METAPREDICATIVE AND EXPLICIT MAHLO: A PROOF-THEORETIC PERSPECTIVE

GERHARD JÄGER

Abstract. After briefly discussing the concepts of *predicativity*, *metapredicativity* and *impredicativity*, we turn to the notion of *Mahloness* as it is treated in various contexts. Afterwards the appropriate Mahlo axioms for the framework of explicit mathematics are presented. The article concludes with relating explicit Mahlo to certain nonmonotone inductive definitions.

§1. Introduction. More than 100 years ago Cantor developed the theory of infinite sets (Cantor's paradise). Shortly afterwards, Russell found his famous paradox, and, as a consequence, many mathematicians became very concerned about the foundations of mathematics, and the expression *foundational crisis* was coined.

To overcome this crisis, Hilbert proposed the program of *Beweistheorie* as a method of rescuing *Cantor's paradise*. A few years later, however, Gödel showed that Hilbert's program – at least in its original strong form – cannot work. Again, after only a short while a first new idea was brought in by Gentzen, and a break-through along these lines was obtained by his proof-theoretic analysis of first order arithmetic. Then, during the last decades, Gentzen's work has been extended to stronger and stronger subsystems of second order arithmetic and set theory, most prominently by the schools of Schütte and Takeuti, leading to what today is denoted as *infinitary* and *finitary proof theory*, respectively.

A position completely different from Hilbert's was taken by Brouwer who advocated the restriction of mathematics to those principles which could be justified on constructive grounds. Starting off from his pioneering work various "dialects" of constructive mathematics have been put forward (e.g. in the Netherlands directly following Brouwer and Heyting, the Russian form(s) of constructivism, Bishop's approach, Martin-Löf type theory, Feferman's explicit mathematics).

In a certain sense the research directions originating from Hilbert's and Brouwer's original ideas come together again under the heading of *reductive proof theory* which tries to justify classical theories and classical principles by reducing them to a (more) constructive framework. For further reading

Meeting

and detailed information about this topic we refer, for example, to Beeson [6], Feferman [17] and Troelstra and van Dalen [61].

§2. The predicative, impredicative and metapredicative. The general picture described so far is, however, oversimplified in that it leaves out many important intermediate approaches. A particularly interesting line of thought was initiated by Poincaré's conviction that many foundational problems are caused by making use of so called *impredicative* definitions.[1] On the other hand, he did not consider the use of classical logic as being critical.

Poincaré was followed by Weyl, and they focused on the *arithmetical foundations of mathematics* (using their own terminology): their starting point being the usual structure $(\mathbb{N}, \dots)$ of the natural numbers with the schema of complete induction; moreover, all predicatively definable subsets of $\mathbb{N}$ are permitted.

This informal Poincaré-Weyl program was later brought into precise mathematical and logical terms by Feferman; see Feferman [13]. A further guiding line is his attempt to answer the following question: *what is implicit in the structure of the natural numbers together with the principle of induction?*

During the sixties Feferman and Schütte independently characterized *predicative mathematics* and showed that the associated ordinal is the famous Feferman-Schütte ordinal Γ_0. They achieved this by employing autonomous progressions of theories or ramified systems of second order arithmetic; for details see, for example, Feferman [8] and Schütte [51]. The theories capturing exactly predicative mathematics have their least standard model at

$$(\mathbb{N},\, L_{\Gamma_0} \cap Pow(\mathbb{N}),\, \dots)$$

and are equivalent to the system of second order arithmetic $\mathsf{AUT}(\Pi^0_1) + (\mathsf{BR})$ which comprises autonomously iterated Π^0_1 comprehension and the bar rule. Starting off from the Feferman-Schütte notion of predicativity, we can now distinguish the following three collections of theories:

A. Predicatively reducible systems. They comprise all those theories which are (finitely) reducible to a predicative system, i.e. whose proof-theoretic ordinal is less than or equal to Γ_0.

Trivially, all predicative theories are predicatively reducible. On the other hand, the least standard model of a predicatively reducible but not predicative theory can well be of the form

$$(\mathbb{N},\, L_\alpha \cap Pow(\mathbb{N}),\, \dots) \quad \text{or} \quad (L_\alpha,\, \in,\, \dots)$$

for some $\alpha > \Gamma_0$. A typical such theory is the system Σ^1_1-AC of second order arithmetic with the Σ^1_1 axiom of choice; its proof-theoretic ordinal is the ordinal $\varphi\varepsilon_0 0 < \Gamma_0$, although its least standard model is only reached at the first nonrecursive ordinal ω_1^{CK}.

[1] The definition of a set S is called impredicative if it refers to a totality of sets to which S itself belongs to.

In most cases the proof-theoretic analysis of a predicatively reducible theory can be obtained by forms of partial cut elimination and consecutive asymmetric interpretations or use of Skolem operators.

Examples of further important predicatively reducible systems are the theory $\Sigma^1_1\text{-}\mathsf{AC} + (\mathsf{BR})$, Friedman's theory ATR_0 of arithmetic transfinite recursion (cf. e.g. [20]), Avigad fixed point theory FP_0 (cf. e.g. [5]), Feferman's theory $\widehat{\mathsf{ID}}_{<\omega}$ of finitely many iterated fixed point axioms (cf. e.g. [12]) and Jäger's theory KPi^0 for a recursively inaccessible universe without foundation (cf. e.g. [28]).

B. Impredicative systems. Traditionally, all theories which are not predicatively reducible have been subsumed under the heading *impredicative*. Such an approach, however, has some undesired consequences.

Friedman's theory ATR_0, mentioned above, has proof-theoretic ordinal Γ_0, thus is predicatively reducible. But for obtaining this result it is essential that in ATR_0 complete induction on the natural numbers is restricted to sets. If complete induction on the natural numbers is permitted for arbitrary formulas, then results of Friedman (see Simpson [53]) and Jäger [25] show that the corresponding theory, called ATR, is of proof-theoretic strength Γ_{ε_0}. As a consequence, ATR is not predicatively reducible.

Recall, in addition, that the schema of complete induction on the natural numbers is at the core of predicativity a la Poincaré, Weyl and Feferman. Thus the predicatively reducible theory ATR_0 would be shifted into the impredicative by adding a purely predicative principle.

Moreover, the proof-theoretic analysis of ATR requires more of less the same concepts and machinery as the proof-theoretic analysis of ATR_0. Hence also from the point of view of methods involved, a very sharp dividing line between ATR_0 and ATR seems out of place.

To overcome such atrocities, we suggest to use the proof-theoretic techniques involved as criterion for structuring the range of systems which are not predicatively reducible. This then leads to the following "definition" of impredicativity: *The collection of impredicative systems comprises all those theories whose proof-theoretic (ordinal) analysis requires the use of impredicative methods.*

Of course, this is far from being a formal definition since we refer to the notion of *impredicative method*, and it is nowhere exactly pinned down what that means. However, given a specific ordinal analysis of a theory, we are convinced that all proof-theorists would agree on whether this analysis is done via impredicative techniques or not.

Our experience shows that typical impredicative methods always refer to some sort of collapsing techniques and collapsing functions, either directly applied to infinitary proofs or to the ordinals assigned to proofs or to both.

The first system for whose proof-theoretic analysis such impredicative methods have been used is the famous first order theory ID_1 for non-iterated positive arithmetic inductive definitions. On the other hand, recent work reveals (see below) that also an alternative approach is possible.

C. Metapredicative systems. The division into predicatively reducible and impredicative systems provided so far leaves some space between these two collections which is filled by the so-called metapredicative systems introduced now: *The collection of metapredicative systems comprises all those theories which are not predicatively reducible and whose proof-theoretic analysis can be carried through without making use of any impredicative methods.*

This description of metapredicativity is as informal as that of impredicativity. Therefore it would be very interesting to answer the following two questions:

(1) Is there a formal counterpart of this informal notion?
(2) If so, what is the limit of metapredicativity?

A satisfactory answer to question (1) has not yet been given, thus also (2) is still open. A first and necessary step in the direction of learning more about metapredicativity certainly consists of

- analyzing typical metapredicative systems and
- identifying their structural properties.

