Proof Theory and Martin-Löf Type Theory

Anton Setzer

January 23, 2008

Abstract

We give an overview of the historic origins of ordinal theoretic proof theory. We indicate why ordinal theoretic proof theory needs to be supplemented by a second step, namely the proof of the well-foundedness of the ordinal notation systems used in theories, for which we have an argument, that everything derived in them is valid. Such theories serve as a substitute for Hilbert's finitary methods, and we explain, in which sense extensions of Martin-Löf type theory (MLTT) are theories, which can be used for this purpose. Next we go through some ordinals along the proof theoretic scale, and describe variants of MLTT used in order to prove the well-foundedness of the ordinal notation systems and therefore the consistency of theories of that strength. The variants of MLTT discussed are basic MLTT with natural numbers, the extension by the W-type, the extension by one universe, MLTT with general inductive-recursive definitions (which includes the so called "red" Mahlo universe), the so called "black" Mahlo universe, the hyper-Mahlo universe, the hyper-α-Mahlo universe, the autonomous Mahlo universe, and the Π_3-reflecting universe. We indicate how with increasing strength the insight into the correctness of the variants of MLTT becomes more difficult, which is by Gödel's incompleteness theorem unavoidable.

-

1 Introduction

It is the goal of ordinal theoretic proof theory to reduce the consistency of theories for formalising mathematical proofs to the well-foundedness of ordinal notation systems. In order to obtain a satisfactory solution to the consistency problem, this reduction needs to be supplemented by a second step, namely by proofs of the well-foundedness of the ordinal notation systems in "safe" theories, for which one has an argument that everything shown in these theories is valid. Because of Gödel's incompleteness theorem, a mathematical correctness proof only can prove relative correctness, but never provide an absolute proof of correctness. In order to obtain an argument which provides absolute trust, the only possibility is to have a philosophical correctness argument. Although there are other theories, which could serve as such safe theories, the theories, for which such arguments have best been worked out at present are extensions of Martin-Löf type theory (*MLTT*).

In this article we will give an overview of the techniques used in this program of reducing the consistency of mathematical theories to MLTT. We will start by giving an introduction into the origins and some basic techniques of ordinal theoretic proof theory. This will motivate the notion of the proof theoretic strength. We will then indicate how to use MLTT in order to obtain a complete consistency proof, which is necessarily based on a philosophical argument. Then we will go through some key steps on the proof theoretic scale: we consider the following variants of MLTT: MLTT with natural numbers; the additional extension by the W-type; the additional extension by one universe; MLTT with inductive-recursive definitions; the red and black Mahlo universes. We will finally look briefly at the hyper-Mahlo, hyper-α-Mahlo, autonomous Mahlo, and Π_3-reflecting universes.

Notations. When introducing new notions, we write them in italic, if they occur as text, and underline them, if they occur as mathematical formulae.

We write terms in functional style, e.g. in the form $\underline{\mathrm{C}\ a_0 \cdots a_n}$ rather than in mathematical style $\mathrm{C}(a_0, \ldots, a_n)$. In running text we put, if necessary, brackets around it, i.e. we write $(\mathrm{C}\ a_0 \cdots a_n)$ instead of $\mathrm{C}\ a_0 \cdots a_n$.

When developing the semantics of variants of MLTT, we will write $\underline{[\![A]\!]}$ for the interpretation of set A. Although the standard model of type theory is the *PER* (*partial equivalence relation*) model, where $[\![A]\!]$ is a set of pairs of terms, namely those, which are equal elements of A, we usually do for simplicity as if $[\![A]\!]$ were just a set of terms and identify therefore $[\![A]\!]$ with its domain.

We write $\underline{a[x := b]}$ for the result of substituting in a the variable x by b. We write $\underline{a[x]}$ for an expression a possibly depending on a variable x. Once we have used $a[x]$, $\underline{a[b]}$ denotes $a[x := b]$, where we assume that any clashes between bound and free variables are resolved by applying α-conversion (all α-equivalent expressions are identified).

2 The Rôle of Type Theory in a Proof Theoretic Program

Proof theory is a discipline of mathematical logic, which was founded by David Hilbert ([40],[41]) at the beginning of the 20th century. At that time various axiom systems for carrying out mathematical proofs had been developed, of which some had turned out to be inconsistent. In order to guarantee that the axiom systems, which were actually used, don't contain any inconsistencies, Hilbert proposed to prove the consistency of mathematical axiom systems [39]. He observed that, if one shows the consistency of a theory for formalising mathematics in the same or an even stronger theory, one has not achieved anything: if the original theory is inconsistent, it proves everything, even its own consistency. So in order to achieve something, one has to do more: namely show the consistency using methods, which are considered to be safe. According to Hilbert, finitary methods were safe [40]. By finitary methods he considered finitary calculations which can be carried out on a piece of paper.

There are two main approaches for carrying out consistency proofs. One is to introduce a model of the system in question in the Meta-theory. However, it seems to be implausible to assume that one can prove this way the consistency of a theory by using finitary methods, since such methods do not allow the use of sets. Hilbert realized this and suggested therefore that one should instead analyse proofs and show this way directly that it is not possible to derive in the formal system in question a contradiction. He called the mathematical discipline, in which such investigations are carried out, proof theory ([40],[41]).

Gödel's second incompleteness theorem and the failure of Hilbert's original programme. In 1931 Gödel [35] showed in his second incompleteness theorem that Hilbert's original programme cannot be carried out – assuming minimal conditions on a theory T, he could show that a consistent theory T does not prove its own consistency. Since finitary methods should be formalisable in any reasonable theory T, it follows that the consistency of theories fulfilling those minimal conditions cannot be shown by finitary means. Most natural theories except for extremely weak ones fulfil the premise of the last sentence – Hilbert's original programme had failed.

Gentzen's proof of the consistency of Peano Arithmetic. 1936 Gerhard Gentzen ([32],[33]) showed the consistency of *Peano Arithmetic* (PA) in PRA (primitive recursive arithmetic, as defined later) extended by quantifier free transfinite induction up to ϵ_0. This was the birth of ordinal theoretic proof theory. His methods have been improved since. The shortest and most elegant presentation is due to Buchholz (e.g in [22]), which is as follows:

One develops PA in a one-sided calculus, the *Tait calculus* [97], in which we derive finite sets of closed formulae Γ, where the intended meaning of Γ is essentially the disjunction of the formulae in Γ. A statement $A_1, \ldots, A_n \vdash B_1, \ldots, B_k$ is translated into the one sided sequent form as $\neg A_1, \ldots, \neg A_n, B_1, \ldots, B_k$ (this elegant approach works only for classical logic).

We develop an infinitary theory $\underline{\mathrm{PA}^*}$, in which we will interpret proofs from PA. PA^* uses the Tait calculus, and formulae in PA^* are closed formulae from PA. Each formula corresponds to a finitary or infinitary disjunction or conjunction, and we write $\underline{A \simeq \bigwedge_{i\in I} A_i}$ for formula A being associated with the conjunction of A_i for $i \in I$, similarly $\underline{A \simeq \bigvee_{i\in I} A_i}$ for the association of a disjunction. For a true prime formula A we have $A \simeq \bigwedge_{i\in\emptyset} A_i$ (there is no A_i), for a false prime formula we have $A \simeq \bigvee_{i\in\emptyset} A_i$, we have $A_0 \wedge A_1 \simeq \bigwedge_{i\in\{0,1\}} A_i$, $A_0 \vee A_1 \simeq \bigvee_{i\in\{0,1\}} A_i$, $\forall x.A(x) \simeq \bigwedge_{i\in\mathbb{N}} A(i)$, $\exists x.A(x) \simeq \bigvee_{i\in\mathbb{N}} A(i)$. By the deMorgan rules we can consider negation as a defined operation (assuming we have for each atomic formula its negation as an atomic formula as well). Note that the formulae used are finitary expressions, which are associated with possibly infinitary conjunctions and disjunctions. When writing $\underline{\bigwedge_{i\in I} A_i}$ or $\underline{\bigvee_{i\in I} A_i}$, we mean in the following any formula A s.t. $A \simeq \bigwedge_{i\in I} A_i$ or $A \simeq \bigvee_{i\in I} A_i$, respectively.

PA^* has 3 rules:

$$\frac{\Gamma, A_i \quad (\text{all } i \in I)}{\Gamma, \bigwedge_{i\in I} A_i}\,(\wedge-\text{intro}) \qquad\qquad \frac{\Gamma, A_i \quad (\text{some } i \in I)}{\Gamma, \bigvee_{i\in I} A_i}\,(\vee-\text{intro})$$

$$\frac{\Gamma, A \qquad \Gamma, \neg A}{\Gamma}\,(\text{Cut})$$

The formula A in the rule (Cut) is called the *cut formula* of this rule. Derivations in PA^* are well-founded derivations constructed from those rules, where *well-foundedness* means that there is no infinite descending sequence in the proof tree.

One can see that PA^* allows to interpret the logical rules of classical logic. The assumption rule (in standard sequent calculus written as $A \vdash A$) reads $\neg A, A$ and is provable in this calculus by induction on the rank of A (where the *rank* is the number of logical connectives in A – note that the formulae are closed formulae from PA, which are finite expressions). For instance, the step from $\neg A(n), A(n)$ for all $n \in \mathbb{N}$ to $\neg(\forall x.A(x)), \forall x.A(x)$, which is $\exists x.\neg A(x), \forall x.A(x)$, reads

$$\frac{\cdots \qquad \dfrac{\neg A(n), A(n)}{\exists x.\neg A(x), A(n)} \qquad \cdots \quad (n \in \mathbb{N})}{\exists x.\neg A(x), \forall x.A(x)}$$

By $\vee$-introduction we obtain immediately $A \vee \neg A$ as well.

Modus Ponus, which derives from $A \to B$, which is $\neg A \vee B$, and A the formula B, is interpreted as follows (we use here the fact that PA^* is closed under *weakening*, i.e. from a proof of Γ, we can obtain a proof of Γ, Δ; this is used here to weaken the proof of A to a proof of A, B; note that $\neg(\neg A \vee B) = A \wedge \neg B$):

$$\frac{\neg A \vee B \qquad \dfrac{A \qquad \neg B, B}{A \wedge \neg B, B}}{B}\,(\text{Cut})$$

Similarly, arithmetical and equality axioms can be interpreted. True atomic formulae correspond to the empty conjunction, which is provable by the rule ($\wedge -$ intro) with no premises in this instance; this allows then by the rules for $\wedge$ and $\vee$ to prove all true propositional closed formulae. By the infinitary introduction rule for $\forall$ we can prove all true closed formulae of the form $\forall x.A(x)$, where $A(x)$ is quantifier free. This allows as well to prove the equality axioms (where congruence $s = t \to A(s) \to A(t)$ is restricted to quantifier free formulae, which is sufficient to prove congruence for all formulae). Now we can interpret proofs in PA as infinitary proofs in PA*:

The only principle we haven't interpreted in PA* yet is the induction rule. So assume PA $\vdash A(0)$ and PA $\vdash \forall x.A(x) \to A(x+1)$, from which the induction rule derives PA $\vdash \forall x.A(x)$. By induction hypothesis we have PA$^* \vdash A(0)$ and PA$^* \vdash \forall x.A(x) \to A(x+1)$. Then we obtain in PA* proofs of $A(n)$ for all $n \in \mathbb{N}$. One can easily show $(\forall x.A(x) \to A(x+1)) \to (A(n) \to A(n+1))$ for all n, therefore we obtain $A(n) \to A(n+1)$. Now using iterated Modus Ponus, we can, using PA$^* \vdash A(0)$, prove PA$^* \vdash A(n)$ for all $n \in \mathbb{N}$. By $\wedge$-introduction we obtain therefore PA$^* \vdash \forall x.A(x)$. Note that this proof has infinite height, since, in order to prove the premise $A(n)$, at least n applications of Modus Ponus were used.

