsf{N})(x \in X)).$$

Formula induction, on the other hand, is the schema

$$(\mathbb{L}\text{-}\mathsf{I_N}) \qquad A(0) \wedge (\forall x \in \mathsf{N})(A(x) \rightarrow A(x')) \ \rightarrow \ (\forall x \in \mathsf{N})A(x)$$

for each $\mathbb{L}$ formula A.

From Feferman [6] we know that $\mathsf{EETJ} + (\mathsf{T}\text{-}\mathsf{I_N})$ is proof-theoretically equivalent to Peano arithmetic PA and to the system $\Sigma^1_1\text{-}\mathsf{AC}_0$ of second order arithmetic; $\mathsf{EETJ} + (\mathbb{L}\text{-}\mathsf{I_N})$ has the same proof-theoretic strength as $\Sigma^1_1\text{-}\mathsf{AC}$.[1] The proof of these two results can be easily adapted to $\mathsf{s\text{-}EETJ}$.

Theorem 2 *The theory* $\mathsf{s\text{-}EETJ} + (\mathsf{T}\text{-}\mathsf{I_N})$ *is proof-theoretically equivalent to* PA *and* $\Sigma^1_1\text{-}\mathsf{AC}_0$*; the theory* $\mathsf{s\text{-}EETJ} + (\mathbb{L}\text{-}\mathsf{I_N})$ *is proof-theoretically equivalent to* $\Sigma^1_1\text{-}\mathsf{AC}$.

The next step is to introduce the concept of a universe in explicit mathematics. To put it very simply: a universe is supposed to be a type which consists of names only and reflects the theory $\mathsf{s\text{-}EETJ}$. For a detailed formulation of the appropriate axioms we introduce some auxiliary notation and let $\mathcal{C}(W, a)$ be the closure condition which is the disjunction of the following $\mathbb{L}$ formulas:

(1) $a = \mathsf{nat} \,\vee\, a = \mathsf{id}$,

(2) $(\exists x)(a = \mathsf{co}(x) \,\wedge\, x \in W)$,

(3) $(\exists x, y)(a = \mathsf{is}(x, y) \,\wedge\, x \in W \,\wedge\, y \in W)$,

(4) $(\exists x)(a = \mathsf{dom}(x) \,\wedge\, x \in W)$,

(5) $(\exists x, f)(a = \mathsf{inv}(x, f) \,\wedge\, x \in W)$,

(6) $(\exists x, f)(a = \mathsf{j}(x, f) \,\wedge\, x \in W \,\wedge\, (\forall y \,\dot{\in}\, x)(fy \in W))$.

Thus the fixed point property $(\forall x)(\mathcal{C}(W, x) \leftrightarrow x \in W)$ states that W is a type which is closed under elementary comprehension and join in the strict sense.

V. Basic axioms for universes. These axioms state that universes consist of names only, satisfy the fixed point property imposed by $\mathcal{C}$ and are transitive in a certain sense.

Ontological axioms

(1) $\mathsf{U}(W) \,\wedge\, s \in W \;\rightarrow\; \Re(s)$,

(2) $\mathsf{U}(V) \,\wedge\, \mathsf{U}(W) \,\wedge\, (\exists x)(\Re(x, V) \wedge x \in W) \;\rightarrow\; V \subset W$.

Fixed points

(3) $\mathsf{U}(W) \;\rightarrow\; (\forall x)(\mathcal{C}(W, x) \,\leftrightarrow\, x \in W)$.

[1] $\Sigma^1_1\text{-}\mathsf{AC}_0$ is $\Sigma^1_1\text{-}\mathsf{AC}$ with induction on the natural numbers restricted to sets.

The second ontological axiom is a sort of transitivity axiom, stating, that the universe V is a subuniverse of the universe W provided that some name of V belongs to W.

EU is defined to be the theory, formulated in $\mathbb{L}$, which extends the system $\mathsf{s\text{-}EETJ}$ by these basic axioms for universes and type induction ($\mathsf{T\text{-}I_N}$). Observe, however, that within EU the existence of universes cannot be proved. Even if we could show that all elements of a type W are names and that W satisfies the required fixed point property, there would be no possibility in EU to conclude that $\mathsf{U}(W)$.

This is different in Jäger, Kahle, and Studer [18], where universes are introduced as a defined concept. In this article we also discussed the question of what sort of ordering principles can or should be imposed on universes, and it turned out that one must not ask for too much.

Universes – or, more precisely, names of universes – will play an important rôle in the following in formulating reflection principles for explicit mathematics.

3 Theories for admissible sets

It is often very illuminating to compare systems of explicit mathematics to theories for admissible sets (with the natural numbers as urelements). In this section we recall some basic ingredients of these set theories and say a few words about some important axioms. Full information and all missing details are supplied in Jäger [12, 14, 15].

Theories for admissible sets can conveniently be formulated in a language $\mathcal{L}^\star = \mathcal{L}_1(\in, \mathsf{N}, \mathsf{S}, \mathsf{Ad})$ which extends some standard first order language $\mathcal{L}_1$ with the usual vocabulary for all primitive recursive functions and relations by the membership relation symbol $\in$, the set constant N for the set of natural numbers and the unary relation symbols S and Ad for sets and admissible sets, respectively. The system KPu of Kripke-Platek set theory (above the natural numbers as urelements) has the following axioms:

1. *Ontological axioms.* They claim that (i) each object is either a natural number or a set, (ii) each admissible is transitive, contains the set N and reflects the Kripke-Platek axioms, (iii) the admissibles are linearly ordered.

2. *Number-theoretic axioms.* They comprise the usual axioms for all primitive recursive functions and relations.

3. *Kripke-Platek axioms.* They provide pairing, for any set a transitive superset, Δ_0 separation, and Δ_0 collection.

4. *Induction principles.* They consist of the schema ($\mathcal{L}^\star$-I_N) of complete induction on the natural numbers and the schema ($\mathcal{L}^\star$-$\mathsf{I}_\in$) of $\in$ induction, both for arbitrary formulas of $\mathcal{L}^\star$.

The theory KPu^r is the subsystem of KPu in which the induction schemata ($\mathcal{L}^\star$-I_N) and ($\mathcal{L}^\star$-$\mathsf{I}_\in$) are restricted to Δ_0 formulas; KPu^0 is KPu with ($\mathcal{L}^\star$-I_N) restricted to Δ_0 formulas and $\in$ induction omitted completely.