Years ago theories like ATR – lying in strength between Γ_0 and ID_1 – have been considered as exceptional cases; today we know that many interesting systems can be found in this area.

§3. More about metapredicative systems. The first metapredicative system which called for attention was the above mentioned theory ATR with proof-theoretic ordinal Γ_{ε_0} and was considered to be a rather isolated phenomenon at the time of its analysis. It was only later that more and more related theories turned up, and the systematic approach to metapredicativity began with the proof-theoretic analysis of the transfinitely iterated fixed point theories $\widehat{\mathsf{ID}}_\alpha$ and $\widehat{\mathsf{ID}}_{<\beta}$ for (recursive) ordinals $\alpha \geq \omega$ and $\beta > \omega$ in Jäger, Kahle, Setzer and Strahm [32].

These first order systems extend Feferman's theories $\widehat{\mathsf{ID}}_n$ into the transfinite and are a very good tool for calibrating the initial part of metapredicativity. ATR, for example, is proof-theoretically equivalent to $\widehat{\mathsf{ID}}_\omega$. In addition, Jäger and Strahm [37] established the role of the schema of Σ^1_1 dependent choice in the contexts of ATR_0 and ATR and proved that $\mathsf{ATR}_0 + (\Sigma^1_1\text{-}\mathsf{DC})$ has the same proof-theoretic strength as $\widehat{\mathsf{ID}}_{<\omega^\omega}$ whereas $\mathsf{ATR}+(\Sigma^1_1\text{-}\mathsf{DC})$ is proof-theoretically equivalent to $\widehat{\mathsf{ID}}_{<\varepsilon_0}$.

By using this result we could also answer a question of Simpson [54] about the strength of some of his second order systems for transfinite Σ^1_1 and Π^1_1 induction:

$$\Sigma^1_1\text{-}\mathsf{TI}_0 + \Pi^1_1\text{-}\mathsf{TI}_0 \;\equiv\; \widehat{\mathsf{ID}}_{<\omega^\omega} \quad \text{and} \quad \Sigma^1_1\text{-}\mathsf{TI} + \Pi^1_1\text{-}\mathsf{TI} \;\equiv\; \widehat{\mathsf{ID}}_{<\varepsilon_0}.$$

Afterwards Strahm studied autonomously iterated fixed point theories; see Strahm [57]. In his article he also establishes the relationship between autonomous fixed point iteration and transfinite fixed point recursion in the spirit of Friedman's arithmetic transfinite recursion.

In the context of admissible set theory without foundation, we obtain an interesting metapredicative system, if complete induction on the natural numbers is added to the previously mentioned theory KPi^0,

$$\mathsf{KPi}^0 + (\text{full induction on } \mathbb{N}) \equiv \widehat{\mathsf{ID}}_{<\varepsilon_0}.$$

To obtain this result we first have to show in KPi^0 that for each set a there exists a least admissible containing a (cf. Jäger [31]). Moreover, because of full complete induction on $\mathbb{N}$, we have transfinite induction for all formulas up to each $\alpha < \varepsilon_0$ and can thus build hierarchies of admissibles of the same height. Then it is easy to embed $\widehat{\mathsf{ID}}_{<\varepsilon_0}$. The upper bound is established by an extension of the methods in Jäger [28]. It is interesting to see that these arguments are sensitive to the question of whether in KPi^0 the admissible sets are linearly ordered or not.

Later in this article the formalism of explicit mathematics will be (very) briefly introduced. However, for the reader already familiar with this approach we include some remarks about metapredicative systems of explicit mathematics.

In the following we write EETJ for the theory comprising of the basic first order axioms plus elementary comprehension and join. Furthermore, we have a so-called *limit axiom*,

$$(\mathsf{Lim}) \qquad (\forall x)\,[\,\Re(x) \rightarrow (x \mathrel{\dot\in} \ell x \wedge \mathcal{U}(x))\,]$$

stating that each name x of a type is contained in a universe named ℓx. Then $\mathsf{EETJ}+(\mathsf{Lim})$ plus complete induction on $\mathbb{N}$ for types is a predicatively reducible theory of strength Γ_0; for details see Marzetta [41] and Kahle [39]. According to Strahm [56], the addition of compete induction on $\mathbb{N}$ for arbitrary formulas gives a metapredicative theory, again proof-theoretically equivalent to $\widehat{\mathsf{ID}}_{<\varepsilon_0}$. A non-uniform version of the limit axiom (Lim) is considered in Marzetta [41].

Although all theories discussed so far are strictly weaker than *metapredicative Mahlo*,[2] we must point out that this is by no means the limit of the area of metapredicativity. Recent work of Jäger and Strahm has shown that all instances of Π_n reflection ($n < \omega$) can safely be added to KPi^0 without surpassing metapredicativity. This yields, among other things, a metapredicative justification for ID_1.

Axioms for explicit mathematics corresponding to the set-theoretic Π_n reflections have also been studied by Jäger and Strahm. And there are certain stability properties for higher order operations in explicit mathematics which all give rise to metapredicative theories and reach far beyond metapredicative Mahlo.

[2] This notion will be described below.

The least standard models of all these metapredicative systems are based on sets L_α for comparatively large ordinals α, and these theories formulate (represent) many aspects of these sets which are sufficient for mathematical practice. On the other hand, since the available induction principles are very restricted, their proof-theoretic strengths stays rather low. So we can summarize some *charatieristics* of of metapredicative systems as follows:

- They cover a good deal of ordinary mathematics, for example large parts of analysis, discrete mathematics, category theory.
- They allow a philosophically careful (justified) proof-theoretic treatment from below in the sense that no collapsing techniques have to be used.
- In connection with metapredicative theories we can have *large (complicated) sets* but have to deal with *low consistency strength* only.

Further evidence for these remarks will now be given by looking closer at the role of Mahloness in various formal frameworks.

§4. A short survey of Mahloness. Mahlo axioms play an important role in present day proof theory. They have been studied during the last years from different perspectives, and some of these will be sketched below. In a sense, Mahloness draws the borderline of the part of proof theory that is so well understood, that the interaction between different standpoints becomes clear.

The Mahlo axioms go back to Mahlo's pioneering work from around 1911; see Mahlo [40]. Today an ordinal α is called a *Mahlo ordinal* if and only if

$$(\forall f : \alpha \to \alpha)\,(\exists \beta \in \mathrm{Reg})\,[\,\beta < \alpha \,\wedge\, f : \beta \to \beta\,],$$

Reg denoting the class of all regular cardinal numbers. The least Mahlo ordinal M_0 outgrows all inaccessible, hyperinaccessible, ... ordinals. M_0 cannot be reached from below by any sort of iteration of inaccessibility.

The usual approach of obtaining the *recursive analogues* of classical cardinal numbers also was applied to Mahlo ordinals by simply replacing

- *regular cardinals* by *admissible sets* and
- *arbitrary* functions by *recursive* functions.

Following the tradition in recursion theory, we will not directly use the corresponding reformulation of the definition of Mahlo ordinal as above, but instead work in the context of admissible set theory.

Let KPu^0 be the system of *Kripke-Platek set theory* above the natural numbers as urelements with induction on the natural numbers restricted to Δ_0 formulas and without $\in$ induction.[3] In this context, the recursive version of Mahlo can be characterized via Π_2 reflection on admissibles.

[3] The omission of $\in$ induction will be crucial for obtaining metapredicative systems; it has no effect, of course, if transitive standard models are considered.

Accordingly, KPm^0 is defined to be the set theory which extends KPu^0 by the following schema

$$A(\vec{u}) \;\rightarrow\; (\exists x)[\mathsf{Ad}(x) \,\wedge\, \vec{u} \in x \,\wedge\, A^x(\vec{u})]$$

for all Π_2 formulas $A(\vec{u})$ with the parameters show. An ordinal α is called *recursively Mahlo* if L_α, or more precisely the structure $(\mathbb{N}, L_\alpha(\mathbb{N}), \in, \dots)$, is a model of KPm^0, and we write μ_0 for the least recursively Mahlo ordinal.