The *cut rank* of a proof in PA* is the supremum of the rank of cut formulae occurring in it. One can see that the cut rank of a proof in PA*, which is the translation of a proof in PA, is finite.

If we omit the cut rule, PA* is consistent, since falsity is interpreted as the empty disjunction, and there is no rule for deriving the empty disjunction. Therefore, in order to prove the consistency of PA, it suffices to show that from each proof of falsity with finite cut rank we obtain a cut free proof of falsity. In fact, the cut elimination result holds for arbitrary proofs of finite cut rank (the theorem can be extended to infinite cut rank as well) in PA*: all proofs can be transformed into cut free proofs of the same set of formulae.

The height of well-founded trees and ordinal notation systems. The proof of cut elimination is carried out by induction over the (well-founded) derivations. Since we have used infinitary proofs, the height of a proof can no longer be measured by a natural number. Proofs are well-founded, and the *height of well-founded trees* can be measured by ordinals. In set theory, *ordinals* form a class <u>Ord</u>, which is *well-ordered* (which means linearly ordered and well-founded), so there are no infinitely descending sequences. For any other well-ordered set $(A, \prec)$ we can find an ordinal α such that $(\{\beta \in \mathrm{Ord} \mid \beta < \alpha\}, <)$ (which is equal to α) is isomorphic to $(A, \prec)$. In this situation α is called the *order type* of $(A, \prec)$. Using this fact one can derive that the height of any well-founded tree, which is a set, can be measured by a unique ordinal α.

Ordinals from set theory are needed only for heuristic reasons – in ordinal theoretic proof theory one works with ordinal notation systems instead. An *ordinal notation system* is a primitive recursive subset OT of $\mathbb{N}$ together with a primitive-recursive binary relation $\prec$ s.t. $(\mathrm{OT}, \prec)$ is well-ordered. Note that elements of OT are therefore just finitary objects. Any $a \in \mathrm{OT}$ denotes the ordinal α, where α is the order type of $\{b \in A \mid b \prec a\}$ – this relates ordinal notations to set theoretic ordinals.

Gentzen was able to replace the induction over proof trees, which is needed in the proof of the cut elimination theorem for PA*, by transfinite induction over an ordinal notation system, the order type of which is denoted by ϵ_0. All infinitary derivations can be replaced by finitary objects (Gentzen worked directly in a finitary system), the most elegant approach for achieving this is due to Buchholz ([22]).

All other arguments can be carried out in *primitive recursive arithmetic* <u>*PRA*</u> (which is PA extended by symbols for primitive recursive functions and their defining axioms, and induction restricted to quantifier free formulae), which is usually

considered as a formalisation of Hilbert's finitary methods. Furthermore, transfinite induction over ϵ_0 can be restricted to quantifier free formulae. Gentzen proved the consistency of PA by using PRA extended by quantifier free transfinite induction up to ϵ_0. It is easy to convince oneself directly that transfinite induction up to ϵ_0 is a valid principle, and in this sense Gentzen obtained a consistency proof of PA.
Proof theoretic strength. Since PRA can be interpreted in PA, PA extended by transfinite induction up to ϵ_0 proves the consistency of PA. Therefore, by Gödel's incompleteness theorem, PA does not prove transfinite induction up to ϵ_0 (unless it is inconsistent). One can show that for any $\alpha < \epsilon_0$ PA proves transfinite induction up to α. We define the *proof theoretic strength* $\underline{|\mathrm{T}|}$ of a theory T as the supremum of all α s.t. T proves transfinite induction up to α (more precisely one needs to restrict oneself to so called natural ordinal notation systems). Then we obtain $|\mathrm{PA}| = \epsilon_0$.
The need for a constructive foundation of proof theory. Gentzen's approach has later been extended to systems of increasing strength and the most powerful result was the analysis of second order arithmetic restricted to the comprehension axiom for Π^1_2-formulae by M. Rathjen ([65],[66],[73],[74]) and, independently, T. Arai ([3],[4],[5],[6],[7],[8]). In fact, Rathjen has carried out an analysis up to the strength of $(\Pi^1_2 - \mathrm{CA}) + (\mathrm{BI})$, but at present the fully published versions reach $(\Delta^1_2 - \mathrm{CA}) + (\mathrm{BI}) + (\Pi^1_2 - \mathrm{CA})^-$. Here $\underline{(\Pi^1_2 - \mathrm{CA})}$ is the comprehension axiom for Π^1_2-formulae, other comprehension axioms are denoted similarly, $\underline{(\Pi^1_2 - \mathrm{CA})^-}$ is the parameter free Π^1_2-comprehension axiom, and $\underline{\mathrm{BI}}$ stands for Bar induction. T. Arai has carried out an analysis up to the strength of $(\Sigma^1_3 - \mathrm{DC}) + (\mathrm{BI})$ ($\underline{\mathrm{DC}}$ stands for the axiom of dependent choice), but most of his work beyond Kripke-Platek set theory plus Π_3-reflection exists at present only in the form of (carefully worked out) preprints.

With increasing proof theoretic strength the ordinal notation systems used become increasingly complicated, and it is no longer possible to get a direct insight into their well-foundedness. (There is an approach [85, 87] by the author called ordinal systems, to obtain intuitive arguments for as strong as possible ordinal notation systems directly.) What can be done instead is to prove the well-foundedness in another theory, a theory $\mathrm{T}_{\mathrm{Good}}$, for which we have a more direct insight into that all its theorems are valid. If PRA can be interpreted in those theories, which is expected, we obtain that $\mathrm{T}_{\mathrm{Good}}$ proves that T is consistent. This way we obtain therefore a consistency proof of T, assuming we believe that everything proved in $\mathrm{T}_{\mathrm{Good}}$ is valid.

The most successful (but not only) theories providing such a direct insight into their validity are constructive theories, and at present the best developed theories for this purpose are extensions of *Martin-Löf Type Theory* (*MLTT*). The argument that everything proved in MLTT is valid is given by *meaning explanations*. Meaning explanations rely substantially on the philosophy of language, which goes beyond the author's expertise, and therefore no meaning explanations are given in this article. We note that at present no meaning explanations have been given for theories of strength the black Mahlo universe and beyond.

One should note that there is another main approach to providing theories, which could serve as $\mathrm{T}_{\mathrm{Good}}$, namely Feferman's theories of explicit mathematics [31]. However, their philosophical foundations are not as well developed, and most research on these theories has been focused on classical and therefore non-constructive variants.

Therefore, MLTT can be used as a substitute for Hilbert's finitary methods. In order to prove the consistency of strong theories for carrying out mathematical proofs we need therefore proof theoretically strong extensions of MLTT.

For this program we only need to prove lower bounds for the proof theoretic strength of MLTT. However, in order to determine the limits of a certain extension, it is important to know as well an upper bound. Therefore, a constructive underpinning of proof theory is achieved by developing proof-theoretically strong extensions of MLTT, for which we have an insight into their correctness, and by determining their precise proof theoretic strength.

Note that this is as well of relevance for a full constructivist, who is only interested in proofs in constructive theories. A constructivist should be interested in proving as many theorems as possible, and therefore has an interest in obtaining an as strong as possible constructive theory. It is as well important to compare this strength with non-constructive theories, in order to determine the limits of what can be proved in the constructive theory.

3 From Natural Numbers to Universes

The basic framework of MLTT. For historic reasons, the lowest type level in MLTT is called "Set". Outside MLTT this type level will usually be denoted by "Type", whereas in MLTT "Type" denotes a type level on top of Set. We will not make use of "Type" until we reach the theory of inductive-recursive definitions. The use of Type outside MLTT explains, why we will, as usual in type theory, refer later to Π-types, Σ-types etc., even so they are elements of Set and should therefore be called Π-sets, Σ-sets, etc.

In MLTT one has *non-dependent judgements* $A : \mathrm{Set}$ (A is a set), $A = B : \mathrm{Set}$ (A and B are equal sets), $a : A$ (a is an element of set A), and $a = b : A$ (a and b are equal elements of set A). Furthermore, we have *dependent judgements* $x_1 : A_1, \ldots, x_n : A_n \Rightarrow \theta$, where θ has the form of a non-dependent judgement, but might depend on $x_1, \ldots, x_n$. Here A_i might depend on $x_1, \ldots, x_{i-1}$. We usually write capital Greek letters Γ, Δ for contexts of the form $x_1 : A_1, \ldots, x_n : A_n$.

When introducing sets by declaring what their elements are, we will in the following mean in fact canonical elements of these sets. *Canonical elements* are those, which are in head normal form, i.e. they start with a constructor. Arbitrary elements of such a set are those, which reduce to canonical elements.

The basic theory of MLTT. *Basic MLTT* (note that Martin-Löf usually adds $\mathbb{N}$, the W-type, and universes to his type theory) consists of the rules for the following sets, which we call the *basic set constructions*:

- the *finite sets* N_n having n elements $\mathrm{A}^n_0, \ldots, \mathrm{A}^n_{n-1}$;
- the *Σ-type* $\Sigma x : A.B[x]$, the elements of which are pairs $(\mathrm{p}\ a\ b)$, where $a : A$ and $b : B[a]$;
- the *Π-type* $\Pi x : A.B[x]$ (dependent function type) with elements $\lambda y.t[y]$ s.t. $y : A \Rightarrow t[y] : B[y]$; we define $A \to B := \Pi x : A.B$ for some fresh variable x, and see that the non-dependent function type is an instance of the Π-type;
- the *disjoint union* $A + B$ of sets $A + B$ with elements $(\mathrm{inl}\ a)$ for $a : A$ and $(\mathrm{inr}\ b)$ for $b : B$;
- the *identity type* $(a =_A a')$ for $A : \mathrm{Set}$, $a : A$, $a' : A$, with element $\mathrm{refl}\ a : a =_A a$ for $a : A$; $(a =_A a')$ stands for the proposition that a and a' are equal elements of A.

MLTT with natural numbers. The weakest type theory we consider is MLTT extended by the rules for the set of natural numbers $\mathbb{N}$, and a microscopic universe Atom. Atom is needed, since without it one can show that one cannot prove Peano's fourth axiom, namely $\neg(0 =_{\mathbb{N}} 1)$. In order to define Atom, we first give better names to N_2 and its elements by defining $\mathrm{Bool} := \mathrm{N}_2$, and calling its two elements tt and ff. Then the rules for Atom express that, depending on $b : \mathrm{Bool}$, we have $\mathrm{Atom}\ b : \mathrm{Set}$, with equality rules Atom ff $=$ N_0, Atom tt $=$ N_1. Note that Atom converts a Boolean value into the corresponding atomic formula (therefore the name Atom): (Atom b) is inhabited (i.e. provable) if and only if b is true. The introduction rules for $\mathbb{N}$ express that it contains $\underline{0}$, and that, whenever $n : \mathbb{N}$, then $\underline{\mathrm{S}\, n} : \mathbb{N}$. The elimination rule for $\mathbb{N}$ corresponds to the principle of induction (or primitive recursion into an arbitrary dependent type). Let the resulting type theory be called $\underline{\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N}}$.