The standard models of KPu are the admissible sets which contain the set of natural numbers as element. If we add the limit axioms (L),

$$(\mathsf{L}) \qquad (\exists z)(\mathsf{Ad}(z) \wedge a \in z),$$

stating that any set is element of an admissible set, we obtain three new theories

$$\mathsf{KPi} := \mathsf{KPu} + (\mathsf{L}), \quad \mathsf{KPi}^r := \mathsf{KPu}^r + (\mathsf{L}), \quad \mathsf{KPi}^0 := \mathsf{KPu}^0 + (\mathsf{L}),$$

depending on the induction principles which are included. They deal with recursively inaccessible sets, i.e. admissible limits of admissibles.

A further strengthening is achieved if Π_2 reflection, which is proof-theoretically equivalent to Δ_0 collection over $\mathsf{KPu} - (\Delta_0$ collection), is upgraded to Π_2 reflection on admissibles. The Mahlo axioms (M) postulate

$$(\mathsf{M}) \qquad A(\vec{a}) \rightarrow (\exists z)(\mathsf{Ad}(z) \wedge \vec{a} \in z \wedge A^z(\vec{a}))$$

for all Π_2 formulas $A(\vec{u})$ whose parameters belong to the list $\vec{u}$. The extensions of KPu and its two subsystems KPu^r and KPu^0 are the theories

$$\mathsf{KPm} := \mathsf{KPu} + (\mathsf{M}), \quad \mathsf{KPm}^r := \mathsf{KPu}^r + (\mathsf{M}), \quad \mathsf{KPm}^0 := \mathsf{KPu}^0 + (\mathsf{M})$$

whose least standard model is the structure $L_\mu(\mathbb{N})$ with μ being the first recursively Mahlo ordinal.

Finally, there is the schema of Π_3 reflection which gives proof-theoretic strength far beyond that of (M). It demands

$$(\Pi_3\ \mathsf{Ref}) \qquad A(\vec{a}) \rightarrow (\exists z)(\mathsf{Ad}(z) \wedge \vec{a} \in z \wedge A^z(\vec{a})).$$

for any Π_3 formula $A(\vec{u})$ of $\mathcal{L}^\star$, again with all its parameters from the list $\vec{u}$; of course, it would be sufficient to ask for a transitive witness.

4 Ordinal notations

Most proof-theoretic ordinals which we explicitly mention in this paper are easily expressed by making use of a *n-ary* Veblen functions φ_n for all natural numbers n greater than 0. The usual Veblen hierarchy, generated by the *binary* function φ, starting off with the function $\varphi 0 \beta = \omega^\beta$, is well known from the literature, cf. Pohlers [25] or Schütte [30]. The *ternary* φ_3 function is obtained as a straightforward generalization of the binary case by defining $\varphi_3(\alpha, \beta, \gamma)$ inductively as follows:

(i) $\varphi_3(0, \beta, \gamma)$ is just $\varphi\beta\gamma$.

(ii) If $\alpha > 0$, then $\varphi_3(\alpha, 0, \gamma)$ denotes the γth ordinal which is strongly critical with respect to all functions $\lambda\xi, \eta.\varphi_3(\delta, \xi, \eta)$ for $\delta < \alpha$.

(iii) If $\alpha > 0$ and $\beta > 0$, then $\varphi_3(\alpha, \beta, \gamma)$ denotes the γth common fixed point of the functions $\lambda\xi.\varphi_3(\alpha, \delta, \xi)$ for $\delta < \beta$.

For example, $\varphi_3(1, 0, \alpha)$ is Γ_α, and more generally, $\varphi_3(1, \alpha, \beta)$ denotes a Veblen hierarchy over $\lambda\alpha.\Gamma_\alpha$. It is straightforward how to extend these ideas in order to obtain φ_n functions of all finite arities, and even further to Schütte's Klammersymbole [29].

Instead of $\varphi_3(\alpha, \beta, \gamma)$ we simply write $\varphi\alpha\beta\gamma$. Φ_0 is chosen to be the least ordinal greater than 0 which is closed under all functions φ_n for $n > 0$. This ordinal Φ_0 can also be written as $\Theta\Omega^\omega 0$ if one prefers to work in the context of ordinal notation systems based on the Aczel-Buchholz-Feferman Θ-functions. A full exposition of this approach can be found in Buchholz [3].

5 Classical and recursive interpretations of explicit mathematics

Explicit mathematics is a very flexible formalism which permits classical and recursive interpretations. This "feature" of explicit mathematics is the main theme of Feferman [5], but it is already inherent in the standard model constructions described, for example, in Feferman [4]. More recently, in Feferman [8] a so-called *operational set theory* is introduced, which is a partial adaptation of explicit mathematics notions to the set-theoretical framework. It provides a step towards Feferman's unfolding of set theory and deals with generalizations of small large cardinals and their interpretations in classical as well as admissible set theory.

Without going into details, we briefly repeat the main ideas underlying the classical and recursive interpretations of explicit mathematics:

(1) The types of $\mathbb{L}$ are *extensional*, all individual terms of $\mathbb{L}$, on the other hand, are regarded as *intensional objects*; some of the intensional individuals name extensional types.

(2) Because of self-application, any individual term also acts as a (possibly partial) intentional operation on the universe and, in this sense, is a representation of a (partial) extensional function on the universe.

(3) In the set-theoretic interpretations of explicit mathematics, we may choose, for example, some sufficiently large initial segment V_α of the cummulative hierarchy as universe of the individuals. Via some suitable coding mechanism, these individuals represent the ordinary set-theoretic functions belonging to V_α plus further partial functions from V_α to V_α. Let us write $\|s\|$ for the partial function represented by s. The application operation in these interpretations is so that the individual s applied to the individual t yields the value $\|s\|(t)$ if t belongs to the domain of $\|s\|$; otherwise, s applied to t is undefined.

(4) The universe of any recursion-theoretic interpretation of explicit mathematics is the set $\mathbb{N}$ of the natural numbers. Then we fix some standard indexing $\{e\}$ for $e = 0, 1, \ldots$ for the partial recursive functions on $\mathbb{N}$. Application of e to n is then simply treated as $\{e\}(n)$.

(5) Thus, in set-theoretic interpretations the individuals represent standard set-theoretic (partial) functions, whereas in recursion-theoretic functions the individuals code partial recursive functions on the natural numbers. From the set-theoretic perspective the join axioms are therefore a sort of replacement axioms, in contrast to the recursion-theoretic approach in which join corresponds to Δ_0 replacement or Δ_0 reflection; cf., e.g., the definition of Kleene's $\mathcal{O}$.