Now we are going to mention two alternative characterizations of μ_0. The first one is in terms of Gandy's *superjump* $\mathbb{S}$, introduced in Gandy [21]. The superjump is an important type 3 functional whose associated closure ordinal has been studied in Aczel and Hinman [4] and Harrington [22].

THEOREM 1 (Aczel, Harrington, Hinman). *μ_0 is the least ordinal which is not recursive in Gandy's superjump $\mathbb{S}$, i.e.*

$$\mu_0 \;=\; \omega_1^{\mathbb{S}}.$$

For the next description of μ_0 we turn to nonmonotone inductive definitions. Let Φ be some arbitrary operator which maps the power set $Pow(\mathbb{N})$ to itself. Then Φ can be considered as (inducing) an inductive definition whose stages are introduced by recursion on the ordinals as follows:

$$I_\Phi^\alpha \;:=\; I_\Phi^{<\alpha} \cup \Phi(I_\Phi^{<\alpha}) \quad \text{and} \quad I_\Phi^{<\alpha} \;:=\; \bigcup\{I_\Phi^\xi : \xi < \alpha\}.$$

Obviously we have $I_\Phi^\alpha \subset I_\Phi^\beta$ for $\alpha \leq \beta$. A simple cardinality argument thus implies the existence of a least ordinal α such that I_Φ^α and $I_\Phi^{<\alpha}$ are identical. This ordinal is often called the *closure ordinal* of Φ and denoted by $||\Phi||$,

$$||\Phi|| \;:=\; \text{least } \alpha \text{ such that } I_\Phi^{<\alpha} \;=\; I_\Phi^\alpha.$$

Correspondingly, if $\mathcal{C}$ is a collection of operators, then the closure ordinal $||\mathcal{C}||$ of $\mathcal{C}$ is defined to be the ordinal $\sup\{\, ||\Phi|| : \Phi \in \mathcal{C}\,\}$.

An interesting way of combining two operators Φ and Ψ was introduced in Richter [47]. A new *combined operator* $[\Phi, \Psi]$ is generated from Φ and Ψ by setting for all subsets X of $\mathbb{N}$:

$$[\Phi,\Psi](X) \quad := \quad \begin{cases} \Phi(X) & \text{if} \quad \Phi(X) \not\subset X, \\ \Psi(X) & \text{if} \quad \Phi(X) \subset X. \end{cases}$$

Some further notation: let $[\,\text{POS-}\Pi^0_\infty, \Pi^0_1\,]$ be the collection of all combined operators whose first component is definable by an X-positive arithmetical formula and whose second component by a (not necessarily X-positive) Π^0_1 formula. Then Richter [47] contains the following theorem.

THEOREM 2 (Richter). $\quad \mu_0 \;=\; ||\,[\,POS\text{-}\Pi^0_\infty, \Pi^0_1\,]\,||.$

Mahloness in explicit mathematics has a natural classical and recursion-theoretic interpretation. We omit its discussion now and refer to Section 6, in which we will review it in detail and give the exact formulations.

Mahlo axioms are presently also of much interest in connection with constructive set theories and constructive type theories. Older approaches towards constructive versions of Zermelo-Fraenkel set theory are due to Friedman, Myhill and Scott, among others, and deal with (sub)systems of ZF, but with intuitionistic logic instead of classical logic (cf. [42]).

More recently, Aczel, in a series of papers [1, 2, 3], propagates alternative systems of constructive set theory CZF which incorporate (constructive) variants of the usual set theoretic principles, although their consistency strength stays comparatively small. His work has been extended in Rathjen [45] so that the constructive versions of large cardinal axioms, including Mahlo, find their place. Crosilla [7] deals with an extension of CZF without foundation for inaccessibility.

CZF and its extensions use intuitionistic logic as well, and we can obtain their constructive justification by interpretations into *Martin-Löf type theory* MLTT, which provides a philosophically motivated framework for constructive reasoning. The axioms reflecting the idea of Mahlo sets in MLTT are originally due to Setzer [52].

Let us end this section with recapitulating some of the most important proof-theoretic results about Mahlo in its various settings.

A. Full recursive Mahlo. The ordinal analysis of the canonical formalization of full recursive Mahlo, i.e. of the set theory for Π_2 reflection on admissibles

$$\mathsf{KPm} \;:=\; \mathsf{KPm}^0 + (\text{full induction}),$$

has been given by Rathjen [43], making use of methods of traditional impredicative proof theory. The corresponding system $\mathsf{FID}([\text{POS-}\Pi^0_\infty, \Pi^0_1])$ for first order inductive definitions treating combined operators from $[\text{POS-}\Pi^0_\infty, \Pi^0_1]$ was studied in Jäger [23] and Jäger and Studer [38].

B. Full and metapredicative explicit Mahlo. Full and metapredicative explicit Mahlo will be introduced in Section 6. Then we also state the respective results concerning their proof-theoretic strength.

In a nutshell: full explicit Mahlo is obtained from Feferman's T_0 by adding the Mahlo axioms ($\mathsf{M1}$) and ($\mathsf{M2}$) formulated below; metapredicative explicit Mahlo is obtained from full explicit Mahlo by deleting inductive generation (for details see Section 6).

C. Constructive Mahlo. As mentioned above, the design and analysis of the first extension of Martin-Löf type theory with one Mahlo universe was given by Setzer. Related formalizations in MLTT and CZF and their treatment are due to Rathjen. In the context of Martin-Löf type theory a sort of Mahlo rule is formulated in Rathjen [46].

D. Metapredicative Mahlo in set theory. KPm^0 is a natural formalism for metapredicative Mahlo in a set-theoretic context. Even if the schema ($\mathcal{L}^*$-$\mathsf{I_N}$) of complete induction on the natural numbers for all formulas of the language $\mathcal{L}^*$ of KPm^0 is added, we do not leave the area of metapredicativity.

THEOREM 3 (Jäger and Strahm). *We have the following proof-theoretic ordinals for metapredicative Mahlo in set theory:*

$$|\mathsf{KPm}^0| \;=\; \varphi\omega 00 \quad \textit{and} \quad |\mathsf{KPm}^0 + (\mathcal{L}^*\text{-}\mathsf{I}_\mathsf{N})| \;=\; \varphi\varepsilon_0 00.$$

This theorem is proved in Jäger and Strahm [35] and Strahm [55]. The lower bounds are established by well-ordering proofs, the upper bounds by interpretations of the respective theories into suitable ordinal theories with fixed point operators; the treatment of those is via partial cut elimination and asymmetric interpretations.

§5. The basics of explicit mathematics. Explicit mathematics was introduced by Feferman around 1975. The three basic papers which illuminate explicit mathematics from various angles are Feferman [9, 10, 11]. The original aim of explicit mathematics was to provide a natural formal framework for Bishop-style constructive mathematics. Soon it turned out, however, that the range of applications of explicit mathematics is much wider and includes, for example, also the following subjects:

Reductive proof theory. Systems of explicit mathematics play an important role in studying the relationship between subsystems of analysis, subsystems of set theory and theories for inductive definitions and for the reduction of classical theories to constructively (better) justified formalisms.

Abstract recursion theory. Several of its basic first order features (λ abstraction, fixed point theorem) are recursion-theoretic in nature, which are not tied to any specific structure; it is also a good tool in developing a proof theory of higher order functionals (cf. e.g. Feferman and Jäger [18, 19] and Jäger and Strahm [36]).

Type systems. Flexible (polymorphic) type systems find a natural place in explicit mathematics; the most practically needed type constructs can be modeled in systems of low proof-theoretic strength (cf. e.g. Feferman [14], Jäger [30]).

Programming. Feferman [15, 16] deals with properties of functional programs; Studer [59, 58] employs explicit mathematics for foundational questions in object-oriented programming.