It is easy to show that all formulae provable in *Heyting Arithmetic* $\underline{\mathrm{HA}}$ (which is PA with classical logic replaced by intuitionistic logic; we have $|\mathrm{PA}| = |\mathrm{HA}|$) are provable in MLTT + Atom + $\mathbb{N}$ as well, and from this fact one can derive $|\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N}| \geq |\mathrm{HA}|$. Furthermore, each judgement $\Gamma \Rightarrow \theta$ provable in MLTT + Atom + $\mathbb{N}$ can be interpreted as a closed provable formula in HA, from which one can derive that $|\mathrm{MLTT}+\mathrm{Atom}+\mathbb{N}| \leq |\mathrm{HA}|$. Therefore we have $|\mathrm{MLTT}+ \mathrm{Atom} + \mathbb{N}| = |\mathrm{HA}| = |\mathrm{PA}| = \epsilon_0$. Note that we have therefore reduced PA to an extension of MLTT and therefore given a constructive justification of PA.

If one looks at this type theory from a consistency point of view, one observes that one has to understand first the basic setup of type theory. Once one has done this, the only problematic construction which remains to be understood is the set of natural numbers $\mathbb{N}$. In order to understand this set, one needs to understand the meaning of a least set closed under finitary introduction rules. This is rather unproblematic, since each element of $\mathbb{N}$ can be introduced by only finitely many applications of the introduction rules. Note that there is some analogy with time: if we start with 0 and in regular time intervals apply the successor function to this element, we will reach each natural number after a finite amount of time.

The W-type. The next step to increase the strength of MLTT is the addition of the *W-type*. Assume $A : \mathrm{Set}$, $x : A \Rightarrow B[x] : \mathrm{Set}$. Then the formation rule for the W-type expresses that $\underline{\mathrm{W}x : A.B[x] : \mathrm{Set}}$. The elements of $\mathrm{W}x : A.B[x]$ are labelled well-founded trees, with labels in A, where nodes with label $a : A$ have branching degree $B[a]$. So whenever we have $a : A$ and $f : B[a] \rightarrow \mathrm{W}x : A.B[x]$, we can introduce a new element $\underline{(\mathrm{sup}\ \ a\ f)} : \mathrm{W}x : A.B[x]$. Here (sup a f) is a tree with root (sup a f), which is labelled with a, has therefore branching degree $B[a]$, and subtrees (f b) for $b : B[a]$. Let $\underline{\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}}$ be the extension of the previous theory by the rules for the W-type.

We introduce a standard model of this type theory. The most difficult construction is to interpret the W-type. We define $\underline{[\![\, \mathrm{W}x : A.B \,]\!]} := [\![\, \mathrm{W} \,]\!]([\![\, A \,]\!], \lambda a \in [\![\, A \,]\!].[\![\, B \,]\!]_{x \mapsto a})$. Here $\underline{[\![\, \mathrm{W} \,]\!]}$ takes a set A of terms and a function F, mapping A to a set of terms, and returns the least set X of terms closed under reductions, such that, if $a \in A$ and $f \in F(a)\ [\![\rightarrow]\!]\ X$, then (sup a f) $\in X$. Here $\underline{f \in F(a)\ [\![\rightarrow]\!]\ X}$ means that for $x \in F(a)$, $f\ x \in X$. *Closure under reduction* means that if b is a term which reduces to $c \in X$, then $b \in X$. More precisely the above has to reformulated in terms of PERs rather than sets of terms. If we unfold the operators $[\![\, \mathrm{W} \,]\!]$ and $[\![\rightarrow]\!]$ in order to obtain a direct definition of $[\![\, \mathrm{W}x : A.B \,]\!]$, we see that $[\![\, \mathrm{W}x : A.B \,]\!]$ can be interpreted as a set defined by a strictly positive inductive definition.

Using this idea one can interpret all statements provable in MLTT+Atom+$\mathbb{N}$+W in the theory of finitely iterated inductive definitions $\mathrm{ID}_{<\omega}$, and therefore obtain $|\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}| \leq |\mathrm{ID}_{<\omega}|$. Here $\underline{\mathrm{ID}_{<\omega}}$ extends PA by adding predicate symbols indexed by strictly positive formulae and axioms stating that the sets given by those predicates are the least fixed points given by the operators corresponding to those formulae. (See [23] for a monograph on the proof theory of iterated inductive definitions.)

It is known [23] that the classical theory $\mathrm{ID}_{<\omega}$ has the same strength as the theory $\underline{\mathrm{ID}^{\mathrm{int},\mathcal{O}}_{<\omega}}$, which is the restriction of $\mathrm{ID}_{<\omega}$ to intuitionistic logic, and where we take as inductive definitions only the predicates corresponding to the nth constructive number classes. An easy way to introduce those number classes $\underline{\mathrm{O}_i}$ is to refer to their representation in type theory. There we define $\mathrm{O}_0 := \mathrm{N}_0$ (the empty set), $\mathrm{O}_1 := \mathrm{N}_1$ (the singleton set), $\mathrm{O}_2 := \mathbb{N}$, and for $n \geq 3$ $\mathrm{O}_n := \mathrm{W}x : \mathrm{N}_n.B[x]$, where $B[\mathrm{A}_i^n] = \mathrm{O}_i$. (More precisely we define $B[x]$ using the identity type in such a way that $B[\mathrm{A}_i^n]$ is isomorphic to O_i.) Then the nth constructive number class can be represented as O_{n+1}. There is a straightforward encoding of O_i as a set $\underline{\mathrm{O}'_i}$ of natural numbers, and O'_3 turns out to be Kleene's O. O'_{n+2} can be modelled by an n-times nested inductive definition. $\underline{\mathrm{ID}^{\mathrm{int},\mathcal{O}}_{<\omega}}$ is the theory having as inductive definitions exactly those defining O'_{n+2} for $n \in \mathbb{N}$. It is easy to model $\mathrm{ID}^{\mathrm{int},\mathcal{O}}_{<\omega}$ in $\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}$, and we therefore obtain $|\mathrm{ID}_{<\omega}| = |\mathrm{ID}^{\mathrm{int},\mathcal{O}}_{<\omega}| \leq |\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}| \leq |\mathrm{ID}_{<\omega}|$. It is known as well that $|(\Pi^1_1 - \mathrm{CA})_0| = |\mathrm{ID}_{<\omega}|$. Here $\underline{(\Pi^1_1 - \mathrm{CA})_0}$ is the second order theory of Π^1_1-comprehension with induction formulated as an axiom, i.e. induction can only applied to sets definable by comprehension. Since all proof theoretic equivalences can be shown in HA (although this is not done explicitly, it is well known in the proof theoretic community that in principle it can be done) and therefore as well in $\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}$, it follows that $\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}$ proves the consistency of approximations of $\mathrm{ID}_{<\omega}$, $\mathrm{ID}^{\mathrm{int},\mathcal{O}}_{<\omega}$, $(\Pi^1_1 - \mathrm{CA})_0$, and therefore we obtain a constructive consistency proof for those theories. By the results of reverse Mathematics (see for instance the monographs [94, 95]) it is known that $(\Pi^1_1 - \mathrm{CA})_0$ allows to prove almost all theorems of ordinary mathematics, so already at this stage most of mathematics can be secured constructively.

Relationship to Kripke-Platek set theory. For the understanding of the proof theory of further extensions of type theory, it is important to understand the relationship to extensions of Kripke-Platek set theory ([55, 17, 42]) and admissible ordinals. One approach to introducing admissible ordinals (we follow here [17], p. 3) is to define an idealised computer, which performs computations involving less than α steps. The functions F from ordinals to ordinals computed by such a computer are called *α-recursive*. An ordinal α is *admissible*, if for every α-recursive function F we have that α is closed under F, i.e. $F \upharpoonright \alpha : \alpha \to \alpha$. Barwise [18] has shown that for any admissible ordinal there exist a structure $\mathcal{M}$ with a definable pairing function such that α is the limit of all closure ordinals of first-order positive inductive definitions in $\mathcal{M}$, and that for any such $\mathcal{M}$ this limit is admissible.

Kripke Platek set theory (KP) is a weak axiomatisation of set theory. It is designed in such a way that in the constructible hierarchy we have L_α is a model of KP if and only if α is admissible. ($\underline{\mathrm{L}_\alpha}$ is the αth constructible set as introduced by Gödel [36] – $\underline{\mathrm{L}} := \bigcup_{\alpha \in \mathrm{Ord}} \mathrm{L}_\alpha$ forms a model of ZFC + GHC). One can easily show that the height of O_{2+i} is the ith admissible ordinal $\underline{\tau_i}$, sometimes denoted by $\underline{\omega_i^{ck}}$. In general we have that, if $[\![A]\!] \in \mathrm{L}_\alpha$, and if for $a \in [\![A]\!]$ we have $[\![B]\!]_{x \mapsto a} \in \mathrm{L}_\alpha$, then $[\![\mathrm{W}x : A.B]\!] \in \mathrm{L}_{\alpha^+}$. Using this idea one can develop a model of any approximation of $\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}$ in the theory KPl. Here KPl is a variant of KP, which essentially formulates the existence of finitely many admissible ordinals, so if $\underline{\tau_{<\omega}} := \sup_{n<\omega} \tau_n$, then $\mathrm{L}_{\tau_{<\omega}}$ is the standard model of KPl. Then one can see that $|\mathrm{KPl}| = |\mathrm{ID}_{<\omega}|$, and we obtain therefore a constructive consistency proof of KPl.

If one looks at $\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}$ from a consistency point of view, we observe that one needs to understand in addition to the constructions studied before the W-type. For this we need to understand $\mathrm{W}x : A.B$ as the least set closed under the introduction rule, which introduces elements of the form $(\sup\ a\ f)$. This is much more problematic than understanding $\mathbb{N}$ as discussed before, since the elements can no longer be introduced in finitely many steps. Note as well that the analogy with time, as we had it for the natural numbers, is broken – one needs to understand the analogy of some kind of transfinite time, which goes beyond our direct experience.

Universes. The next set construction added to type theory, which substantially increases its strength, is a universe. A *universe* is a family of sets. This is given by a set $\underline{\mathrm{U}}$: Set of codes for sets, together with a decoding function $\underline{\mathrm{T}}$, s.t. $u : \mathrm{U} \Rightarrow \mathrm{T}\ u : \mathrm{Set}$. Then $(\mathrm{T}\ u)$ is the set denoted by the code u.

The introduction rules express that U is closed under the previous set constructions N_n, Σ, Π, $+$, $=$, $\mathbb{N}$, W, which we call the *standard set constructions*. This is expressed in case of closure under $\mathbb{N}$ as follows: we have a code $\underline{\widehat{\mathbb{N}}} : \mathrm{U}$ with equality rule $\mathrm{T}\ \widehat{\mathbb{N}} = \mathbb{N} : \mathrm{Set}$. In case of closure under the W-type we have: if $a : \mathrm{U}$, $b : \mathrm{T}\ a \to \mathrm{U}$, then $\underline{\widehat{\mathrm{W}}\ a\ b} : \mathrm{U}$, and $\mathrm{T}\ (\widehat{\mathrm{W}}\ a\ b) = \mathrm{W}x : \mathrm{T}\ a.\mathrm{T}\ (b\ x) : \mathrm{Set}$.