As mentioned above, the formalism and the axioms of explicit mathematics abstract from any particular interpretation. Nevertheless, set-theoretic and recursion-theoretic models do have a certain significance, especially for furthering the intuitive understanding of the various notions. Applied to our definition of universe in explicit mathematics, the remarks (1)–(5) have the following consequences:

(6) Typical universes in the set-theoretic V_α-interpretations are generated by initial segments V_κ so that κ is a regular cardinal number.

(7) In the recursion-theoretic world, universes correspond to admissible (i.e. recursively regular) sets above $\mathbb{N}$.

This is not surprising since the recursive analogue of a regular cardinal is that of an admissible ordinal.

Universes in explicit mathematics are also related to universes in Martin-Löf type theory, cf., e.g., Martin-Löf [23, 24], and to fixed points – not necessarily least fixed points – of positive arithmetic operators as studied, for example, in Feferman [7] and Jäger et al. [17].

6 The limit axiom

It is in the spirit of explicit mathematics to work with *uniform* versions of reflection principles. We do not simply formulate that there exists a type which reflects a certain property but use an appropriate reflector to denote such a type, uniformly in the relevant parameters. Actually, the reflectors provide us with canonical names of universes. In the following we write

$$\mathcal{U}(s) \ := \ (\exists X)(\Re(s, X) \wedge \mathsf{U}(X))$$

in order to have a compact notation for the fact that the individual term s is a name of a universe.

A very first and elementary form of reflecting the axioms of EETJ is to claim that we have "many" universes in the sense that each name of a type belongs to a universe,

$$(\mathsf{Lim}) \qquad \forall a(\Re(a) \ \rightarrow \ \mathcal{U}(\ell(a)) \wedge a \mathrel{\dot{\in}} \ell(a)).$$

Keeping the set-theoretic and recursion-theoretic interpretations of explicit mathematics in mind, it arises quite naturally that the canonical models of $\mathsf{EU} + (\mathsf{Lim})$ comprehend the notion of (recursive) inaccessibility; (recursive) regularity because of modelling EETJ, and limit of (recursive) regularity because of (Lim).

However, the induction principles available in $\mathsf{EU} + (\mathsf{Lim})$ are very weak and so the situation is analogous to that of the theory KPi^0 of iterated admissible sets: the least standard model of KPi^0 is the structure $L_\iota(\mathbb{N})$, ι the least recursively inaccessible ordinal, whereas its proof-theoretic strength is characterized by the famous Feferman-Schütte ordinal Γ_0.

Theorem 3 (Jäger, Kahle, Strahm) *The theory* $\mathsf{EU} + (\mathsf{Lim})$ *is a framework for predicative mathematics; its extension by formula induction on* N *goes beyond predicativity. More precisely, we have:*

1. *The theory* $\mathsf{EU} + (\mathsf{Lim})$ *is proof-theoretically equivalent to the theories* KPi^0, $\widehat{\mathsf{ID}}_{<\omega}$ *and* ATR_0*; it thus has the proof-theoretic ordinal* Γ_0*, i.e.* $\varphi 100$.

2. *The theory* $\mathsf{EU} + (\mathsf{Lim}) + (\mathbb{L}\text{-}\mathsf{I_N})$ *is proof-theoretically equivalent to the theories* $\mathsf{KPi}^0 + (\mathcal{L}^\star\text{-}\mathsf{I_N})$ *and* $\widehat{\mathsf{ID}}_{<\varepsilon_0}$*; it thus has the proof-theoretic ordinal* $\varphi 1 \varepsilon_0 0$.

Here $\widehat{\mathsf{ID}}_{<\omega}$ and $\widehat{\mathsf{ID}}_{<\varepsilon_0}$ are theories for iterated fixed points of positive arithmetic operators (cf. e.g. Feferman [7]), ATR_0 is Friedman's theory of arithmetic transfinite recursion (cf., e.g., Friedman [10] or Simpson [32]). For a proof of this theorem we refer to Kahle [22] and Strahm [33]; a construction in Jäger and Strahm [20] provides a natural model of $\mathsf{EU} + (\mathsf{Lim})$.

Feferman's famous theory T_0 was the starting point of explicit mathematics; it extends the theory $\mathsf{EETJ} + (\mathbb{L}\text{-}\mathsf{I_N})$ by the powerful principle of inductive generation (IG): an new generator i is added to $\mathbb{L}$, and for every type W named w and every binary relation R on W with name r there exists the type of the R-accessible elements of W and is named $\mathsf{i}(w, r)$; induction along $\mathsf{i}(w, r)$ is permitted for arbitrary formulas of $\mathbb{L}$. The restriction $(\mathsf{IG})\!\restriction$ of (IG) claims induction along $\mathsf{i}(w, r)$ for types only. So we set:

$$
\begin{aligned}
\mathsf{T}_0\!\restriction &:= \mathsf{EETJ} + (\mathsf{IG})\!\restriction + (\mathsf{T}\text{-}\mathsf{I_N}),\\
\mathsf{T}_0 &:= \mathsf{EETJ} + (\mathsf{IG}) + (\mathbb{L}\text{-}\mathsf{I_N}),\\
\mathsf{s}\text{-}\mathsf{T}_0\!\restriction &:= \mathsf{EU} + (\mathsf{IG})\!\restriction,\\
\mathsf{s}\text{-}\mathsf{T}_0 &:= \mathsf{EU} + (\mathsf{IG}) + (\mathbb{L}\text{-}\mathsf{I_N}).
\end{aligned}
$$

It is clear that $\mathsf{s}\text{-}\mathsf{T}_0\!\restriction$ has the same proof-theoretic strength as $\mathsf{T}_0\!\restriction$ and $\mathsf{s}\text{-}\mathsf{T}_0$ the same as T_0. In Jäger and Studer [21] it is shown that adding the limit axiom (Lim) to the theories $\mathsf{s}\text{-}\mathsf{T}_0\!\restriction$ and $\mathsf{s}\text{-}\mathsf{T}_0$ does not change their respective proof-theoretic strength. Together with the results in Feferman [4], Jäger [13] and Jäger and Pohlers [19] we thus obtain the following theorem.

Theorem 4 (Jäger, Studer) *Adding the limit axiom* (Lim) *to* $\mathsf{s}\text{-}\mathsf{T}_0\!\restriction$ *and* $\mathsf{s}\text{-}\mathsf{T}_0$ *does not increase the proof-theoretic strength of these theories. Hence we have:*

1. *The theory* $\mathsf{s}\text{-}\mathsf{T}_0\!\restriction + (\mathsf{Lim})$ *is proof-theoretically equivalent to* $\mathsf{T}_0\!\restriction$*, thus also to* KPi^r *and* $\Delta^1_2\text{-}\mathsf{CA}_0$.