In the following we do not work with Feferman's original formulation of explicit mathematics but use instead the framework of theories of types and names introduced in Jäger [29]. Their general "ontology" can then be described as follows:

- *individuals* are explicitly given and can be interpreted as objects, operations, (constructive) functions, programs and the like;
- self-application is possible; we define new operations (terms) by means of principles such as λ abstraction and the fixed point theorem;
- induction is then often used in order to show that these new operations have the desired properties.

- *Types* are abstractly defined collections of operations; they have names and are addressed via these names.

The focus of explicit mathematics is on the explicit presentation of operations rather than their constructive justification; it is possible to be explicit without being constructive (and vice versa).

Explicit mathematics starts off from a language $\mathbb{L}$ of two sorts, those being individuals $(a, b, c, x, y, z, \dots)$ and types $(U, V, W, X, Y, Z, \dots)$. There are several constants $\mathsf{k}, \mathsf{s}, \mathsf{p}, \mathsf{p}_0, \mathsf{p}_1, \mathsf{p}_\mathsf{N}, \mathsf{s}_\mathsf{N}, \dots$ whose meaning will be explained later plus one binary function symbol Ap for application. *Terms* are generated from the individual variables and constants by this form of application,

$$\text{Terms } (r, s, t, \dots): \quad \text{variables} \quad | \quad \text{constants} \quad | \quad \mathsf{Ap}(s, t).$$

In the following we often abbreviate $\mathsf{Ap}(s, t)$ as $(s \cdot t)$ or simply as (st) or st. We also adopt the convention of association to the left so that $s_1 s_2 \dots s_n$ stands for $(\dots(s_1 s_2)\dots s_n)$. In addition, we often write $s(t_1, \dots, t_n)$ for $st_1 \dots t_n$. Further we put $t' := \mathsf{s}_\mathsf{N} t$ and $1 := 0'$.

In addition, we have two unary relation symbols $\downarrow$ and N where $r{\downarrow}$ and $\mathsf{N}(r)$ express that r is defined (has a value) and r is a natural number, respectively. The only further relation symbols of our language of explicit mathematics are the binary $=$ for equality between individuals and between types, $\in$ for elementhood of individuals in types and $\Re$ for the naming relation; if $\Re(r, U)$ then we say that the individual r represents (is a name of) the type U. Therefore we have the atomic formulas

$$r{\downarrow}, \quad \mathsf{N}(r), \quad r = s, \quad U = V, \quad \Re(r, U),$$

and from those our formulas are generated as usual. A formula is called *elementary* if it contains neither the relation symbol $\Re$ nor bound type variables.

Finally, our logic is the classical *Beeson-Feferman logic of partial terms* with equality in both sorts as described, for example, in Beeson [6] and Troelstra and van Dalen [61]. Since it is not guaranteed that terms have values, a *partial equality* $\simeq$ à la Kleene is introduced by

$$(s \simeq t) \;:=\; (s{\downarrow} \vee t{\downarrow}) \;\to\; (s = t).$$

To simplify the notation, we frequently also use the following abbreviations concerning the predicate N:

$$\begin{aligned}
t \in \mathsf{N} &:= \mathsf{N}(t), \\
t \notin \mathsf{N} &:= \neg\mathsf{N}(t), \\
(\exists x \in \mathsf{N})A &:= (\exists x)(x \in \mathsf{N} \wedge A), \\
(\forall x \in \mathsf{N})A &:= (\forall x)(x \in \mathsf{N} \to A), \\
t \in (\mathsf{N} \to \mathsf{N}) &:= (\forall x \in \mathsf{N})(tx \in \mathsf{N}), \\
t \in (\mathsf{N}^{k+1} \to \mathsf{N}) &:= (\forall x \in \mathsf{N})(tx \in (\mathsf{N}^k \to \mathsf{N})).
\end{aligned}$$

5.1. Basic theory BON of operations and numbers. BON was introduced in Feferman and Jäger [18]. The nonlogical axioms of BON formalize that the individuals form a partial combinatory algebra, that we have pairing and projection and the usual closure conditions on the natural numbers as well as definition by numerical cases. We divide the axioms into the following five groups:

I. Partial combinatory algebra.

(1) $\mathsf{k}ab = a$,
(2) $\mathsf{s}ab\downarrow \,\wedge\, \mathsf{s}abc \simeq (ac)(bc)$.

II. Pairing and projection.

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

III. Natural numbers.

(4) $0 \in \mathsf{N} \,\wedge\, (\forall x \in \mathsf{N})(x' \in \mathsf{N})$,
(5) $(\forall x \in \mathsf{N})(x' \neq 0 \wedge \mathsf{p_N}(x') = x)$,
(6) $(\forall x \in \mathsf{N})(x \neq 0 \,\rightarrow\, \mathsf{p_N}x \in \mathsf{N} \wedge (\mathsf{p_N}x)' = x)$.

IV. Definition by numerical cases.

(7) $a \in \mathsf{N} \wedge b \in \mathsf{N} \,\wedge\, a = b \,\rightarrow\, \mathsf{d_N}uvab = u$,
(8) $a \in \mathsf{N} \wedge b \in \mathsf{N} \,\wedge\, a \neq b \,\rightarrow\, \mathsf{d_N}uvab = v$.

V. Primitive recursion on N.

(9) $f \in (\mathsf{N}^2 \rightarrow \mathsf{N}) \,\wedge\, a \in \mathsf{N} \,\rightarrow\, \mathsf{r_N}fa \in (\mathsf{N} \rightarrow \mathsf{N})$,
(10) $f \in (\mathsf{N}^2 \rightarrow \mathsf{N}) \,\wedge\, a \in \mathsf{N} \,\wedge\, b \in \mathsf{N} \,\wedge\, h = \mathsf{r_N}fa \,\rightarrow$
$h0 = a \,\wedge\, h(b') = fb(hb)$.

There are two crucial principles which follow already from the the axioms of a partial combinatory algebra, i.e. from axioms (1) and (2) of BON: λ *abstraction* and the *fixed point (recursion) theorem*. These are of course standard results which have been discussed in the relevant literature a long time ago; cf. e.g. Beeson [6], Feferman [9] or Troelstra and van Dalen [61].

The existence of $\mathsf{r_N}$ as claimed in (9) and (10) surely allows us to introduce representing terms for all primitive recursive functions. The defining equations and totality of (the representing terms for) these functions are derivable in BON.

In view of the availability of the fixed point theorem in BON one might even suspect that $\mathsf{r_N}$ and the axioms (9) and (10) are superfluous. Unfortunately, this is only the case if sufficiently strong induction principles are available. Actually, all induction principles formulated below would suffice. Nevertheless we decided to include (9) and (10) since these two axioms belong to the now "official" formulation of BON.

5.2. Basic axioms about types. Our next step is to formulate some basic axioms about types and their names, which will be included in all our further systems of explicit mathematics. We first claim that each type has a name, that types with the same name are identical and that the equality of types is extensional.

Naming and extensionality axioms (N&E)**.**

(11) $(\forall X)(\exists a)\Re(a, X)$,
(12) $\Re(a, X) \wedge \Re(a, Y) \;\rightarrow\; X = Y$,
(13) $(\forall a)(a \in X \leftrightarrow a \in Y) \;\leftrightarrow\; X = Y$.

Our systems of explicit mathematics combine intensionality and extensionality: on the level of types we are extensional, and types may be considered as objects in a Platonistic universe, given by abstract definitions. On the other hand, on the level of names we are intensional, and names have to be concretely given (introduced) terms. This idea also manifests itself by the following treatment of elementary comprehension.

Elementary comprehension (ECA)**.** Nowadays we prefer to work with a finite axiomatization (f-ECA) of elementary comprehension. That means that we add further constants to our language corresponding to several basic operations on types so that the following theorem can be proved.

THEOREM 4 (ECA). *For every elementary formula $A(x, \vec{y}, \vec{Z})$ with all its free variables indicated we can define a term t_A so that*

$$\Re(\vec{v}, \vec{V}) \;\rightarrow\; (\exists X)\,[\,X = \{x : A(x, \vec{u}, \vec{V})\} \;\wedge\; \Re(t_A(\vec{u}, \vec{v}), X)\,].$$

Here we assume that $\vec{u}$ is a finite string $u_1, \ldots, u_n$ of individual variables, $\vec{V}$ is a finite string $V_1, \ldots, V_n$ of type variables and $\Re(\vec{v}, \vec{V})$ is short for the conjunction of the formulas $\Re(v_i, V_i)$ (for $i = 1, \ldots, n$). This form of elementary comprehension is uniform in the individual and type parameters of the formula involved.