In the original formulation by Martin-Löf there are no elimination rules for universes. One can add such rules, but they do not add any strength to the theory. In fact, it turns out that these rules are not very useful. If one, for instance, added another universe $\mathrm{U}' : \mathrm{Set}$ together with $u : \mathrm{U}' \Rightarrow \mathrm{T}'\ u : \mathrm{Set}$ to the type theory, such that U' has the same closure properties as U (with constructors $\widehat{\mathbb{N}}'$, $\widehat{\mathrm{W}}'$ etc), then the elimination rules don't seem to allow to define a function $f : \mathrm{U} \to \mathrm{U}'$ s.t. $f\ \widehat{\mathbb{N}} = \widehat{\mathbb{N}}'$, $f\ (\widehat{\mathrm{W}}\ a\ b) = \widehat{\mathrm{W}}\ (f\ a)\ (\lambda x.f\ (b\ x))$, unless one extends type theory substantially (it is believed that this is not possible, but this fact hasn't been shown yet). Let $\underline{\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{U}}$ be the extension of $\mathrm{MLTT} + \mathrm{Atom} + \mathbb{N} + \mathrm{W}$ by the rules for a universe closed under the W-type, and by omitting the rules for Atom (Atom is definable using the universe).

The universe as introduced above is the one originally introduced by Martin-Löf ([47]), which we call the *standard universe*. (Martin-Löf considered in his book [47] finitely iterated universes as well). One can consider universes with different closure properties. In its most general form we obtain the concept of an inductive-recursive definition, as discussed in the next section.

Because of their closure under the W-type, standard universes form some kind of inductive definitions, which is closed under the step to the next inductive definition. More precisely, universes correspond to *recursively inaccessible ordinals*. Recursively inaccessible ordinals are admissible ordinals I, which are closed under the step to the next admissible ordinal, i.e. if $\alpha < \mathrm{I}$, then $\alpha^+ < \mathrm{I}$, where $\underline{\alpha^+}$ is the next admissible ordinal above κ.

Richter has shown ([76], see as well [1] Prop. 3.5.1, [43]) that the first recursively inaccessible is the limit of the closure ordinals of $[\Pi^0_1, \Pi^0_0]$-*non-monotone inductive definitions*. An *operator* is here a function from sets to sets. For any operator Γ we define by recursion on the set theoretic ordinals $\underline{\Gamma^\alpha} := \Gamma(\Gamma^{<\alpha}) \cup \Gamma^{<\alpha}$ where $\underline{\Gamma^{<\alpha}} = \bigcup_{\beta<\alpha} \Gamma^\beta$. The *closure ordinal* of an operator Γ is the least α s.t. $\Gamma(\Gamma^\alpha) = \Gamma^\alpha$ (which can be shown in ZFC to exists by a cardinality argument, since $(\Gamma^\alpha)_{\alpha\in\mathrm{Ord}}$ form an increasing sequence of subsets of $\mathbb{N}$). A $\underline{[\Pi^0_1, \Pi^0_0]}$-non-monotone operator is an operator Γ s.t. $\Gamma(X) = \{x \in \mathbb{N} \mid x \in \Gamma_0(X) \vee (\Gamma_0(X) \subseteq X \wedge x \in \Gamma_1(X))\}$, where Γ_0 is given by a Π^0_1-formula and Γ_1 by a Π^0_0-formula. So iteration of Γ means that we iterate Γ_0, until we have reached a fixed point. Then we apply Γ_1 to it. Then we continue applying Γ_0, until we reach the next fixed point, etc. This characterisation of the first recursively inaccessible ordinal as the limit of such closure ordinals makes precise the statement that the first recursively inaccessible corresponds to an inductive definition closed under the step to the next inductive definition.

We will not make use of this characterisation of the first recursively inaccessible ordinal, but illustrate, in which sense a recursively inaccessible ordinal occurs naturally when constructing a model of $\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{U}$: We define by recursion on α sets of terms $\underline{\mathrm{U}^\alpha}$ and functions $\underline{\mathrm{T}^\alpha}$, mapping elements of U^α to sets of terms, s.t. $(\mathrm{U}^\alpha, \mathrm{T}^\alpha)_{\alpha\in\mathrm{Ord}}$ is an *increasing sequence of family of sets*. This means that, if $\alpha < \beta$, then $\mathrm{U}^\alpha \subseteq \mathrm{U}^\beta$ and $\mathrm{T}^\beta \restriction \mathrm{U}^\alpha = \mathrm{T}^\alpha$. Let $\underline{\mathrm{U}^{<\alpha}} = \bigcup_{\beta<\alpha} \mathrm{U}^\beta$, similarly for $\underline{\mathrm{T}^{<\alpha}}$. Then U^α is defined by closing $(\mathrm{U}^{<\alpha}, \mathrm{T}^{<\alpha})$ under the one step application of the closure operators for the universe and under reductions. So we have if a reduces to $b \in \mathrm{U}^{<\alpha}$, then $a \in \mathrm{U}^{\alpha}$ and $\mathrm{T}^{\alpha}(a) = \mathrm{T}^{<\alpha}(b)$. $\widehat{\mathbb{N}} \in \mathrm{U}^{\alpha}$ and $\mathrm{T}^{\alpha}(\widehat{\mathbb{N}}) = [\![\, \mathbb{N} \,]\!]$. If $a \in \mathrm{U}^{<\alpha}$ and $b : \mathrm{T}^{<\alpha}(a) \,[\![\rightarrow]\!]\, \mathrm{U}^{<\alpha}$, then $\widehat{\mathrm{W}}\ a\ b \in \mathrm{U}^{\alpha}$ and $\mathrm{T}^{\alpha}(\widehat{\mathrm{W}}\ a\ b) = [\![\, \mathrm{W} \,]\!](\mathrm{T}^{<\alpha}(a), \lambda y \in \mathrm{T}^{<\alpha}(a).\mathrm{T}^{<\alpha}(b\ y))$. Then we have that $\mathrm{U}^{\alpha}, \mathrm{T}^{\alpha} \in \mathrm{L}_{\tau_{1+\alpha}}$. The reason why we need to refer to $\tau_{1+\alpha}$ is that $\mathrm{T}^{\alpha}(\widehat{\mathrm{W}}\ a\ b)$ is obtained by iterating the operator for inductively defining the W-type up to the next admissible ordinal, so in each step of the construction of the universe, we have to go to the next admissible ordinal.

We can define $\underline{[\![\, \mathrm{U} \,]\!]} = \mathrm{U}^{<\mathrm{I}}$, $\underline{[\![\, \mathrm{T}\ a \,]\!]} = \mathrm{T}^{<\mathrm{I}}(a)$, where $\underline{\mathrm{I}}$ is the first recursively inaccessible ordinal. The main case in showing that $[\![\, \mathrm{U} \,]\!]$ is closed under the introduction rule for the universe is to prove that if $a \in [\![\, \mathrm{U} \,]\!]$ and $b \in \mathrm{T}^{<\mathrm{I}}(a) \,[\![\rightarrow]\!]\, [\![\, \mathrm{U} \,]\!]$, then $\widehat{\mathrm{W}}\ a\ b \in [\![\, \mathrm{U} \,]\!]$. This can be shown using the fact that $\mathrm{T}^{<\mathrm{I}}(a) \in \mathrm{L}_{\gamma}$ for some $\gamma < \mathrm{I}$ (using the fact that I is closed under $\lambda\alpha.\tau_{\alpha}$) and the fact that I is admissible, from which one can deduce that $a \in \mathrm{U}^{<\alpha}$ and $b \in \mathrm{T}^{<\alpha}(a) \,[\![\rightarrow]\!]\, \mathrm{U}^{<\alpha}$ for some $\alpha < \mathrm{I}$, so $\widehat{\mathrm{W}}\ a\ b \in \mathrm{U}^{\alpha} \subseteq [\![\, \mathrm{U} \,]\!]$.

Using this construction (see [81, 91] for details) the author has given a model of $\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{U}$ in the theory $\underline{\mathrm{KPI}^{+}}$. Here KPI^{+} is a variant of KP, which formalises the existence of one recursively inaccessible ordinal and finitely many admissible ordinals above it. We therefore obtain that $|\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{U}| \leq |\mathrm{KPI}^{+}|$. The author has shown using a technically complex well-ordering proof ([81, 84, 83]) that this bound is sharp, so we have $|\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{U}| = |\mathrm{KPI}^{+}| = \psi_{\Omega_1}(\Omega_{\mathrm{I}+\omega})$. Here $\underline{\Omega_{\alpha}} = \aleph_{\alpha}$, and $\underline{\psi_{\kappa}}$ is the collapsing function, which collapses ordinals to ordinals $< \kappa$. ψ_{κ} is the main function used in impredicative ordinal notation systems and has a technically difficult definition. $\underline{\mathrm{I}}$ is the first inaccessible cardinal. With some effort cardinals can be replaced by their recursive analogues (admissibles and limits of admissibles; see [79]). We note here that a variant of this theory has been analysed [37] by E. Griffor and M. Rathjen. Note that the above gives a constructive justification of KPI^{+}, and, since the theories $\underline{(\Delta^1_2 - \mathrm{CA}) + (\mathrm{BI})}$ (second order analysis with comprehension restricted to Δ^1_2-formulae and the addition of bar induction) and $\underline{\mathrm{KPI}}$ (KP plus inaccessibility of the universe) are proof theoretically slightly weaker, as well of those two theories.

Many more types have been added to MLTT, but the W-type and extended universes are the main ingredients of standard extensions of MLTT, which add proof theoretic strength. E. Palmgren has added superuniverses ([51], (Rathjen [68, 70] has analysed the variants one obtains by omitting the W-type, which are very weak) and higher type universes to type theory ([53]), which go substantially beyond the theory above, but are instances of indexed inductive-recursive definitions as discussed in the next section.

From a consistency point of view it seems that once one has accepted the W-type, it is easy to accept as well a universe. A universe is, although technically slightly more complicated than the W-type, nothing but an inductive definition, in which we define, whenever we have introduced a new element, recursively T applied to this element, and make use of T applied to previous elements.

4 Inductive-Recursive Definitions and the Mahlo Universe

The logical framework. The theory of inductive-recursive definitions is best introduced using the *logical framework (LF)*. There we have two type levels, namely Set and Type. We have Set : Type, and if A : Set then A : Type.

Set and Type are closed under the LF *dependent function type* $\underline{(x : A) \rightarrow B[x]}$. So if A : Type, and $x : A \Rightarrow B[x]$: Type then $(x : A) \rightarrow B[x]$: Type, similarly with Type replaced by Set. $(x : A) \rightarrow B[x]$ has essentially the same rules as $\Pi x : A.B[x]$, and in addition the η-rule. Elements introduced by the introduction rule are denoted by $(y)b[y]$ where $y : A \Rightarrow b[y] : B[y]$. We will not use the Π-type in this article in the presence of the LF, and redefine, when using the LF, $\lambda y.b[y] := (y)b[y]$, which is a more readable notation, and $A \to B := (x : A) \to B$ for some fresh variable x.

In order to easily formulate the theory of inductive recursive definitions, one closes Set and Type as well under the *dependent sum type* $(x : A) \times B[x]$ with the same formation rule as $(x : A) \to B[x]$. $(x : A) \times B[x]$ has essentially the same introduction, elimination and equality rules as $\Sigma x : A.B[x]$, and in addition the η-rule. Elements introduced by the introduction rule are denoted by $\langle a, b\rangle : (x : A) \times B[x]$ for $a : A$ and $b : B[a]$, and the elimination rule is given as projections: for $a : (x : A) \times B[x]$ we have $\pi_0\ a : A$ and $\pi_1\ a : B[\pi_0\ a]$.

We add as well $\{*\}$: Set, which has essentially the same rules as N_1 and in addition the η-rule.