2. *The theory* $\mathsf{s}\text{-}\mathsf{T}_0 + (\mathsf{Lim})$ *is proof-theoretically equivalent to* T_0*, thus also to* KPi *and* $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$.

The theory Δ^1_2-CA is, of course, the usual system of second order arithmetic with Δ^1_2 comprehension; Δ^1_2-CA_0 is Δ^1_2-CA with induction on the natural numbers restricted to sets, and (BI) stands for bar induction.

7 The Mahlo axiom

A next quality in reflection is provided by the Mahlo principle. In classical set theory an ordinal κ is called a *Mahlo ordinal* if it is a regular cardinal and if, for every normal function $\mathfrak{f}$ from κ to κ, there exists a regular cardinal μ less than κ so that $\{\mathfrak{f}(\xi) : \xi < \mu\} \subset \mu$. The statement that there exists a Mahlo ordinal is a powerful set existence axiom going beyond theories like ZFC. It also outgrows the existence of inaccessible cardinals, hyper inaccessibles, hyperhyperinaccessible and the like.

In explicit mathematics we try to stay as close as possible to the set-theoretic formulation: we simply replace regular cardinals by universes and employ the reflector m to obtain uniformity. In order to formulate our Mahlo axiom in a compact way, we have to introduce some handy abbreviations. The corresponding definitions of $[\Re]_n(f)$ and $[s]_n(f)$ for n a natural number greater than 0 will be set out in a more general form than needed for stating the Mahlo axiom; however, this general definition will be used below when we will introduce 2-universes and Π_3 reflection in explicit mathematics. We define the following formulas by induction on $n > 0$:

$$\begin{aligned}
[\Re]_1(f) &:= (\forall x)(\Re(x) \to \Re(fx)), \\
[\Re]_{n+1}(f) &:= (\forall x)([\Re]_n(x) \to [\Re]_n(fx)), \\
f \mathrel{\dot{\in}} [s]_1 &:= (\forall x \mathrel{\dot{\in}} s)(fx \mathrel{\dot{\in}} s), \\
f \mathrel{\dot{\in}} [s]_{n+1} &:= (\forall x \mathrel{\dot{\in}} [s]_n)(fx \mathrel{\dot{\in}} [s]_n).
\end{aligned}$$

In a nutshell, the formulas $[\Re]_n(f)$ and $f \in [s]_n$ express that the operation f represents a (total) type n operation on the collection of all names $\Re$ and the elements of s, respectively. For the Mahlo case below, we only need these definitions for $n = 1$; the case $n = 2$ will be relevant in the next section.

Obviously, $[\Re]_1(f)$ means that f maps names to names, and $f \mathrel{\dot{\in}} [s]_1$ says that f maps elements of (the type named by) s to elements of (the type named by) s. The Mahlo axiom can now be formulated as in Jäger and Strahm [20],

$$(\mathsf{Mah}) \qquad \begin{aligned} &(\forall a, f)(\Re(a) \wedge [\Re]_1(f) \;\to \\ &\qquad \mathcal{U}(\mathsf{m}(f,a)) \wedge a \mathrel{\dot{\in}} \mathsf{m}(f,a) \wedge f \mathrel{\dot{\in}} [\mathsf{m}(f,a)]_1). \end{aligned}$$

It states that for every name a and for every operation f from names to names the reflector m picks the name $\mathsf{m}(f,a)$ of a universe which contains a and is closed under f. Setzer [31] presents a related formulation in the framework of Martin-Löf type theory.

The theory $\mathsf{EU} + (\mathsf{Mah})$ is significantly stronger than $\mathsf{EU} + (\mathsf{Lim})$. To convey an idea why this is the case, we introduce by induction on the natural numbers n a sequence $\ell_0, \ell_1, \ldots$ of terms of $\mathbb{L}$ so that

$$\begin{aligned} \ell_0 &:= \lambda x.\mathsf{m}(\lambda y.y, x), \\ \ell_{n+1} &:= \lambda x.\mathsf{m}(\ell_n, x). \end{aligned}$$

An iteration of this process into the transfinite would be possible, but details are omitted here. Note that the Mahlo axiom (Mah) immediately gives us

$$\mathsf{EU} + (\mathsf{Mah}) \vdash (\forall a)(\Re(a) \rightarrow \mathcal{U}(\ell_0(a)) \wedge a \mathrel{\dot\in} \ell_0(a)),$$

telling us that the term ℓ_0 plays the same rôle in $\mathsf{EU} + (\mathsf{Mah})$ as the reflector ℓ in $\mathsf{EU} + (\mathsf{Lim})$. Furthermore, a trivial inductive argument also shows

$$\mathsf{EU} + (\mathsf{Mah}) \vdash \Re(a) \rightarrow \mathcal{U}(\ell_n(a)) \wedge a \mathrel{\dot\in} \ell_n(a),$$

$$\mathsf{EU} + (\mathsf{Mah}) \vdash \Re(a) \wedge b \mathrel{\dot\in} \ell_n(a) \rightarrow \ell_m(b) \mathrel{\dot\in} \ell_n(a)$$

for all natural numbers n and m less than n. This means that, for any name a, the term $\ell_1(a)$ names an inaccessible universe, the term $\ell_2(a)$ a hyperinaccessible universe, the term $\ell_3(a)$ a hyperhyperinacessible universe, and so on.

The exact proof-theoretic analysis of the two theories $\mathsf{EU}+(\mathsf{Mah})$ and $\mathsf{EU}+(\mathsf{Mah}) + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ is carried through in Jäger and Strahm [20] and Strahm [34]. The first article establishes that the ordinals $\varphi\omega00$ and $\varphi\varepsilon_0 00$ are respective upper bounds of the proof-theoretic ordinals of these theories; the latter paper proves that both bounds are sharp.

Theorem 5 (Jäger, Strahm) *1. The theory* $\mathsf{EU} + (\mathsf{Mah})$ *is proof-theoretically equivalent to the theories* KPm^0 *and* Σ^1_1-TDC_0*; it thus has the proof-theoretic ordinal* $\varphi\omega00$.

2. The theory $\mathsf{EU} + (\mathsf{Mah}) + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ *is proof-theoretically equivalent to the theories* $\mathsf{KPm}^0 + (\mathcal{L}^\star\text{-}\mathsf{I}_\mathsf{N})$ *and* Σ^1_1-TDC*; it thus has the proof-theoretic ordinal* $\varphi\varepsilon_0 00$.