Join (J)**.** The join (J) is the way in which explicit mathematics treats disjoint unions. Suppose that we have a type A and an operation f which maps each element x of A to the name fx of a type, say, B_x. Then we write $\Sigma\{fx : x \in A\}$ for the disjoint union of the types B_x, indexed by A.

We want a uniform formulation of join and therefore choose a new constant j which names the intended disjoint union depending on a name of the index type and the operation from this index type to names; hence the axiom (J) can be written as

$$\Re(a, A) \;\wedge\; (\forall x \in A)(\exists X)\Re(fx, X) \;\rightarrow$$
$$(\exists Y)\,[\,Y = \Sigma\{fx : x \in A\} \;\wedge\; \Re(\mathsf{j}(a, f), Y)\,].$$

This finishes the description of the basic type-theoretic axioms of explicit mathematics. From now on we will write EETJ (elementary explicit typing with join) for the extension of BON by the naming and extensionality axioms (N&E), elementary comprehension (ECA) and join (J),

$$\mathsf{EETJ} \quad := \quad \mathsf{BON} + (\mathsf{N\&E}) + (\mathsf{f\text{-}ECA}) + (\mathsf{J}).$$

5.3. Induction in explicit mathematics. There are many induction principles which are studied in the context of explicit mathematics. In the following we confine ourselves here to type induction and $\mathbb{L}$ induction with respect to the natural numbers. The first is the axiom

$$(\mathsf{T}\text{-}\mathsf{I}_{\mathsf{N}}) \qquad (\forall X)[0 \in X \ \wedge \ (\forall x \in \mathsf{N})(x \in X \rightarrow x' \in X) \ \rightarrow \ (\forall x \in \mathsf{N})(x \in X)].$$

Obviously, type induction is a subcase of the schema of $\mathbb{L}$ induction stating for all formulas A of $\mathbb{L}$

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

Weaker forms of complete induction on the natural numbers, referring to the first order part of explicit mathematics, are studied at length in Jäger and Feferman [18].

5.4. Marriage of convenience. There are two main roads leading to models of explicit mathematics: one, in which the individuals are interpreted as (codes of) partial functions in the sense of classical set theory, and a second, which restricts itself to (codes of) partial *recursive* functions for dealing with the individuals. Details about such model constructions can be found, for example, in Feferman [11].

If we write $\mathrm{Gen}(V_{\aleph_1})$ and $\mathrm{Gen}(L_{\omega_1^{CK}})$ for the set-theoretic and recursion-theoretic model generated from the structure $(V_{\aleph_1}, \in)$ and $(L_{\omega_1^{CK}}, \in)$, respectively, the following observation can be easily established:

$$\mathrm{Gen}(V_{\aleph_1}) \models \mathsf{EETJ} + (\mathbb{L}\text{-}\mathsf{I}_{\mathsf{N}}) \quad \text{and} \quad \mathrm{Gen}(L_{\omega_1^{CK}}) \models \mathsf{EETJ} + (\mathbb{L}\text{-}\mathsf{I}_{\mathsf{N}}).$$

The (full) set-theoretic and the recursion-theoretic interpretations of explicit mathematics are connected by what Feferman [10] calls a *marriage of convenience*. Consider one of the usual extensions S of EETJ studied so far and a formula A provable in S. Then we may interpret A in the full set-theoretic model(s) of S – thus yielding the classical meaning $A^{(\mathsf{set})}$ of A – and in the recursion-theoretic models of S for obtaining the recursive reflection $A^{(\mathsf{rec})}$ of the classical assertion $A^{(\mathsf{set})}$.

5.5. Universes in explicit mathematics. Universes were introduced into explicit mathematics in Feferman [12], Marzetta [41] and Jäger, Kahle and Studer [33] as a powerful method for increasing its expressive and proof-theoretic strength. Informally speaking, universes play a similar role in explicit mathematics as admissible sets in weak set theory and the sets V_κ (for regular cardinals κ) in full classical set theory; explicit universes are also related to universes in Martin-Löf type theory. More formally, universes in explicit mathematics are types which consist of names only and reflect the theory EETJ.

In the following we write $U \models \mathsf{EETJ}$ for the conjunction of the finitely many formulas of $\mathbb{L}$ which express that the type U validates all type-theoretic axioms of EETJ; see Jäger, Kahle and Studer [33] for the exact formulation.

Furthermore, the following shorthand notations are convenient:

$$\Re(a) := (\exists X)\Re(a, X),$$

$$a \mathrel{\dot\in} b := (\exists X)(\Re(b, X) \wedge a \in X),$$

$$\mathsf{U}(W) := W \models \mathsf{EETJ} \wedge (\forall x \in S)\Re(x),$$

$$\mathcal{U}(a) := (\exists X)\,[\,\mathsf{U}(X) \wedge \Re(a, X)\,].$$

Thus $\Re(a)$ means that the individual a names some type; the formula $a \mathrel{\dot\in} b$ expresses that the individual a is an element of the type named by b; $\mathsf{U}(W)$ and $\mathcal{U}(a)$ say that the type W is a universe and the individual a a name of a universe, respectively.

The first important axiom in connection with universes, also studied in Jäger, Kahle and Studer [33], is the earlier mentioned *limit axiom*

$$(\mathsf{Lim}) \qquad (\forall x)\,[\,\Re(x) \rightarrow (x \mathrel{\dot\in} \ell x \wedge \mathcal{U}(x))\,]$$

which states that the individual ℓ uniformly picks for each name x of a type the name ℓx of a universe containing x.

For several proof-theoretic aspects of (Lim) see the above mentioned Jäger, Kahle and Studer [33]; the proof-theoretic strength of (Lim) in a metapredicative context is analyzed in Strahm [56].

§6. The Mahlo axioms in explicit mathematics. The limit axiom (Lim) together with EETJ describes the explicit analogue of (recursive) inaccessibility. Now we go an important step further and adapt the formulation of Mahloness to our explicit context. To simplify the notation we set

$$f \in (\Re \rightarrow \Re) := (\forall x)\,(\,\Re(x) \rightarrow \Re(fx)\,),$$

$$f \in (a \rightarrow a) := (\forall x)\,(\,x \mathrel{\dot\in} a \rightarrow fx \mathrel{\dot\in} a\,)$$

to express that the individual is an operation form names to names and the type named by a to itself, respectively. Then the *Mahlo axioms* are as follows:

$$(\mathsf{M1}) \qquad \Re(a) \wedge f \in (\Re \rightarrow \Re) \;\rightarrow\; \mathcal{U}(\mathsf{m}(a, f)) \wedge a \mathrel{\dot\in} \mathsf{m}(a, f),$$

$$(\mathsf{M2}) \qquad \Re(a) \wedge f \in (\Re \rightarrow \Re) \;\rightarrow\; f \in (\mathsf{m}(a, f) \rightarrow \mathsf{m}(a, f)).$$

m is a fresh individual constant for obtaining a formulation of these two axioms which is uniform in the name a and the operation f from names to names. From now on the theory $\mathsf{EETJ} + (\mathsf{M1}) + (\mathsf{M2})$ is usually written as $\mathsf{EETJ(M)}$.

Let M_0 be the first Mahlo cardinal and μ_0 the first recursively Mahlo ordinal. Then natural models of explicit Mahlo are generated from the full set-theoretic and the corresponding recursion-theoretic model of Mahloness.

$$\mathrm{Gen}(V_{M_0}) \models \mathsf{EETJ(M)} + (\mathbb{L}\text{-}\mathsf{I_N}) \quad \text{and} \quad \mathrm{Gen}(L_{\mu_0}) \models \mathsf{EETJ(M)} + (\mathbb{L}\text{-}\mathsf{I_N}).$$

The proof-theoretic analysis of $\mathsf{EETJ(M)}$ with type and formula induction on the natural numbers is carried through in Jäger and Strahm [35] and in

Strahm [55]. Related results can also be obtained for corresponding systems of explicit mathematics with intuitionistic logic.