The LF allows to introduce type constructors in a more convenient way: For instance, in case of a universe U, T we have $\mathrm{T} : \mathrm{U} \to \mathrm{Set}$ (for which we need that $\mathrm{U} \to \mathrm{Set} : \mathrm{Type}$), rather than having to write $x : \mathrm{U} \Rightarrow \mathrm{T}\ x : \mathrm{Set}$ with an extra rule that if $u = u' : \mathrm{U}$ then $\mathrm{T}\ u = \mathrm{T}\ u' : \mathrm{Set}$.

When modelling type theory with the LF, one has, in order to make sense of types such as $\mathrm{Set} \to \mathrm{Set}$, to introduce an interpretation $[\![\, \mathrm{Set}\,]\!]$ of Set, which will usually be a collection of terms (or more precisely a PER). This means that in the model $[\![\, \mathrm{Set}\,]\!]$ is a closed object. Without the LF, there is no need to define $[\![\, \mathrm{Set}\,]\!]$, all what is needed is to make sure that, if we derive A : Set, then $[\![\, A\,]\!]$ is defined and fulfils certain correctness conditions. Therefore, the LF adds complications to the model construction and to the meaning explanations, and therefore we usually avoid the LF. We note that the LF usually doesn't add any strength. In order to reflect this in the model, one can interpret elements of Set as sets, and elements of Type as classes in set theory.

Inductive recursive definitions (IRD) were introduced by P. Dybjer [26] as a concept, which generalises inductive definitions and universes. IRD allow to define $\mathrm{U} : \mathrm{Set}$ and $\mathrm{T} : \mathrm{U} \to D$ for an arbitrary type D, using strictly positive constructors C. In case of $D = \{*\}$ the function T is trivial (by the η-rule it is $\lambda x.*$), and we obtain strictly-positive inductive definitions. In case of $D = \mathrm{Set}$ we obtain universes.

The idea behind an *IRD* is that we define U inductively. Whenever we introduce an element of U, we recursively define T. Therefore, when referring in the inductive definition of U to elements of U previously introduced, we can make use of T applied to them.

The constructors of U are supposed to be *strictly positive* in U, which means the following: The constructors have the form $\mathrm{C} : (x_1 : A_1, \ldots, x_n : A_n) \to \mathrm{U}$, where the set A_i refers either to sets introduced before one started to define U, T, or $A_i = (y_1 : B_1, \ldots, y_l : B_l) \to \mathrm{U}$, where B_j were introduced before introducing U, T. If A_i doesn't refer to U, then x_i is called a *non-inductive argument*, otherwise it is called an *inductive argument*.

What is crucial is the dependency of A_i on $x_1, \ldots, x_{i-1}$. A_i can directly depend on a non-inductive argument x_j for $j < i$. If x_j is an inductive argument $x_j : A_i = (y_1 : B_1, \ldots, y_l : B_l) \to \mathrm{U}$, we cannot make use of x_j directly: we are still at the stage of forming the set U, and have therefore not defined yet how to introduce other sets from an element in U. However, we can make use of the simultaneously recursively defined function T applied to it, i.e. on $\mathrm{T}\ (x_j\ b_1 \cdots b_l)$, which has been introduced before adding $(\mathrm{C}\ x_1 \cdots x_n)$ to U.

In order to define T recursively, one has to define $\mathrm{T}\ (\mathrm{C}\ a_1 \cdots a_n)$, and this can be defined in an arbitrary way (by using sets introduced before U, T are introduced), but the dependency on the arguments a_i is the same as the dependency of later arguments on previous arguments.

If one wants to make the above precise, one sees that before we can introduce an IRD we are required to carry out some derivations beforehand. For instance, in order to accept the constructor $\mathrm{C} : (x : A, y : B[x] \to \mathrm{U}, z : E[x,y]) \to \mathrm{U}$, $\mathrm{T}\ (\mathrm{C}\ x\ y\ z) = F[x,y,z]$, one needs first to derive the following judgements:

- $A : \mathrm{Set}$;
- $x : A \Rightarrow B[x] : \mathrm{Set}$;
- $E[x,y] = E'[x, \mathrm{T} \circ y]$ for some E' s.t. $x : A, y' : B[x] \to D \Rightarrow E'[x,y'] : \mathrm{Set}$;
- $F[x,y,z] = F'[x, \mathrm{T} \circ y, z]$, s.t. $x : A, y' : B[x] \to D, z : E'[x,y'] \Rightarrow F'[x,y',z] : \mathrm{Set}$.

Therefore, a precise formalisation of a theory of IRDs needs to interleave derivations with the introduction of new IRDs. In order to define such a theory using finitely many rules only, the author has together with Peter Dybjer [27, 29] introduced a data type $\underline{\mathrm{OP}_D}$ of IRDs $\mathrm{U} : \mathrm{Set}$, $\mathrm{T} : \mathrm{U} \to D$. Elements of OP_D are codes for IRDs. If $\gamma : \mathrm{OP}_D$, then one has $\mathrm{U}_\gamma : \mathrm{Set}$ and $\mathrm{T}_\gamma : \mathrm{U}_\gamma \to D$. So, OP_D together with $\lambda\gamma.\mathrm{U}_\gamma$ and $\lambda\gamma.\mathrm{T}_\gamma$ forms a generalised universe (which is a true type and no longer inductive-recursive), the elements of which are inductive-recursively defined sets.

In his original paper [26], P. Dybjer introduced in fact a slight generalisation of IRD, called *indexed inductive-recursive definitions* (*IIRD*s). The generalisation allows to define inductively simultaneously several sets $\mathrm{U}\ i : \mathrm{Set}$ for $i : I$, where $I : \mathrm{Set}$ is an index set introduced before, while simultaneously recursively defining $\mathrm{T}\ i : \mathrm{U}\ i \to D[i]$. Here $i : I \Rightarrow D[i] : \mathrm{Type}$. A closed formalisation of IIRDs was formulated by the author and P. Dybjer in [28, 30].

Apart from the LF we need only rules for introducing IIRDs, all standard set constructions are instances of IIRDs. The theory of IIRDs allows to define practically all standard extensions of MLTT considered in the literature including Palmgren's super universes [51] and higher type universes [53].

At present, we have introduced only a full set-theoretic model of the theories of IRD and IIRD, which doesn't provide any realistic proof theoretic bound. However, it seems not too complicated to transform this model into a model in the theory KPM^+. Here $\underline{\mathrm{KPM}^+}$ is the variant of KP, which formulates the existence of a recursively Mahlo ordinal and finitely many admissibles above it. A *recursively Mahlo ordinal* is an admissible ordinal M such that, whenever we have for a Δ_0-formula φ that $\forall x \in \mathrm{L}_\mathrm{M}.\exists y \in \mathrm{L}_\mathrm{M}.\varphi(x,y)$ holds, then there exists an admissible ordinal $\kappa < \mathrm{M}$ such that $\forall x \in \mathrm{L}_\kappa.\exists y \in \mathrm{L}_\kappa.\varphi(x,y)$. The model construction for the theory of IRD and IIRD is rather complicated, we will sketch a model for the Mahlo universe below, which is a theory slightly stronger than the theory of IIRD, and which captures the essence of IIRD.

If such a model is established, then we obtain $|\mathrm{IRD}| \leq |\mathrm{IIRD}| \leq |\mathrm{KPM}^+| = \psi_{\Omega_1}(\Omega_{\mathrm{M}+\omega})$, where $\underline{\mathrm{M}}$ is a Mahlo cardinal. A lower bound can be obtained by showing that IRD and IIRD allow to interpret the *red Mahlo universe*, which is a variant of the ordinary Mahlo universe, sometimes called the *black Mahlo universe*. We will first develop the black Mahlo universe, and then develop the red Mahlo universe and illustrate, what is known about their strengths. We will see that $\psi_{\Omega_1}(\epsilon_{\mathrm{M}+1}) \leq |\mathrm{IIRD}| \leq |\mathrm{IRD} + \mathrm{ext}|$, where $\psi_{\Omega_1}(\epsilon_{\mathrm{M}+1})$ is slightly smaller than the upper bound $\psi_{\Omega_1}(\Omega_{\mathrm{M}+\omega})$ and $\underline{\mathrm{ext}}$ are the rules of extensionality. The precise strength of IIRD and IRD is unknown at present.

The (black) Mahlo universe is a universe, which makes use of non-positive constructors. It was developed by the author [86] in order to capture the notion of a recursively Mahlo ordinal: Remember that a recursively Mahlo ordinal is an admissible ordinal M s.t. whenever we have for a Δ_0-formula φ that $\psi := \forall x \in \mathrm{L}_\mathrm{M}.\exists y \in \mathrm{L}_\mathrm{M}.\varphi(x,y)$ holds, then there exists an admissible ordinal $\kappa < \mathrm{M}$ s.t. $\psi' := \forall x \in \mathrm{L}_\kappa.\exists y \in \mathrm{L}_\kappa.\varphi(x,y)$ holds as well. If we replace "admissible" by "recursively inaccessible", we obtain an equivalent definition. Recursively inaccessible ordinals correspond in type theory to universes. The existence of a recursively inaccessible M is translated into the existence of a universe $\underline{\mathrm{V}} : \mathrm{Set}$, $\underline{\mathrm{T}_\mathrm{V}} : \mathrm{V} \to \mathrm{Set}$, which is closed under the standard set constructions. L_M are sets definable up to level M, which correspond to elements of Fam(V) for the corresponding universe. Here $\underline{\mathrm{Fam}(\mathrm{V})} := (x : \mathrm{V}) \times (\mathrm{T}_\mathrm{V}\ x \to \mathrm{V})$ is the set of families of sets in V, where $\langle a, b\rangle : \underline{\mathrm{Fam}(\mathrm{V})}$ is the family, which using pseudo set theoretic notation might be denoted by $\{\mathrm{T}_\mathrm{V}\ (b\ x) \mid x : \mathrm{T}_\mathrm{V}\ a\}$. The formula ψ can be represented in type theory as the assumption of some $f : \mathrm{Fam}(\mathrm{V}) \to \mathrm{Fam}(\mathrm{V})$. The existence of κ is translated into the existence of a subuniverse $\underline{\mathrm{U}_f} : \mathrm{Set}$, $\underline{\widehat{\mathrm{T}}_f} : \mathrm{U}_f \to \mathrm{V}$. Let $\underline{\mathrm{T}_f} := \mathrm{T}_\mathrm{V} \circ \widehat{\mathrm{T}}_f$. That κ is recursively inaccessible translates into $\widehat{\mathrm{U}}_f, \widehat{\mathrm{T}}_f$ being closed under the standard set constructions. Let $\underline{\mathrm{Fam}(\mathrm{U}_f)} := (x : \mathrm{U}_f) \times (\mathrm{T}_f\ x \to \mathrm{U}_f)$. Then that L_κ is closed under ψ' is represented as that U_f has constructors, which reflect f: Assume $a : \mathrm{Fam}(\mathrm{U}_f)$, and lift it to an element $a_\mathrm{V} : \mathrm{Fam}(\mathrm{V})$. Then the constructors introduce the two components of an element $b : \mathrm{Fam}(\mathrm{U}_f)$, such that its lifting to Fam(V) is equal to $(f\ a_\mathrm{V})$. Finally, we need to model that $\kappa \in \mathrm{M}$, which is expressed as the existence of $\underline{\widehat{\mathrm{U}}_f} : \mathrm{V}$ s.t. $\mathrm{T}_\mathrm{V}\ \widehat{\mathrm{U}}_f = \mathrm{U}_f$.