The system Σ^1_1-TDC of second order arithmetic with Σ^1_1 transfinite dependent choice and its subsystem Σ^1_1-TDC_0, which is obtained by restricting induction on the natural numbers to sets, was introduced in Rüede [27]; for details see also Rüede [28].

Upper proof-theoretic bounds for the Mahlo axiom in the context of the theories $\mathsf{T}_0\!\restriction$ and T_0 are most naturally obtained via model constructions within theories for specific nonmonotone inductive definitions introduced in Jäger [16]. We will not discuss these in this paper, but refer to Jäger and Studer [21] for the following result.

Theorem 6 (Jäger, Studer) *1. The theory $\mathsf{T}_0\!\restriction + (\mathsf{Mah})$ is contained in the theory KPm^r.*

2. The theory $\mathsf{T}_0 + (\mathsf{Mah})$ is contained in the theory KPm.

Together with recent results of Tupailo [36] we can conclude that $\mathsf{T}_0\!\restriction + (\mathsf{Mah})$ is proof-theoretically equivalent to KPm^r and $\mathsf{T}_0 + (\mathsf{Mah})$ to KPm.

The Mahlo axiom imposes certain closure properties on the collection of all names. Accordingly, a name d of a universe represents a *Mahlo universe*, written as $\mathsf{mah}\text{-}\mathcal{U}(d)$, if these closure properties are satisfied with respect to d, i.e.

$$(\forall a, f)(a \,\dot{\in}\, d \wedge f \,\dot{\in}\, [d]_1 \rightarrow \\ \mathsf{m}(f,a) \,\dot{\in}\, d \wedge \mathcal{U}(\mathsf{m}(f,a)) \wedge a \,\dot{\in}\, \mathsf{m}(f,a) \wedge f \,\dot{\in}\, [\mathsf{m}(f,a)]_1).$$

We want to point out that in general we cannot deduce from the simpler property

$$(\forall a, f)(a \,\dot{\in}\, d \wedge f \,\dot{\in}\, [d]_1 \rightarrow \mathsf{m}(f,a) \,\dot{\in}\, d)$$

that the name d of a universe represents a Mahlo universe. If we only know that $a \,\dot{\in}\, d$ and $f \,\dot{\in}\, [d]_1$, then the Mahlo axiom (Mah) does not permit us to conclude that $\mathcal{U}(\mathsf{m}(f,a))$, $a \,\dot{\in}\, \mathsf{m}(f,a)$ or $f \,\dot{\in}\, [\mathsf{m}(f,a)]_1$. We will come back to similar problems in a broader context later.

8 2-Universes and Π_3 reflection

In Aczel and Richter [1] and Richter and Aczel [26] the two authors are interested in formulating constructive (recursive) analogues for large regular ordinals and to connect those to closure ordinals of monotone and nonmonotone inductive definitions. For this purpose the notion of 2-*regularity* turns out to be particularly interesting.

Let κ be a regular cardinal and $\mathfrak{F}$ a function from κ^κ to κ^κ. We say that $\mathfrak{F}$ is *κ-bounded* if for all $\mathfrak{g} \in \kappa^\kappa$ and for all $\xi < \kappa$ there exists an $\alpha < \kappa$ such that

$$(\forall \mathfrak{h} \in \kappa^\kappa)(\mathfrak{g}{\restriction}\alpha = \mathfrak{h}{\restriction}\alpha \;\rightarrow\; \mathfrak{F}(\mathfrak{g})(\xi) = \mathfrak{F}(\mathfrak{h})(\xi)).$$

Hence κ-boundedness is a sort of continuity condition stating that the value $\mathfrak{F}(\mathfrak{g})(\xi)$ is determined by less than κ values of $\mathfrak{g}$. An ordinal α is a *κ-witness* for $\mathfrak{F}$ if $0 < \alpha < \kappa$ and

$$(\forall \mathfrak{g} \in \kappa^\kappa)(\{\mathfrak{g}(\xi) : \xi < \alpha\} \subset \alpha \;\rightarrow\; \{\mathfrak{F}(\mathfrak{g})(\xi) : \xi < \alpha\} \subset \alpha).$$

Finally, if every κ-bounded function from κ^κ to κ^κ has a κ-witness, Aczel and Richter call κ a *2-regular* ordinal and prove that κ is 2-regular if and only if it is weakly compact.

The two papers mentioned above also give an analogous formulation of 2-regularity in terms of (recursion theory on) admissible sets. The so obtained *2-admissible* ordinals are then shown to coincide with the Π_3 reflecting ordinals.

In the following we adapt the notions of 2-regular and 2-admissible ordinal to the context of explicit mathematics and introduce so-called *2-universes* with the reflector π providing the uniformity denotation of the witnesses and propose the following definition, which, however, will turn out to be insufficient:

$$2\text{-}\widetilde{\mathcal{U}}(d) \;:=\; \begin{cases} \mathcal{U}(d) \;\wedge \\ (\forall a, f, g)(a \,\dot{\in}\, d \;\wedge\; f \,\dot{\in}\, [d]_1 \;\wedge\; g \,\dot{\in}\, [d]_2 \;\rightarrow \\ \qquad \pi(g,f,a) \,\dot{\in}\, d \;\wedge\; \mathcal{U}(\pi(g,f,a)) \;\wedge\; a \,\dot{\in}\, \pi(g,f,a) \;\wedge \\ \qquad f \,\dot{\in}\, [\pi(g,f,a)]_1 \;\wedge\; g \,\dot{\in}\, [\pi(g,f,a)]_2). \end{cases}$$

One should expect that any 2-universe is a Mahlo universe, even a hyper-Mahlo universe. To check this claim, we define by induction on the natural numbers n the following sequence $\mathsf{m}_0, \mathsf{m}_1, \ldots$ of terms of $\mathbb{L}$:

$$\mathsf{m}_0 \;:=\; \lambda f.\lambda x.\pi(\lambda y.y, f, x),$$

$$\mathsf{m}_{n+1} \;:=\; \lambda f.\lambda x.\pi(\mathsf{m}_n, f, x).$$

With m_0 taking over the part of the reflector m of the previous section, it is immediately verified in EU that $2\text{-}\widetilde{\mathcal{U}}(d)$ yields $\mathsf{mah}\text{-}\mathcal{U}(d)$. Also, if $2\text{-}\widetilde{\mathcal{U}}(d)$, then $\mathsf{m}_0 \,\dot{\in}\, [d]_2$. But is each 2-universe d also hyper-Mahlo (in the obvious natural sense)?