THEOREM 5 (Jäger and Strahm). *We have the following two proof-theoretic equivalences:*

1. $\mathsf{EETJ}(\mathsf{M}) + (\mathsf{T}\text{-}\mathsf{I}_{\mathsf{N}}) \;\equiv\; \mathsf{KPm}^0$,
2. $\mathsf{EETJ}(\mathsf{M}) + (\mathbb{L}\text{-}\mathsf{I}_{\mathsf{N}}) \;\equiv\; \mathsf{KPm}^0 + (\mathcal{L}^*\text{-}\mathsf{I}_{\mathsf{N}})$.

All induction principles in these "metapredicative Mahlo" theories are restricted to the natural numbers. Later in this article stronger theories will be considered as well.

§7. Metapredicative Mahlo in second order arithmetic. This section is a short insertion turning to the problem of metapredicative Mahlo in the context of subsystems of second order arithmetic. The basic reference is Rüede's recent PhD thesis [48].

A subsystem of second order arithmetic which is proof-theoretically equivalent to KPm, i.e. KPm^0 with full $\in$ induction, is introduced and analyzed in Rathjen [44]. For obtaining systems of the same strength as KPm^0, Rüede had to proceed differently.

The role of universes is played in his approach by countable coded ω-models of $(\Sigma^1_1\text{-}\mathsf{DC})$, and for such universes M, the "elements" of M are the sets which can be written as projections of M, i.e. for all subsets X of the natural numbers we simply define

$$X \;\varepsilon\; M \;:=\; (\exists y)(X = (M)_y).$$

Countable coded ω-models of $(\Sigma^1_1\text{-}\mathsf{AC})$ would not suffice since for all limits λ of admissible ordinals, the set $L_\lambda \cap Pow(\mathbb{N})$ is a countable ω-model of $(\Sigma^1_1\text{-}\mathsf{AC})$, but not necessarily admissible.

Two schemas are central. The first is Π^1_2 *reflection on countable coded ω-models of* $(\Sigma^1_1\text{-}\mathsf{DC})$ and consists of

$$(\Pi^1_2\text{-}\mathsf{REF})^{(\Sigma^1_1\text{-}\mathsf{DC})} \qquad A(X) \;\rightarrow\; (\exists M)[X \;\varepsilon\; M \;\wedge\; M \models_\omega (\Sigma^1_1\text{-}\mathsf{DC}) \;\wedge\; A^M(X)]$$

for all Π^1_2 formulas $A(X)$ with the only free set variable being X; of course finite strings of set parameters could be permitted as well. The second important schema is Σ^1_1 *transfinite dependent choice*, consisting of

$$(\Sigma^1_1\text{-}\mathsf{TDC}) \qquad \begin{array}{l} \mathsf{WO}(\prec) \;\wedge\; (\forall x)(\forall X)(\exists Y)A(x,X,Y) \;\rightarrow \\ (\forall X)(\exists Z)\,[\,(Z)_0 = X \;\wedge\; (\forall x)(0 \prec x \rightarrow A(x,(Z)_{\prec\restriction x},(Z)_x,)\,] \end{array}$$

for all Σ^1_1 formulas $A(x,X,Y)$. As it turns out, $(\Pi^1_2\text{-}\mathsf{REF})^{(\Sigma^1_1\text{-}\mathsf{DC})}$ is equivalent to $(\Sigma^1_1\text{-}\mathsf{TDC})$ and has the desired proof-theoretic strength.

THEOREM 6 (Rüede). *The two schemas* $(\Pi^1_2\text{-}\mathsf{REF})^{(\Sigma^1_1\text{-}\mathsf{DC})}$ *and* $(\Sigma^1_1\text{-}\mathsf{TDC})$ *are equivalent over* ACA_0 *and have the same proof-theoretic strength as metapredicative Mahlo; i.e. we have*

1. $\mathsf{ACA}_0 + (\Pi^1_2\text{-}\mathsf{REF})^{(\Sigma^1_1\text{-}\mathsf{DC})} \ = \ \mathsf{ACA}_0 + (\Sigma^1_1\text{-}\mathsf{TDC})$,
2. $\mathsf{ACA}_0 + (\Sigma^1_1\text{-}\mathsf{TDC}) \ \equiv \ \mathsf{KPm}^0$.

For a proof of these two results see the above mentioned PhD thesis Rüede [48] or Rüede [49, 50].

§8. Mahlo beyond Feferman's T_0. 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): for every type A named a and every binary relation R on A with name r there exists the type of the R-accessible elements of A and is named $\mathsf{i}(a, r)$. So we set

$$\mathsf{T}_0 \quad := \quad \mathsf{EETJ} + (\mathsf{IG}) + (\mathbb{L}\text{-}\mathsf{I_N}).$$

Originally, T_0 was formulated within intuitionistic logic, but for some time classical logic has been used. The intuitionistic version of T_0 is called T^i_0 nowadays and provides an elegant framework for Bishop-style constructive mathematics. The constructive justification of T^i_0 is via a realizability interpretation.

The proof-theoretic strength of T_0 and T^i_0 is substantial and has been established in the following four articles: Jäger [27] presents a well-ordering proof for T^i_0; Feferman [11] shows that T_0 can be embedded into $(\Delta^1_2\text{-}\mathsf{CA}) + (\mathsf{BI})$; Jäger [24, 26] contain proofs that $(\Delta^1_2\text{-}\mathsf{CA}) + (\mathsf{BI})$ is contained in KPi; Jäger and Pohlers [34], finally, deals with the upper proof-theoretic limit of KPi.

THEOREM 7 (Feferman, Jäger and Pohlers). *The theory T_0 and its intuitionistic version T^i_0 possess the same proof-theoretic strength, which is determined by the following equivalences:*

$$\mathsf{T}^i_0 \ \equiv \ \mathsf{T}_0 \ \equiv \ (\Delta^1_2\text{-}\mathsf{CA}) + (\mathsf{BI}) \ \equiv \ \mathsf{KPi}.$$

Conceptually, T_0 and T^i_0 go beyond what is reachable via iterated monotone inductive definability. Both systems provide a first step towards nonmonotone inductive definitions. This becomes very perspicuous in the context of the new model constructions for explicit mathematics given the help of certain classes of nonmonotone inductive definitions in Jäger [23] and Jäger and Studer [38] and Studer [60].

Nonmonotone inductive definability is even more important if the Mahlo axioms $(\mathsf{M1})$ and $(\mathsf{M2})$ are added to T_0; call the resulting theory $\mathsf{T}_0(\mathsf{M})$ for simplicity. In the next section we will try to convey an idea how nonmonotone inductive definitions can be used for modeling $\mathsf{EETJ}(\mathsf{M})$ and $\mathsf{T}_0(\mathsf{M})$.

The rest of this section is dedicated to some recent results about the proof-theoretic strength of $\mathsf{T}_0(\mathsf{M})$, $\mathsf{T}^i_0(\mathsf{M})$ and other related theories. For the upper bound of $\mathsf{T}_0(\mathsf{M})$ we only have to refer to Jäger and Studer [38].

THEOREM 8 (Jäger and Studer). *The theory $\mathsf{T}_0(\mathsf{M})$ can be interpreted in the theory KPm, i.e. $\mathsf{T}_0(\mathsf{M}) \subset \mathsf{KPm}$.*

Recent work of Tupailo is concerned with providing realizability interpretations of subsystems of second order arithmetic and extensions of Aczel's system CZF into explicit mathematics. The article Tupailo [62] embeds a subsystem of second order arithmetic of the same strength as the familiar $\Delta^1_2\text{-}\mathsf{CA} + (\mathsf{BI})$ into T^i_0 whereas Tupailo [63] deals with the extension of CZF by a form of the Mahlo axiom (Mahlo) suited for the constructive setting.