The precise formulation of the Mahlo universe avoids the LF. For this we uncurry f and split it into two functions. So assume $f_0 : (x : \mathrm{V}, y : \mathrm{T}_\mathrm{V}\ x \to \mathrm{V}) \to \mathrm{V}$ and $f_1 : (x : \mathrm{V}, y : \mathrm{T}_\mathrm{V}\ x \to \mathrm{V}, z : \mathrm{T}_\mathrm{V}(f_0\ x\ y)) \to \mathrm{V}$, and let $\vec{f} := f_0, f_1$. (So, from f above we would obtain $f_0 = \lambda x.\lambda y.\pi_0\ (f\ \langle x, y\rangle)$ and $f_1 = \lambda x.\lambda y.\lambda z.(\pi_1\ (f\ \langle x, y\rangle))\ z$.) In the following, when we say that $\vec{f}$ is a *function from families of sets inside a universe to itself*, we mean that $\vec{f} = f_0, f_1$ for the two uncurried components of such a function. Assuming such $\vec{f}$, there exists $\underline{\mathrm{U}_{\vec{f}}} : \mathrm{Set}$ and $\underline{\widehat{\mathrm{T}}_{\vec{f}}} : \mathrm{U}_{\vec{f}} \to \mathrm{V}$, s.t. with $x : \mathrm{U}_{\vec{f}} \Rightarrow \mathrm{T}_{\vec{f}}\ x := \mathrm{T}_\mathrm{V}\ (\widehat{\mathrm{T}}_{\vec{f}}\ x) : \mathrm{Set}$ we have:

- $\mathrm{U}_{\vec{f}}, \widehat{\mathrm{T}}_{\vec{f}}$ is closed under the standard set constructions. Closure under $\mathbb{N}$ means that we have $\underline{\widehat{\mathrm{N}}_{\vec{f}}} : \mathrm{U}_{\vec{f}}$, $\widehat{\mathrm{T}}_{\vec{f}}\ \widehat{\mathrm{N}}_{\vec{f}} = \widehat{\mathrm{N}}_\mathrm{V} : \mathrm{V}$, where $\mathrm{T}_\mathrm{V}\ \widehat{\mathrm{N}}_\mathrm{V} = \mathbb{N}$. Closure under W means that we have if $a : \mathrm{U}_{\vec{f}}$ and $b : \mathrm{T}_{\vec{f}}\ a \to \mathrm{U}_{\vec{f}}$, then $\underline{\widehat{\mathrm{W}}_{\vec{f}}\ a\ b} : \mathrm{U}_{\vec{f}}$ and $\widehat{\mathrm{T}}_{\vec{f}}\ (\widehat{\mathrm{W}}_{\vec{f}}\ a\ b) = \widehat{\mathrm{W}}_\mathrm{V}\ (\widehat{\mathrm{T}}_{\vec{f}}\ a)\ (\widehat{\mathrm{T}}_{\vec{f}} \circ b)$ where $\mathrm{T}_\mathrm{V}\ (\widehat{\mathrm{W}}_\mathrm{V}\ c\ d) = \mathrm{W}x : (\mathrm{T}_\mathrm{V}\ c).\mathrm{T}_\mathrm{V}\ (d\ x)$.
- $\mathrm{U}_{\vec{f}}, \widehat{\mathrm{T}}_{\vec{f}}$ is closed under f_0, f_1. So we have, if $a : \mathrm{U}_{\vec{f}}$ and $b : \mathrm{T}_{\vec{f}}\ a \to \mathrm{U}_{\vec{f}}$, and $a_\mathrm{V} := \widehat{\mathrm{T}}_{\vec{f}}\ a : \mathrm{V}$, $b_\mathrm{V} := \widehat{\mathrm{T}}_{\vec{f}} \circ b : \mathrm{T}_\mathrm{V}\ a_\mathrm{V} \to \mathrm{V}$ are the corresponding elements in V, then $\underline{\widehat{\mathrm{f}}_{\vec{f},0}\ a\ b} : \mathrm{U}_{\vec{f}}$ and $\widehat{\mathrm{T}}_{\vec{f}}\ (\widehat{\mathrm{f}}_{\vec{f},0}\ a\ b) = f_0\ a_\mathrm{V}\ b_\mathrm{V}$. If, in addition, $c : \mathrm{T}_{\vec{f}}\ (f_0\ a_\mathrm{V}\ b_\mathrm{V})$, then $\underline{\widehat{\mathrm{f}}_{\vec{f},1}\ a\ b\ c} : \mathrm{U}_{\vec{f}}$ and $\widehat{\mathrm{T}}_{\vec{f}}\ (\widehat{\mathrm{f}}_{\vec{f},1}\ a\ b\ c) = f_1\ a_\mathrm{V}\ b_\mathrm{V}\ c$.
- We have a constructor $\underline{\widehat{\mathrm{U}}_{\vec{f}}} : \mathrm{V}$ s.t. $\mathrm{T}_\mathrm{V}\ \widehat{\mathrm{U}}_{\vec{f}} = \mathrm{U}_{\vec{f}}$.

A model of the Mahlo universe can be constructed in KPM^+ as follows (some of the ideas of this model are due to U. Berger (private communication)): For simplicity one treats $\mathrm{U}_{\vec{f}}$ as a subset of V, and therefore treats $\widehat{\mathrm{T}}_{\vec{f}}$ as the identity function, identifies $\widehat{\mathrm{N}}_{\vec{f}}$ with $\widehat{\mathrm{N}}_\mathrm{V}$, $\widehat{\mathrm{W}}_{\vec{f}}$ with $\widehat{\mathrm{W}}_\mathrm{V}$, etc., $\widehat{\mathrm{f}}_{f_0,f_1,0}$ with f_0, and $\widehat{\mathrm{f}}_{f_0,f_1,1}$ with f_1. Furthermore, we omit the subscript V of T_V, $\widehat{\mathrm{W}}_\mathrm{V}$ etc. As for the universe one defines by recursion on α sets $\underline{\mathrm{V}^\alpha}$ and functions $\underline{\mathrm{T}^\alpha}$ with domain V^α by closing them under the same operations as before, s.t. $(\mathrm{V}^\alpha, \mathrm{T}^\alpha)_{\alpha \in \mathrm{Ord}}$ is an increasing sequence of families of sets.

In addition one adds closure under $\widehat{\mathrm{U}}_{\vec{f}}$ as follows: One defines for every pair of terms $\vec{f} = f_0, f_1$ a set $\underline{\mathrm{U}^\alpha_{\vec{f}}} \subseteq \mathrm{V}$. $\mathrm{U}^\alpha_{\vec{f}}$ is the least subset of V^α which is closed under the set constructions and under $\vec{f}$, provided the results are in $V^{<\alpha}$. So if $a \in \underline{U^{<\alpha}_{\vec{f}}}$ $(:= \bigcup_{\beta<\alpha} U^{\beta}_{\vec{f}})$, $b \in T^{<\alpha}(a)[\![\rightarrow]\!] U^{<\alpha}_{\vec{f}}$, and $\widehat{W}\ a\ b \in V^{<\alpha}$, then $\widehat{W}\ a\ b \in U^{\alpha}_{\vec{f}}$, similarly for the other standard set constructions. If for the same a, b we have $f_0\ a\ b \in V^{<\alpha}$, then $f_0\ a\ b \in U^{\alpha}_{\vec{f}}$. If furthermore $c \in T^{<\alpha}(f_0\ a\ b)$ and $f_1\ a\ b\ c \in V^{<\alpha}$, then $f_1\ a\ b\ c \in U^{\alpha}_{\vec{f}}$.

By $U^{\alpha}_{\vec{f}}$ being *closed* we mean that the condition of $V^{<\alpha}$ in the definition of $U^{\alpha}_{\vec{f}}$ is always fulfilled: we have under the above conditions for a and b that $\widehat{W}\ a\ b \in V^{<\alpha}$ and $f_0\ a\ b \in V^{<\alpha}$ hold, and for the c as above we have $f_1\ a\ b\ c \in V^{<\alpha}$. Note that if $U^{\alpha}_{\vec{f}}$ is closed, then, because of the fact that if $\alpha < \beta$ then $V^{\alpha} \subseteq V^{\beta}$ and $T^{\beta} \restriction V^{\alpha} = T^{\alpha}$ we have that $U^{\alpha}_{\vec{f}} = U^{\beta}_{\vec{f}}$ for all $\beta > \alpha$, i.e. $U^{\beta}_{\vec{f}}$ doesn't change any more. If $U^{\alpha}_{\vec{f}}$ is closed, then $U^{\alpha}_{\vec{f}}$ is a complete subuniverse of V closed under $\vec{f}$, and we add $\widehat{U}_{\vec{f}}$ to V^{α}, with $T^{\alpha}(\widehat{U}_{\vec{f}}) = U^{\alpha}_{\vec{f}}$.

Finally, one defines for a recursively Mahlo ordinal M the interpretation of V as $\underline{[\![V]\!]} := V^{M}$, $\underline{[\![T_V\ a]\!]} := T^{M}(a)$, $\underline{[\![\widehat{U}_{\vec{f}}]\!]} := U^{M}_{\vec{f}}$. That $[\![V]\!]$ is closed under the introduction rule for $\widehat{U}_{\vec{f}}$ is shown as follows: Assume f_0, f_1 is a function from elements of $[\![V]\!]$ to itself. This property can be, using the fact that M is admissible, expressed as a formula $\forall\alpha < M.\exists\beta < M.\varphi(\alpha,\beta)$, which means that $U^{M}_{\vec{f}}$ is closed (we use the fact that if α is inaccessible, then V^{α} is closed under the standard set constructions). Then this property holds as well with M replaced by some inaccessible $\kappa <$ M. But then V^{κ} and therefore as well $U^{\kappa}_{\vec{f}}$ are closed under the standard set constructions, and by $\forall\alpha < \kappa.\exists\beta < \kappa.\varphi(\alpha,\beta)$ it follows that $U^{\kappa}_{\vec{f}}$ is closed under f_0, f_1. Therefore, $U^{\kappa}_{\vec{f}}$ is closed and we obtain $\widehat{U}_{\vec{f}} \in V^{\kappa} \subseteq [\![V]\!]$.

It is easy to verify that all other rules (including those for $U_{\vec{f}}$, $\widehat{T}_{\vec{f}}$) hold as well. Note that this model construction is predicative in nature: While defining approximations of V, we were never referring to V as a whole, but only to elements of the approximation. Especially the definition of $U^{\alpha}_{\vec{f}}$ only refers to elements of $V^{<\alpha}$, i.e. elements defined before. Therefore, the model construction reveals that the Mahlo universe is predicative in nature, and gives an insight into its consistency.

One should note however that $[\![V]\!]$ contains more elements then those justified by the introduction rules for the Mahlo universe: When adding $\widehat{U}_{\vec{f}}$ to V^{α} we have only guaranteed that there exists a subuniverse of V^{α} closed under $\vec{f}$. This doesn't guarantee that $\vec{f}$ forms a function from families of $[\![V]\!]$ to itself, only that it is one on families of $[\![U_{\vec{f}}]\!]$.

It would be very satisfactory if one could define a variant of the Mahlo universe, which is in accordance with its standard model, so that one could see immediately that it is predicative in nature, and that one could develop meaning explanations based on this idea. The problem is that the model refers to the collection of all terms (when referring to arbitrary f_0, f_1), and we don't have access to the collection of all terms in MLTT.