Assume $2\text{-}\widetilde{\mathcal{U}}(d)$ and pick $a \mathrel{\dot\in} d$ and $f \mathrel{\dot\in} [d]_1$. In order to prove that d is a hyper-Mahlo universe, we need a Mahlo universe in d which contains a and is closed under f. The canonical candidate for (a name of) this universe is $b := \mathsf{m}_1(f,a)$. Since $\mathsf{m}_0 \mathrel{\dot\in} [d]_2$ according to the remark above, we thus have

$$b \mathrel{\dot\in} d, \quad a \mathrel{\dot\in} b \quad \text{and} \quad f \mathrel{\dot\in} [b]_1.$$

But the problem is to show that b represents a Mahlo universe. What we know is that $\mathsf{m}_0 \mathrel{\dot\in} [b]_2$, but this is not sufficient for making b to a Mahlo universe: If $c \mathrel{\dot\in} b$ and $g \mathrel{\dot\in} [b]_1$, then $\mathsf{m}_0(g,c) \mathrel{\dot\in} b$. However, there seems to be no possibility to derive, for example, that $\mathcal{U}(\mathsf{m}_0(g,c))$. To achieve this, we would need $g \mathrel{\dot\in} [d]_1$, whereas only $g \mathrel{\dot\in} [b]_1$ is available.

An elegant way to overcome this and similar problems and to improve the clumsy definition of Mahlo universe at the end of Section 7 is to call for appropriate *strictness properties* of our reflectors. A related approach – name strictness – is studied in Jäger, Kahle, and Studer [18].

The strict variants of our limit axiom (Lim) and Mahlo axiom (Mah) consist of two pairs (ℓ.1) and (ℓ.2) and (m.1) and (m.2), respectively, the first part of each pair being the actual strictness assertion.

Strict limit axioms

(ℓ.1) $$\Re(\ell(a)) \rightarrow \mathcal{U}(\ell(a)) \wedge a \mathrel{\dot\in} \ell(a),$$

(ℓ.2) $$\Re(a) \rightarrow \Re(\ell(a)).$$

Strict Mahlo axioms

(m.1) $$\Re(\mathsf{m}(f,a)) \rightarrow \mathcal{U}(\mathsf{m}(f,a)) \wedge a \mathrel{\dot\in} \mathsf{m}(f,a) \wedge f \mathrel{\dot\in} [\mathsf{m}(f,a)]_1,$$

(m.2) $$\Re(a) \wedge [\Re]_1(f) \rightarrow \Re(\mathsf{m}(f,a)).$$

Clearly, $\mathsf{EU} + (\ell.1) + (\ell.2)$ proves (Lim), and $\mathsf{EU} + (\mathsf{m}.1) + (\mathsf{m}.2)$ proves (Mah). On the other hand, all natural structures which satisfy (Lim) or (Mah) also satisfy (ℓ.1) plus (ℓ.2) or (m.1) plus (m.2), respectively. As a consequence, all results stated in Section 6 and Section 7 remain true if (Lim) and (Mah) are replaced by $(\ell.1) + (\ell.2)$ and $(\mathsf{m}.1) + (\mathsf{m}.2)$.

One of the immediate advantages, not the most important one, of course, of the strictness of the reflector m is the possibility to give a simpler definition of Mahlo universe, since

$$\mathsf{mah}\text{-}\mathcal{U}(d) \leftrightarrow \mathcal{U}(d) \wedge (\forall a, f)(a \mathrel{\dot\in} d \wedge f \mathrel{\dot\in} d^{d} \rightarrow \mathsf{m}(a,f) \mathrel{\dot\in} d)$$

is provable in $\mathsf{EU} + (\mathsf{m}.1) + (\mathsf{m}.2)$. For the direction from right to left, we only hat to observe that $\mathsf{m}(a, f) \mathrel{\dot\in} d$ implies that $\mathsf{m}(a, f)$ is a name (since d is a universe) and then apply $(\mathsf{m}.1)$.

The real benefit of the strictness of reflectors is in connection with the reflector π. Following the pattern of the strict limit and Mahlo axioms, we now present the axioms for a strict 2-universe.

Strict 2-universe axioms

$$(\pi.1) \qquad \begin{array}{c} \Re(\pi(g,f,a)) \;\rightarrow \\ \mathcal{U}(\pi(g,f,a)) \wedge a \mathrel{\dot\in} \pi(g,f,a) \wedge f \mathrel{\dot\in} [\pi(g,f,a)]_1 \wedge g \mathrel{\dot\in} [\pi(g,f,a)]_2, \end{array}$$

$$(\pi.2) \qquad \Re(a) \wedge [\Re]_1(f) \wedge [\Re]_2(g) \;\rightarrow\; \Re(\pi(g,f,a)).$$

They make sure that the collection $\Re$ of the names has all the properties of a 2-universe.

The extension of EU by the axioms $(\pi.1)$ and $(\pi.2)$ is called $\mathsf{EU} + (2\text{-}\mathsf{Uni})$. The proof-theoretic analysis of this theory also reveals the correspondence between the axioms $(\pi.1)$ and $(\pi.2)$ of explicit mathematics and the schema of Π_3 reflection in admissible set theory. In a series of unpublished notes, the following theorem is proved.

Theorem 7 (Jäger, Strahm) *1. The theory* $\mathsf{EU} + (2\text{-}\mathsf{Uni})$ *is proof-theoretically equivalent to the theory* $\mathsf{KPi}^0 + (\Pi_3\text{-}\mathsf{Ref})$*; it thus has the proof-theoretic ordinal* Φ_0*, i.e.* $\Theta\Omega^{\omega}0$.

2. The theory $\mathsf{EU} + (2\text{-}\mathsf{Uni}) + (\mathbb{L}\text{-}\mathsf{I_N})$ *is proof-theoretically equivalent to the theory* $\mathsf{KPi}^0 + (\Pi_3\text{-}\mathsf{Ref}) + (\mathcal{L}^\star\text{-}\mathsf{I_N})$*; it thus has the proof-theoretic ordinal* $\Theta\Omega^{\varepsilon_0}0$.

In the sequel let us briefly sketch some of the crucial steps which are required in proving the first part of the above theorem. In order to establish the *upper bound* for $\mathsf{EU} + (2\text{-}\mathsf{Uni})$, we proceed via suitable theories of ordinals and non-monotone inductive definitions as in the corresponding treatment of $\mathsf{EU} + (\mathsf{Mah})$ in Jäger and Strahm [20]. The ordinal theory required now features a suitable form of Π_3 reflection. More precisely, we work with the strengthening $\mathsf{OP3}$ of OMA, where Π_2 reflection on admissible ordinals is replaced by Π_3 *reflection on admissible ordinals*. The embedding of $\mathsf{EU} + (2\text{-}\mathsf{Uni})$ into $\mathsf{OP3}$ is similar in spirit to the embedding of $\mathsf{EU} + (\mathsf{Mah})$ in OMA; it proceeds via a formalized model construction building in the new reflector π. Of course, the 2-universe axioms of $\mathsf{EU} + (2\text{-}\mathsf{Uni})$ are now validated by making use of the Π_3 reflection axioms of $\mathsf{OP3}$.