In view of Rathjen [43] and some of his unpublished observations about constructive set theory it follows that the well-ordering proof for the proof-theoretic ordinal of KPm can be carried through in $\mathsf{CZF} + (\mathsf{Mahlo})$. Combining all these results we thus have the following theorem.

THEOREM 9. *The theories* $\mathsf{T}_0(\mathsf{M})$ *and* $\mathsf{T}^i_0(\mathsf{M})$ *have the same proof-theoretic strength as the theory* KPm,

$$\mathsf{CZF} + (\mathsf{Mahlo}) \ \equiv\ \mathsf{T}^i_0(\mathsf{M}) \ \equiv\ \mathsf{T}_0(\mathsf{M}) \ \equiv\ \mathsf{KPm}.$$

Setzer's approach to Mahloness within the framework of Martin-Löf type theory has also to be mentioned in this context; in [52] he studies a type theory of comparable strength.

§9. Modeling $\mathsf{EETJ}(\mathsf{M}) + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ and $\mathsf{T}_0(\mathsf{M})$. We conclude this survey with a brief description of how explicit Mahlo can be modeled in systems of nonmonotone inductive definitions of the same strength. We confine ourselves to the basic ideas; all further details concerning the theory $\mathsf{T}_0(\mathsf{M})$ for full recursive Mahlo can be found in Jäger and Studer [38]; the metapredicative variant is in Jäger and Strahm [35].

The treatment of the applicative part of $\mathsf{EETJ}(\mathsf{M}) + (\mathbb{L}\text{-}\mathsf{I}_\mathsf{N})$ and $\mathsf{T}_0(\mathsf{M})$ is as usual: the individuals are supposed to range over the set of natural numbers $\mathbb{N}$, we assume that $\{e\}$ for $e \in \mathbb{N}$ is a usual indexing of the partial recursive functions and let individual application be translated by setting

$$(e \bullet n) \ :\simeq\ \{e\}(n).$$

Then standard applications of the well-known S-m-n theorem provide natural numbers so that the axioms of BON are satisfied. The interpretation of types and names is more interesting. For this purpose we choose suitable formulas

$$\mathfrak{A}(X,a,b,c) \ \in\ \mathsf{POS}\text{-}\Pi^0_\infty \quad \text{and} \quad \mathfrak{B}(X,a,b,c) \ \in\ \Pi^0_1,$$

with the corresponding operators $\Phi_\mathfrak{A}$ and $\Phi_\mathfrak{B}$ and work with the combined operator $\Theta := [\Phi_\mathfrak{A}, \Phi_\mathfrak{B}]$ generated from $\Phi_\mathfrak{A}$ and $\Phi_\mathfrak{B}$. The sets I^α_Θ are the stages of the inductive definition induced by Θ, and I_Θ is defined as

$$I_\Theta \ :=\ \bigcup \{\, I^\alpha_\Theta : \alpha < ||\Theta||\,\},$$

with α ranging – in the formalized versions – over ordinals or linear orderings, depending on whether we want to treat, respectively, the stronger or the weaker (metapredicative) system.

The set I_Θ consists of triples of natural numbers which code names and their elements in the following sense:

$$
\begin{aligned}
(e,0,0) \in I_\Theta \quad &\sim \quad e \text{ is (name of) a type,} \\
(e,n,1) \in I_\Theta \quad &\sim \quad e \text{ is a type and } n \,\dot{\in}\, a, \\
(e,n,2) \in I_\Theta \quad &\sim \quad e \text{ is a type and } n \,\dot{\notin}\, a.
\end{aligned}
$$

The operator form $\mathfrak{A}(X,a,b,c)$ is reminiscent of the definition of Kleene's $\mathcal{O}$ and is used to deal with

- elementary comprehension and join,
- *elements* of universes (and of accessible parts).

The operator form $\mathfrak{B}(X,a,b,c)$, which contains negative occurrences in an essential way, provides for

- *names* of universes (and accessible parts),
- *extensions* of the complements of universes (and accessible parts).

Formalization of this approach in the suitable systems of nonmonotone inductive definitions immediately provides the sharp upper proof-theoretic bounds (cf. [38, 35]).

According to Theorem 9, our system $\mathsf{T}_0(\mathsf{M})$ has the same proof-theoretic strength as the system of constructive set theory $\mathsf{CZF} + (\mathsf{Mahlo})$ and thus is justified in the sense of reductive proof theory. However, the question remains whether there is a direct constructive justification of the system $\mathsf{T}_0^i(\mathsf{M})$ and, if so, what such a justification would mean.

It may be an interesting general approach to analyze whether nonmonotone inductive definitions of the form discussed above or even more powerful ones could be instrumentalized as tool for the constructive justification of strong systems. It could be the role of nonmonotone operators to provide for the possibility of carrying through *sufficiently long* iterations (of such operators) and the corresponding ordinals. More research in this direction is left for future publications.

REFERENCES

[1] PETER ACZEL, *The type-theoretic interpretation of constructive set theory*, ***Logic Colloquium '77*** (A. MacIntyre, L. Pacholski, and J. Paris, editors), North-Holland, 1978, pp. 55–66.

[2] ———, *The type-theoretic interpretation of constructive set theory: choice principles*, ***The L. E. J. Brouwer Centenary Symposium*** (A. S. Troelstra and D. van Dalen, editors), North-Holland, 1982, pp. 1–40.

[3] ———, *The type-theoretic interpretation of constructive set theory: inductive definitions*, ***Logic, Methodology and Philosophy of Science VII*** (R. B. Marcus, G. J. W. Dorn, and P. Weingartner, editors), North-Holland, 1986, pp. 17–49.

[4] PETER ACZEL and PETER G. HINMAN, *Recursion in the superjump*, ***Generalized Recursion Theory, Proceedings of the 1972 Symposium, Oslo*** (J. E. Fenstad and P. G. Hinman, editors), vol. 94, North-Holland, 1974, pp. 3–41.

[5] Jeremy Avigad, *On the relationship between ATR_0 and $\widehat{ID}_{<\omega}$*, ***The Journal of Symbolic Logic***, vol. 61 (1996), no. 3, pp. 768–779.

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

[7] Laura Crosilla, *Realizability Models for Constructive Set Theories with Restricted Induction*, ***Ph.D. thesis***, School of Mathematics, University of Leeds, 2000.

[8] Solomon Feferman, *Systems of predicative analysis*, ***The Journal of Symbolic Logic***, vol. 29 (1964), no. 1, pp. 1–30.

[9] ———, *A language and axioms for explicit mathematics*, ***Algebra and Logic*** (J.N. Crossley, editor), Lecture Notes in Mathematics, vol. 450, Springer, 1975, pp. 87–139.

[10] ———, *Recursion theory and set theory: a marriage of convenience*, ***Generalized Recursion Theory II, Oslo 1977*** (J. E. Fenstad, R. O. Gandy, and G. E. Sacks, editors), North-Holland, 1978, pp. 55–98.

[11] ———, *Constructive theories of functions and classes*, ***Logic Colloquium '78*** (M. Boffa, D. van Dalen, and K. McAloon, editors), North-Holland, 1979, pp. 159–224.

[12] ———, *Iterated inductive fixed-point theories: application to Hancock's conjecture*, ***The Patras Symposion*** (G. Metakides, editor), North-Holland, 1982, pp. 171–196.

[13] ———, *Weyl vindicated: "Das Kontinuum" 70 years later*, ***Temi e Prospettive della Logica e della Filosofia della Scienza Contemporanee*** (C. Celluci and G. Sambin, editors), CLUEB, 1988, pp. 59–93.

[14] ———, *Polymorphic typed lambda-calculi in a type-free axiomatic framework*, ***Logic and Computation*** (W. Sieg, editor), Contemporary Mathematics, vol. 106, American Mathematical Society, 1990, pp. 101–136.

[15] ———, *Logics for termination and correctness of functional programs*, ***Logic from Computer Science*** (Y. N. Moschovakis, editor), Springer, 1991, pp. 95–127.