Let $\underline{\text{MLTT} + \mathbb{N} + \text{W} + \text{Mahlo}}$ be the type theory formulating the Mahlo universe closed under the W-type. Using the model [82, 91] the author has shown that $|\text{MLTT} + \mathbb{N} + \text{W} + \text{Mahlo}| \leq |\text{KPM}^{+}| = \psi_{\Omega_1}(\Omega_{M+\omega})$, and, using a well-ordering proof [86], he has shown as well that $\psi_{\Omega_1}(\Omega_{M+\omega}) \leq |\text{MLTT} + \mathbb{N} + \text{W} + \text{Mahlo}|$, therefore the bound is sharp. Here $\underline{M}$ is a Mahlo cardinal. If one accepts the consistency of the Mahlo universe one gets therefore a constructive justification of the rather strong theories KPM^{+} and KPM (where $\underline{\text{KPM}}$ is the extension of KP by the fact that the set theoretic universe has the Mahlo property).

The red Mahlo Universe. When looking at IIRDs, we see that the type Set has the same closure properties as the Mahlo universe: Assume $\vec{f} = f_0, f_1$ are the two components of a function $((A : \mathrm{Set}) \times (A \to \mathrm{Set})) \to ((A : \mathrm{Set}) \times (A \to \mathrm{Set}))$. Then the rules of induction-recursion show that there exists a universe $\underline{\mathrm{U}_{\vec{f}}} : \mathrm{Set}$, $\underline{\mathrm{T}_{\vec{f}}} : \mathrm{U}_{\vec{f}} \to \mathrm{Set}$, s.t. $\mathrm{U}_{\vec{f}}$ is closed under the standard set constructions and under $\vec{f}$.

We can now define a theory of the red Mahlo universe as follows: It contains the rules for the standard set constructions, the rules LF, and the rules <u>RedMahlo</u> expressing that for every $\vec{f}$ as above there exists a universe $\mathrm{U}_{\vec{f}}$, $\mathrm{T}_{\vec{f}}$ closed under $\vec{f}$ and the standard set constructions. Furthermore, we add *large elimination* (i.e. elimination into a type, which might depend on the argument, rather than into a set) for the W-type and for $\mathbb{N}$. We call the resulting theory $\underline{\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{LF} + \mathrm{RedMahlo}}$. The well-ordering proof for the black Mahlo universe can be easily be adapted in order to show $|\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{LF} + \mathrm{RedMahlo}| \geq \psi_{\Omega_1}(\epsilon_{\mathrm{M}+\omega}) = |\mathrm{KPM}|$ (see the proof by the author in [29]). It should not be too difficult to define a model of $\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{LF} + \mathrm{RedMahlo}$ by interpreting sets in type theory as PERs which are sets, types as PERs which are classes, and $\mathrm{U}_{\vec{f}}$ as the least set closed under $\vec{f}$ and the standard set constructions (assuming $\vec{f}$ is in KPM a term representing a function from families of sets to families of sets). This would show that $|\mathrm{MLTT} + \mathbb{N} + \mathrm{W} + \mathrm{LF} + \mathrm{RedMahlo}| \leq |\mathrm{KPM}|$, and therefore that this bound is sharp.

The red Mahlo universe is a subtheory of the theory of IIRD, which in turn can at least in the presence of an extensional equality be simulated by IRD (unpublished result by P. Dybjer and the author). Assuming large elimination, the red Mahlo universe is as well directly a subtheory of IRD. That the theory of the red Mahlo universe can be interpreted in IIRD even without large elimination can be seen as follows: $\mathrm{MLTT} + \mathbb{N} + \mathrm{W}$, and U_f are instances of IIRD. Furthermore, large elimination can be simulated by using IIRDs. If one wants to define $f : (y : \mathrm{W}x : A.B[x]) \to D[y]$ where $y : \mathrm{W}x : A.B[x] \Rightarrow D[y] : \mathrm{Type}$, which is defined induction on $\mathrm{W}x : A.B[x]$, then one defines first inductive recursively $\mathrm{U}' : (\mathrm{W}x : A.B) \to \mathrm{Set}$ together with $\mathrm{T}' : (y : \mathrm{W}x : A.B) \to \mathrm{U}'\ y \to D[y]$. If the desired equality rule is $f\ (\mathrm{sup}\ \ a\ \ b) = g\ \ a\ \ b\ \ (f \circ b)$ where $g : (a : A, b : B[a] \to \mathrm{W}x : A.B, (y : B[a]) \to D[b\ y]) \to D[(\mathrm{sup}\ \ a\ \ b)]$ then we take as constructor of U' the function $\mathrm{C} : (a : A, b : B[a] \to \mathrm{W}x : A.B, (y : B[a]) \to \mathrm{U}'[b\ y]) \to \mathrm{U}'\ (\mathrm{sup}\ \ a\ \ b)$ with $\mathrm{T}'\ (\mathrm{sup}\ \ a\ b)\ (\mathrm{C}\ a\ b\ c) = g\ a\ b\ (\lambda y.\mathrm{T}'\ (b\ y)\ (c\ y))$. Now we can define $f' : (y : \mathrm{W}x : A.B) \to \mathrm{U}'\ y$ by $f'\ (\mathrm{sup}\ \ a\ b) = \mathrm{C}\ a\ b\ (\lambda y.f'\ (b\ y))$, define $f = \lambda y.\mathrm{T}'\ y\ (f'\ y)$, and verify that f fulfils the equation $f\ (\mathrm{sup}\ \ a\ b) = g\ a\ b\ (f \circ b)$.

From this it follows that $|\mathrm{KPM}| \leq |\mathrm{MLTT}+\mathbb{N}+\mathrm{W}+\mathrm{LF}+\mathrm{RedMahlo}| \leq |\mathrm{IIRD}| \leq |\mathrm{IRD}+\mathrm{ext}|$ and $|\mathrm{MLTT}+\mathbb{N}+\mathrm{W}+\mathrm{LF}+\mathrm{RedMahlo}| \leq |\mathrm{IRD}+\mathrm{Largeelim}|$, where <u>ext</u> are the rules of extensionality, and <u>Largeelim</u> are the rules of large elimination. This doesn't provide a sharp bound for the theories of IRD and IIRD, we only obtain that their strength (assuming extensionality or large elimination in case of IRD) is in the interval $[|\mathrm{KPM}|, |\mathrm{KPM}^+|]$. Since the black Mahlo universe has strength KPM^+, we can say that it captures the essence of IIRD (for instance, assuming the model of IIRD is developed in KPM^+, it shows the well-foundedness of a sequence of ordinals with supremum the upper bound of the proof theoretic strength of IIRD and therefore the consistency of approximations of IIRD).

5 The Π_3-Reflecting Universe

The Hyper Mahlo Universe. The first step towards the Π_3-reflecting universe is the hyper-Mahlo Universe. The *hyper-Mahlo Universe* is a universe $\underline{\mathrm{U}_2}, \underline{\mathrm{T}_2}$ s.t. for every function $\vec{f} = f_0, f_1$ from families of sets in U_2 to itself, there exist a subuniverse $\underline{\mathrm{U}_{\vec{f},1}} : \mathrm{Set}$ together with $\underline{\widehat{\mathrm{T}}_{\vec{f},1}} : \mathrm{U}_{\vec{f},1} \to \mathrm{U}_2$, which is Mahlo, closed under $\vec{f}$

(all universes in this section will be closed under the standard set constructions as well), and represented in $\mathrm{U}_{\vec{f},1}$ as a code $\underline{\widehat{\mathrm{U}}_{\vec{f},2}}$.

Let $\underline{\mathrm{T}_{\vec{f},1}\ a} = \mathrm{T}_2\ (\widehat{\mathrm{T}}_{\vec{f},1}\ x)$. That $\mathrm{U}_{\vec{f},1}$ is a Mahlo universe is expressed by the rule that for every function $\vec{g}$ from families of sets in $\mathrm{U}_{\vec{f},1}$ to itself there exists a subuniverse $\underline{\mathrm{U}_{\vec{f},\vec{g},0}}$ together with $\underline{\widehat{\mathrm{T}}_{\vec{f},\vec{g},0}} : \mathrm{U}_{\vec{f},\vec{g},0} \to \mathrm{U}_{\vec{f},1}$, which is closed under $\vec{g}$ and represented as a code $\underline{\widehat{\mathrm{U}}_{\vec{f},\vec{g},0}}$ in $\mathrm{U}_{\vec{f},1}$. Note that this means that we need to have an element representing $\mathrm{U}_{\vec{f},\vec{g},0}$ in U_2 as well, in order to define $\widehat{\mathrm{T}}_{\vec{f},1}\ \widehat{\mathrm{U}}_{\vec{f},\vec{g},0}$, which means that there needs to be an additional constructor for U_2. There are two ways for achieving this: One possibility would be to add a constructor which introduces directly a code for $\widehat{\mathrm{T}}_{\vec{f},1}\ \widehat{\mathrm{U}}_{\vec{f},\vec{g},0}$ in U_2. When moving to hyper^n-Mahlo universes and beyond it becomes technically rather complicated to make sure that all elements created by a subuniverse are passed on to all universes containing it. The other more simple possibility is to consider $\widehat{\mathrm{T}}_{\vec{f},1}$ as a constructor of elements of U_2 rather than a recursively defined function. The disadvantage of this is that we get doubling of elements, e.g. $\widehat{\mathbb{N}}_2$ and $\widehat{\mathrm{T}}_{\vec{f},1}\ \widehat{\mathbb{N}}_{\vec{f},1}$ are now two different codes for the natural numbers in U_2. This doesn't cause any problems from a proof theoretic point of view, and reduces the technicalities in the following. So in the following, all subuniverses are given as "*inductive subuniverses*", which means that $\widehat{\mathrm{T}}$ is a constructor.

Although a lot of work needs to be carried out in order to analyse this theory, it seems very plausible that this theory has the strength of $\underline{\mathrm{KP} - \mathrm{Hyper} - \mathrm{M}^+}$, which is the extension of KP by the existence of a recursively hyper-Mahlo ordinal and finitely many admissible ordinals above it. Here a *recursively hyper-recursively Mahlo ordinal* is an ordinal, which fulfils the same condition as a recursively Mahlo ordinal M, except that the $\kappa < \mathrm{M}$ in the definition of a recursively Mahlo ordinal, which was claimed to exist, can be chosen to be recursively Mahlo rather than simply admissible.

The hyper-n- and hyper-α-Mahlo universes. The next step is to define a *hyper-n-Mahlo universe* which is done by simply iterating the step towards the hyper-Mahlo universe further. Similarly, one can define, assuming a fixed type of ordinals Ord, *hyper-α-Mahlo universes* for α : Ord, which correspond to recursively hyper-α-Mahlo ordinals. Here the definition of *recursively hyper-α-Mahlo ordinals* is similar to that of recursively hyper-Mahlo, except that the κ can for any $\beta < \alpha$ be chosen to be hyper-β-Mahlo.