In a next step we have to compute ordinal-theoretic upper bounds for $\mathsf{OP3}$. The central idea is to reduce the system $\mathsf{OP3}$ to the family of theories OMA_n ($n \in \mathbb{N}$), where OMA_n is defined to be OMA plus Π_2 reflection on n-hyper-Mahlo ordinals; thus, OMA_0 is just OMA. The first step in this reduction uses a Tait-style reformulation of $\mathsf{OP3}$, so that a standard partial cut-elimination argument can be established. The main step then consists in showing that, for a suitable class of statements, derivability in $\mathsf{OP3}$ entails provability in the family of theories OMA_n ($n \in \mathbb{N}$). More percisely, it can be shown that a statement which has a quasi-normal derivation of length n in $\mathsf{OP3}$ can already been validated in the theory OMA_n.

The last step in the upper bound computation now consists in the analysis of the theories OMA_n. The details of this analyis are long and tedious, but essentially follow the pattern already used for the analysis of OMA. The upshot is that for the proof-theoretic treatment of OMA_n, the $(n+3)$-ary φ-function is sufficient, in much the same way as the ternary φ-function has been used in the case $n = 0$ for OMA. This concludes our brief sketch of the upper bound computation for $\mathsf{EU} + (2\text{-}\mathsf{Uni})$.

In order to show that the above-mentioned upper bounds are sharp, one has to carry through well-ordering proofs in $\mathsf{EU} + (2\text{-}\mathsf{Uni})$. The crucial observation is that the axiom $(2\text{-}\mathsf{Uni})$ enables us to prove that there are Mahlo universes, hyper-Mahlo universes, hyperhyper-Mahlo universes, and so on. By induction we can prove that

$$
\begin{aligned}
\mathsf{EU} + (2\text{-}\mathsf{Uni}) \vdash\ & \Re(a) \land [\Re]_1(f) \\
& \rightarrow \mathcal{U}(\mathsf{m}_n(f,a)) \land a \mathrel{\dot\in} \mathsf{m}_n(f,a) \land f \mathrel{\dot\in} [\mathsf{m}_n(f,a)]_1, \\
\mathsf{EU} + (2\text{-}\mathsf{Uni}) \vdash\ & \Re(a) \land [\Re]_1(f) \land b \mathrel{\dot\in} \mathsf{m}_n(f,a) \land g \mathrel{\dot\in} [\mathsf{m}_n(f,a)]_1 \\
& \rightarrow \mathsf{m}_m(g,b) \mathrel{\dot\in} \mathsf{m}_n(f,a)
\end{aligned}
$$

for all natural numbers n and m less than n. Consequently, any $\mathsf{m}_1(f,a)$, provided that $\Re(a)$ and $[\Re]_1(f)$, names a Mahlo universes, $\mathsf{m}_2(f,a)$ a hyper-Mahlo universe, $\mathsf{m}_3(f,a)$ a hyperhyper-Mahlo universe, and so on.

The existence of n-hyper-Mahlo universes for each $n \in \mathbb{N}$ allows one to derive transfinite induction for all initial segments of the ordinal Φ_0, i.e., the first ordinal > 0 which is closed under all n-ary φ-functions. The well-ordering proofs use and generalize the techniques developed in Strahm [34].

We close the sketch of the proof of Theorem 7 by mentioning that no techniques of impredicative proof theory are required for the analysis of these systems; hence they are metapredicative in the sense of Jäger [11].

To end this section, let us come back to the definition of 2-universe. When working within $\mathsf{EU} + (\text{2-}\mathsf{Uni})$, we propse the following definition:

$$2\text{-}\mathcal{U}(d) \;:=\; \begin{cases} \mathcal{U}(d) \wedge \\ (\forall a, f, g)(a \mathrel{\dot{\in}} d \wedge f \mathrel{\dot{\in}} [d]_1 \wedge g \mathrel{\dot{\in}} [d]_2 \;\rightarrow\; \Re(\pi(g, f, a))). \end{cases}$$

In virtue of $(\pi.1)$ we immediately realize that $2\text{-}\widetilde{\mathcal{U}}(d)$ is entailed by $2\text{-}\mathcal{U}(d)$; actually $\mathsf{EU} + (\text{2-}\mathsf{Uni})$ proves the equivalence of $2\text{-}\mathcal{U}(d)$ and $2\text{-}\widetilde{\mathcal{U}}(d)$. The property $(\pi.1)$ is also instrumental in showing that $2\text{-}\mathcal{U}(d)$ yields that d codes a universe which is Mahlo, hyper-Mahlo, hyperhyper-Mahlo, ..., as we would like it to be. Clearly, $\mathsf{EU} + (\text{2-}\mathsf{Uni})$ does not prove that there exists a d so that $2\text{-}\mathcal{U}(d)$.

9 Final remark

There is no principal problem to generalize the previous approach to n-universes for any natural number n. Also the step into the transfinite, i.e. the notion of α-universe for $\alpha \geq \omega$, should be possible and thus provide an explicit equivalent of Π_α in the admissible setting.

It is much more demanding to address the task of studying much stronger reflection principles, such as Π^1_1 reflection or strict Π^1_1 reflection, in explicit mathematics. In Feferman's operational set theory, the principles *Op reflection* and $\forall$-*Op reflection* indicate a possible direction; see Feferman [8]. As for (pure) explicit mathematics, strong reflection principles will be treated in future publications.

References

[1] Peter Aczel and Wayne Richter, *Inductive definitions and analogues of large cardinals*, Conference in Mathematical Logic, London '70 (Wilfried Hodges, ed.), Lecture Notes in Mathematics, vol. 255, Springer, Berlin, 1972, pp. 1–9.

[2] Michael J. Beeson, *Foundations of Constructive Mathematics: Metamathematical Studies*, Springer, Berlin, 1985.

[3] Wilfried Buchholz, *Normalfunktionen und konstruktive Systeme von Ordinalzahlen*, Proof Theory Symposium, Kiel 1974 (Justus Diller and Gert H. Müller, eds.), Lecture Notes in Mathematics, vol. 500, Springer, Berlin, 1975, pp. 4–25.