[16] ———, *Logics for termination and correctness of functional programs II: Logics of strength PRA*, ***Proof Theory*** (P. Aczel, H. Simmons, and S. S. Wainer, editors), Cambridge University Press, 1992, pp. 195–225.

[17] ———, *Does reductive proof theory have a viable rationale?*, 2000, Preprint, Department of Mathematics, Stanford University.

[18] Solomon Feferman and Gerhard Jäger, *Systems of explicit mathematics with non-constructive μ-operator. Part I*, ***Annals of Pure and Applied Logic***, vol. 65 (1993), no. 3, pp. 243–263.

[19] ———, *Systems of explicit mathematics with non-constructive μ-operator. Part II*, ***Annals of Pure and Applied Logic***, vol. 79 (1996), no. 1, pp. 37–52.

[20] 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.

[21] Robin O. Gandy, *General recursive functionals of finite type and hierarchies of functions*, ***Ann. Fac. Sci. Univ. Clermont Ferrant***, vol. 35 (1967), pp. 5–24.

[22] Leo Harrington, *The superjump and the first recursively Mahlo ordinal*, ***Generalized Recursion Theory, Proceedings of the 1972 Symposium, Oslo*** (J. E. Fenstad and P. G. Hinman, editors), vol. 94, North-Holland, 1974, pp. 3–41.

[23] Gerhard Jäger, *First order theories for nonmonotone inductive definitions: recursively inaccessible and Mahlo*, ***The Journal of Symbolic Logic***, to appear.

[24] ———, *Die konstruktible Hierarchie als Hilfsmittel zur beweistheoretischen Untersuchung von Teilsystemen der Mengenlehre und Analysis*, ***Ph.D. thesis***, Universität München, 1979.

[25] ———, *Theories for iterated jumps*, 1980, Technical notes, Mathematical Institute, Oxford University.

[26] ———, *Iterating admissibility in proof theory*, ***Logic Colloquium '81. Proceedings of the Herbrand Symposion.***, North-Holland, 1982.

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

[28] ———, *The strength of admissibility without foundation*, ***The Journal of Symbolic Logic***, vol. 49 (1984), no. 3, pp. 867–879.

[29] ———, *Induction in the elementary theory of types and names*, ***Computer Science Logic '87*** (E. Börger, H. Kleine Büning, and M. M. Richter, editors), Lecture Notes in Computer Science, vol. 329, Springer, 1988, pp. 118–128.

[30] ———, *Type theory and explicit mathematics*, ***Logic Colloquium '87*** (H.-D. Ebbinghaus, J. Fernandez-Prida, M. Garrido, M. Lascar, and M. Rodriguez Artalejo, editors), North-Holland, 1989, pp. 117–135.

[31] ———, *The next admissible in set theories without foundation*, 1995, Technical notes. Institut für Informatik und angewandte Mathematik, Universität Bern.

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

[33] Gerhard Jäger, Reinhard Kahle, and Thomas Studer, *Universes in explicit mathematics*, ***Annals of Pure and Applied Logic***, to appear.

[34] Gerhard Jäger and Wolfram Pohlers, *Eine beweistheoretische Untersuchung von* $(\Delta^1_2\text{-}\mathsf{CA})+(\mathsf{BI})$ *und verwandter Systeme*, ***Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch-naturwissenschaftliche Klasse***, 1982, pp. 1–28.

[35] Gerhard Jäger and Thomas Strahm, *Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory*, ***The Journal of Symbolic Logic***, to appear.

[36] ———, *The proof-theoretic strength of the Suslin operator in applicative theories*, 1999, Preprint, Institut für Informatik und angewandte Mathematik, Universität Bern.

[37] ———, *Fixed point theories and dependent choice*, ***Archive for Mathematical Logic***, vol. 39 (2000), no. 7, pp. 493–508.

[38] Gerhard Jäger and Thomas Studer, *Extending the system* T_0 *of explicit mathematics: the limit and Mahlo axioms*, ***Annals of Pure and Applied Logic***, to appear.

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

[40] Paul Mahlo, *Über lineare transfinite Mengen*, ***Berichte über die Verhandlungen der königlich sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse***, vol. 63, 1911, pp. 187–225.

[41] Markus Marzetta, *Predicative Theories of Types and Names*, ***Ph.D. thesis***, Institut für Informatik und angewandte Mathematik, Universität Bern, 1993.

[42] John Myhill, *Constructive set theory*, ***The Journal of Symbolic Logic***, vol. 40 (1975), no. 3, pp. 347–382.

[43] Michael Rathjen, *Proof-theoretic analysis of* KPM, ***Archive for Mathematical Logic***, vol. 30 (1991), pp. 377–403.

[44] ———, *The recursively Mahlo property in second order arithmetic*, ***Mathematical Logic Quaterly***, vol. 42 (1996), pp. 59–66.

[45] ———, *The higher infinite in proof theory*, ***Logic Colloquium '95*** (J. Makowsky and E. Ravve, editors), Lecture Notes in Logic, vol. 11, Springer, 1998, pp. 275–304.

[46] ———, *The superjump in Martin-Löf type theory*, ***Logic Colloquium '98*** (S. Buss, P. Hájek, and P. Pudlák, editors), Lecture Notes in Logic, vol. 13, Association for Symbolic Logic, 2000, pp. 363–386.

[47] W. Richter, *Recursively Mahlo ordinals and inductive definitions*, ***Logic Colloquium '69*** (R. O. Gandy and C. E. M. Yates, editors), North-Holland, 1971, pp. 273–288.

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

[49] ———, *Metapredicative subsystems of analysis*, 2000, Technical notes. Institut für Informatik und angewandte Mathematik, Universität Bern.

[50] ———, *The proof-theoretic analysis of Σ^1_1 transfinite dependent choice*, 2000, Technical notes. Institut für Informatik und angewandte Mathematik, Universität Bern.

[51] Kurt Schütte, *Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik*, ***Archiv für Mathematische Logik und Grundlagen der Mathematik***, vol. 7 (1964), pp. 45–60.

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

[53] Stephen G. Simpson, *Set-theoretic aspects of* ATR_0, ***Logic Colloquium '80*** (D. van Dalen, D. Lascar, and J. Smiley, editors), North-Holland, Amsterdam, 1982, pp. 255–271.

[54] ———, *Σ^1_1 and Π^1_1 transfinite induction*, ***Logic Colloquium '80*** (D. van Dalen, D. Lascar, and J. Smiley, editors), North-Holland, Amsterdam, 1982, pp. 239–253.

[55] Thomas Strahm, *Wellordering proofs for metapredicative Mahlo*, ***The Journal of Symbolic Logic***, submitted.

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

[57] ———, *Autonomous fixed point progressions and fixed point transfinite recursion*, ***Logic Colloquium '98*** (S. Buss, P. Hájek, and P. Pudlák, editors), Lecture Notes in Logic, vol. 13, Association for Symbolic Logic, 2000, pp. 449–464.

[58] Thomas Studer, *Constructive Foundations for Featherweight Java*, submitted.

[59] ———, *A semantics for $\lambda^{\{\}}_{str}$: a calculus with overloading and late-binding*, ***Journal of Logic and Computation***, to appear.

[60] ———, *Explicit mathematics: W-type, models*, ***Master's thesis***, Institut für Informatik und angewandte Mathematik, 1997.

[61] Anne S. Troelstra and D. van Dalen, ***Constructivism in Mathematics***, vol. I and II, North-Holland, 1988.

[62] Sergei Tupailo, *Realization of analysis into explicit mathematics*, ***The Journal of Symbolic Logic***, to appear.

[63] ———, *Realization of constructive mathematics into explicit mathematics: a lower bound for an impredicative Mahlo universe*, 2000, Preprint, Institut für Informatik und angewand Mathematik, Universität Bern.

INSTITUT FÜR INFORMATIK UND ANGEWANDTE MATHEMATIK
UNIVERSITÄT BERN
NEUBRÜCKSTRASSE 10
CH-3012 BERN, SWITZERLAND
E-mail: jaeger@iam.unibe.ch