[4] Solomon Feferman, *A language and axioms for explicit mathematics*, Algebra and Logic (John N. Crossley, ed.), Lecture Notes in Mathematics, vol. 450, Springer, Berlin, 1975, pp. 87–139.

[5] ______, *Recursion theory and set theory: a marriage of convenience*, Generalized Recursion Theory II, Oslo 1977 (Jens E. Fenstad, Robin O. Gandy, and Gerald E. Sacks, eds.), North-Holland, Amsterdam, 1978, pp. 55–98.

[6] ______, *Constructive theories of functions and classes*, Logic Colloquium '78 (Maurice Boffa, Dirk van Dalen, and Kenneth McAloon, eds.), North-Holland, Amsterdam, 1979, pp. 159–224.

[7] ______, *Iterated inductive fixed-point theories: application to Hancock's conjecture*, The Patras Symposion (George Metakides, ed.), North-Holland, Amsterdam, 1982, pp. 171–196.

[8] ______, *Notes on Operational Set Theory, I. Generalizations of "small" large cardinal in classical and admissible set theory*, 2001, Notes.

[9] Solomon Feferman and Gerhard Jäger, *Systems of explicit mathematics with non-constructive μ-operator. Part II*, Annals of Pure and Applied Logic **79** (1996), no. 1, 37–52.

[10] Harvey Friedman, *Some systems of second order arithmetic and their use*, Proceedings of the International Congress of Mathematicians, Vancouver 1974, vol. 1, Canadian Mathematical Congress, 1975, pp. 235–242.

[11] Gerhard Jäger, *Metapredicative and explicit Mahlo: a proof-theoretic perspective*, Logic Colloquium 2000, To appear.

[12] ______, *Die konstruktible Hierarchie als Hilfsmittel zur beweistheoretischen Untersuchung von Teilsystemen der Mengenlehre und Analysis*, Ph.D. thesis, Mathematisches Institut, Univeristät München, 1979.

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

[14] ______, *The strength of admissibility without foundation*, The Journal of Symbolic Logic **49** (1984), no. 3, 867–879.

[15] ______, *Theories for Admissible Sets: A Unifying Approach to Proof Theory*, Studies in Proof Theory, Lecture Notes, vol. 2, Bibliopolis, Napoli, 1986.

[16] ______, *First order theories for nonmonotone inductive definitions: recursively inaccessible and Mahlo*, The Journal of Symbolic Logic **66** (2001), no. 3, 1073–1089.

[17] Gerhard Jäger, Reinhard Kahle, Anton Setzer, and Thomas Strahm, *The proof-theoretic analysis of transfinitely iterated fixed point theories*, The Journal of Symbolic Logic **64** (1999), no. 1, 53–67.

[18] Gerhard Jäger, Reinhard Kahle, and Thomas Studer, *Universes in explicit mathematics*, Annals of Pure and Applied Logic **109** (2001), no. 3, 141–162.

[19] Gerhard Jäger and Wolfram Pohlers, *Eine beweistheoretische Untersuchung von* (Δ^1_2-CA) + (BI) *und verwandter Systeme*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse **1** (1982), 1–28.

[20] Gerhard Jäger and Thomas Strahm, *Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory*, The Journal of Symbolic Logic **66** (2001), no. 2, 935–958.

[21] Gerhard Jäger and Thomas Studer, *Extending the system* T_0 *of explicit mathematics: the limit and Mahlo axioms*, Annals of Pure and Applied Logic **114** (2002), no. 1–3, 79–101.

[22] Reinhard Kahle, *Uniform limit in explicit mathematics with universes*, Tech. Report IAM-97-002, Institut für Informatik und angewandte Mathematik, Universität Bern, 1997.

[23] Per Martin-Löf, *An intuitionistic theory of types: predicative part*, Logic Colloquium '73 (Harvey E. Rose and John Sheperdson, eds.), North-Holland, Amsterdam, 1975, pp. 73–118.

[24] ______, *Intutionistic Type Theory*, Studies in Proof Theory, Lecture Notes, vol. 1, Bibliopolis, Napoli, 1984.

[25] Wolfram Pohlers, *Proof Theory: An Introduction*, Lecture Notes in Mathematics, vol. 1407, Springer, Berlin, 1989.

[26] Wayne Richter and Peter Aczel, *Inductive definitions and reflecting properties of admissible ordinals*, Generalized Recursion Theory (Jens E. Fenstad and Peter G. Hinman, eds.), North-Holland, Amsterdam, 1974, pp. 301–381.

[27] Christian Rüede, *Metapredicative Subsystems of Analysis*, Ph.D. thesis, Institut für Informatik und angewandte Mathematik, Univeristät Bern, 2000.

[28] ———, *The proof-theoretic analysis of Σ^1_1 transfinite dependent choice*, Annals of Pure and Applied Logic **122** (2003), no. 1–3, 195–234.

[29] Kurt Schütte, *Kennzeichnung von Ordnungszahlen durch rekursiv erklärte Funktionen*, Mathematische Annalen **127** (1954), 15–32.

[30] ———, *Proof Theory*, Studies in Logic and Foundations of Mathematics, vol. 146, Springer, Berlin, 1977.

[31] Anton Setzer, *Extending Martin-Löf type theory by one Mahlo universe*, Archive for Mathematical Logic **39** (2000), no. 3, 155–181.

[32] Stephen G. Simpson, *Subsystems of Second Order Arithmetic*, Perspectives in Mathematical Logic, Springer, Berlin, 1999.

[33] Thomas Strahm, *First steps into metapredicativity in explicit mathematics*, Sets and Proofs (S. Barry Cooper and John Truss, eds.), Cambridge University Press, Cambridge, 1999, pp. 383–402.

[34] ———, *Wellordering proofs for metapredicative Mahlo*, The Journal of Symbolic Logic **67** (2002), no. 1, 260–278.

[35] Anne S. Troelstra and Dirk van Dalen, *Constructivism in Mathematics*, vol. I, North-Holland, Amsterdam, 1988.

[36] Sergei Tupailo, *Realization of constructive set theory into explicit mathematics: a lower bound for impredicative Mahlo universe*, Annals of Pure and Applied Logic **120** (2003), no. 1–3, 165–196.

Address

Gerhard Jäger, Thomas Strahm
Institut für Informatik und angewandte Mathematik
Universität Bern
Neubrückstrasse 10, CH-3012 Bern, Switzerland
{`jaeger`, `strahm`}@`iam.unibe.ch`

May 30, 2005