The autonomous Mahlo universe. (See as well the article [89] by the author.) A *recursively autonomous Mahlo ordinal* is an ordinal κ, which is recursively hyper-κ-Mahlo. If we translate recursively inaccessible ordinals as universes, and the ordinals of a universe $\mathrm{V}, \mathrm{T}_\mathrm{V}$ as $\mathrm{W}x : \mathrm{V}.\mathrm{T}_\mathrm{V}\ x$, we arrive at the following definition of an autonomous Mahlo universe: It is a universe $\underline{\mathrm{V}}, \underline{\mathrm{T}_\mathrm{V}}$, which is hyper-$d$-Mahlo for every $d : (\mathrm{W}x : \mathrm{V}.\mathrm{T}_\mathrm{V}\ x)$. Since $(\mathrm{W}x : \mathrm{V}.\mathrm{T}_\mathrm{V}\ x)$ can only be defined once the definition of V is complete, we replace this set by the set $\underline{\mathrm{Deg}}$ of *Mahlo degrees*, which is the least set introduced by introduction rule $\underline{\mathrm{sup}'} : (x : \mathrm{V}, y : \mathrm{T}_\mathrm{V}\ x \to \mathrm{Deg}) \to \mathrm{Deg}$. Deg is isomorphic to $\mathrm{W}x : \mathrm{V}.\mathrm{T}_\mathrm{V}\ x$, but can be defined simultaneously with V and the other constructions defined in the following. We note that elements of Deg depend only locally on V: For every element of d : Deg we can define an a : V and $b : \mathrm{T}_\mathrm{V}\ x \to \mathrm{V}$, s.t. all elements of V used in Deg are of the form $(b\ x)$ for some $x : \mathrm{T}_\mathrm{V}\ x$.

The autonomous Mahlo universe $\mathrm{V}, \mathrm{T}_\mathrm{V}$ is hyper-d-Mahlo for every d : Deg. If we consider $\mathrm{V}, \mathrm{T}_\mathrm{V}$ as being inductively defined by the introduction rules, we see that as new elements are added to V, new elements of Deg, i.e. new Mahlo degrees become available, which result in the possibility of using the Mahlo reflection rule for V into new Mahlo degrees. The existence of the hyper-Mahlo universe claims that this process reaches a fixed point.

The precise formulation of the autonomous Mahlo universe is as follows: We first define depending on d : Deg a set $\underline{\mathrm{Univ}_d}$ of codes for universes of Mahlo degree d, together with for $u : \mathrm{Univ}_d$ the universes $\underline{\mathrm{U}_{d,u}}$: Set, $\underline{\mathrm{T}_{d,u}} : \mathrm{U}_{d,u} \to$ Set given by u. Let $d = \mathrm{sup}'\ a\ b$. That $\mathrm{U}_{d,u}, \mathrm{T}_{d,u}$ has Mahlo degree d means that for every function $\vec{f}$ from families of sets in $\mathrm{U}_{d,u}$ to itself and for every $c : \mathrm{T}_{\mathrm{V}}\ a$, i.e. for every subdegree $d' := b\ c$ of d, there exists a an element $\underline{\mathrm{u}_{0,d,u,\vec{f},c}} : \mathrm{Univ}_{d'}$. Let temporarily $u' := \mathrm{u}_{0,d,u,\vec{f},c}$. Then there exists a constructor $\underline{\widehat{\mathrm{T}}_{0,d,u,\vec{f},c}} : \mathrm{U}_{d',u'} \to \mathrm{U}_{d,u}$ preserving the denoted sets, and $\mathrm{U}_{d',u'}$ is closed under $\vec{f}$. Furthermore, $\mathrm{U}_{d',u'}$ is represented in $\mathrm{U}_{d,u}$, i.e. there exists $\underline{\widehat{\mathrm{U}}_{0,d,u,\vec{f},c}} : \mathrm{U}_{d,u}$ s.t. $\mathrm{T}_{d,u}\ \widehat{\mathrm{U}}_{0,d,u,\vec{f},c} = \mathrm{U}_{d',u'}$.

That $\mathrm{V}, \mathrm{T}_{\mathrm{V}}$ is hyper-d-Mahlo for every d : Deg means that for every such d and function $\vec{f}$ from families of sets in V to itself there exists an element $\underline{\mathrm{u}_{1,d,\vec{f}}} : \mathrm{Univ}_d$. Let temporarily $u' := \mathrm{u}_{1,d,\vec{f}}$. Then there exists a constructor $\underline{\widehat{\mathrm{T}}_{1,d,\vec{f}}} : \mathrm{U}_{d,u'} \to \mathrm{V}$ with $\mathrm{T}_{\mathrm{V}}\ (\widehat{\mathrm{T}}_{1,d,\vec{f}}\ c) = \mathrm{T}_{d,u'}\ c$. $\mathrm{U}_{d,u'}$ is closed under $\vec{f}$. Furthermore, there exists $\underline{\widehat{\mathrm{U}}_{1,d,\vec{f}}} : \mathrm{V}$ s.t. $\mathrm{T}_{\mathrm{V}}\ (\widehat{\mathrm{U}}_{1,d,\vec{f}}\ x) = \mathrm{U}_{1,d,\vec{f}}$.

The Π_3-reflecting Universe. The subuniverses of the autonomous Mahlo universe have *static Mahlo degrees* d : Deg. However, we cannot assign directly to the autonomous Mahlo universe a static Mahlo degree itself – the subdegrees depend on families of sets in $\mathrm{V}, \mathrm{T}_{\mathrm{V}}$ (remember that for each d : Deg we could determine an $a : \mathrm{Fam}(\mathrm{V})$ s.t. d refers only to elements of V in a). In the Π_3-reflecting universe this is generalised by having *dynamic Mahlo degrees*. The new introduction rule for *(dynamic) Mahlo degrees* $\underline{\mathrm{Deg}}$ is that if $f_0 : (a : \mathrm{V}, b : \mathrm{T}_{\mathrm{V}}\ a \to \mathrm{V}) \to \mathrm{V}$, $f_1 : (a : \mathrm{V}, b : \mathrm{T}_{\mathrm{V}}\ a \to \mathrm{V}, c : \mathrm{T}_{\mathrm{V}}\ (f_0\ a\ b)) \to \mathrm{Deg}$ are the two components of a function from families of sets in V to families of elements of Deg indexed by V, then there exists $\underline{\mathrm{sup}'}\ f_0\ f_1$: Deg. $\underline{\mathrm{Univ}_d}$, $\underline{\mathrm{U}_{d,u}}$, $\underline{\mathrm{T}_{d,u}}$ are formed as before, but we have as well a constructor $\underline{\widehat{\mathrm{T}}_{2,d,u}} : \mathrm{U}_{d,u} \to \mathrm{V}$ s.t. $\mathrm{T}_{\mathrm{V}}\ (\widehat{\mathrm{T}}_{2,d,u}\ a) = \mathrm{T}_{d,u}\ a$. Assume $d = \mathrm{sup}'\ f_0\ f_1$, $a : \mathrm{U}_{d,u}$, $b : \mathrm{T}_{d,u}\ a \to \mathrm{U}_{d,u}$, i.e. $\langle a, b\rangle : \mathrm{Fam}(\mathrm{U}_{d,u})$. Let $a_{\mathrm{V}} := \widehat{\mathrm{T}}_{2,d,u}\ a : \mathrm{V}$, $b_{\mathrm{V}} := \widehat{\mathrm{T}}_{2,d,u} \circ b : \mathrm{T}_{\mathrm{V}}\ a \to \mathrm{V}$. Then $\widehat{\mathrm{f}}_{0,d,u}\ a\ b : \mathrm{U}_{d,u}$ and $\mathrm{T}_{d,u}\ (\widehat{\mathrm{f}}_{0,d,u}\ a\ b) = \mathrm{T}_{\mathrm{V}}\ (f_0\ a_{\mathrm{V}}\ b_{\mathrm{V}})$, so the first component of d is reflected in $\mathrm{U}_{d,u}$. Furthermore, assume $c : \mathrm{T}_{\mathrm{V}}\ (f_0\ a_{\mathrm{V}}\ b_{\mathrm{V}})$. Then $d' := f_1\ a_{\mathrm{V}}\ b_{\mathrm{V}}\ c$ is a subdegree of d relative to $\mathrm{U}_{d,u}$ as given by $\langle a, b\rangle : \mathrm{Fam}(\mathrm{U}_{d,u})$. Assume $\vec{g}$ is a function from families of sets in $\mathrm{U}_{d,u}$ to itself. Then there exists a subuniverse of $\mathrm{U}_{d,u}$ of Mahlo degree d' closed under $\vec{g}$ and represented in $\mathrm{U}_{d,u}$. Furthermore, assume d : Deg, $\vec{f}$ a function from families of V to itself. Then there exists a subuniverse of V which is an element of Univ_d, represented in V and closed under $\vec{f}$. This concludes the definition of the Π_3-reflecting universe.

In [89] the author gives a model of the Π_3-reflecting universe in the theory $\mathrm{KP}\Pi_3^+$, which is defined following a similar methodology as the model of the Mahlo universe, although it is technically much more complicated. Here $\underline{\mathrm{KP}\Pi_3^+}$ is the theory, which formulates the existence of a Π_3-reflecting ordinal α, and finitely many admissibles above it. That α is Π_3-*reflecting* means that for every Π_3-formula ψ relative to L_α, where $\psi = \forall x \in \mathrm{L}_\alpha.\exists y \in \mathrm{L}_\alpha.\forall z \in \mathrm{L}_\alpha.\varphi(x, y, z)$, there exists a $u \in \mathrm{L}_\alpha$ which is transitive, not empty, and which reflects ψ, which means $\forall x \in u.\exists y \in u.\forall z \in u.\varphi(x, y, z)$. The model of the Π_3-reflecting universe shows that the Π_3-reflecting universe has at most the strength of $|\mathrm{KP}\Pi_3^+|$. We have a sketch of a well-ordering proof which shows that the Π_3-reflecting universe has the strength of $\mathrm{KP}\Pi_3^+$, therefore this bound is sharp. Therefore we obtain, assuming one accepts the consistency of the Π_3-reflecting universe, a constructive justification of $\mathrm{KP}\Pi_3^+$.

6 Future Work

We have seen in the course of this article that there are many theorems about the precise strength of certain theories, which still need to be worked out in detail. Especially, the precise strength of the theories of IRD and IIRD still needs to be determined. The next steps towards stronger theories seem to be to develop Π_n-reflecting, Π_α-reflecting, and Π^1_1-reflecting universes. The author has some sketches of type theories of that strength, but he doesn't know yet whether they have the strength of KP extended by corresponding recursively large admissibles. A big step would be to reach the strength of $(\Pi^1_2 - \mathrm{CA}) + (\mathrm{BI})$. M. Rathjen has argued in [72] that there is a limit for the proof theoretic strength of extensions of MLTT, which is due to the fact that sets in type theory can be interpreted as non-monotone inductive definitions. This is an interesting point of view, but one might see a paradigm shift happening in type theory, which allows MLTT to go beyond this boundary, see the author's review [90] of Rathjen's article. It remains to be seen whether MLTT can move beyond those boundaries, but we are optimistic that, if a mathematical theory can be analysed proof theoretically, then it will be possible to develop an extension of MLTT, which reaches its strength.

7 Conclusion

We have seen how to develop extensions of MLTT of increasing strength. Whereas for the first theories a direct insight into their consistency was easy, this became increasingly more difficult when moving to stronger theories. This is by Gödel's incompleteness theorem unavoidable. If we formulate any consistency argument mathematically precisely, it becomes a proof in a theory, the strength of which is stronger than the theory in question. Therefore, the stronger the theory, the stronger the theory needs to be in which such a consistency argument is formalised. There are of course technical problems, which add to the complexity of a consistency argument without requiring any strength, and it is a mathematical task to develop theories so that the mathematical technicalities of the consistency argument are as simple as possible. But once one has rid theories of any unnecessary mathematical ballast, the real problem of getting an insight is condensed to having to use principles, which have a certain proof theoretic strength. All we can do is to prove mathematically that certain theories we want to use are equivalent or reducible to certain other reference theories, for which we have an as easy consistency argument as possible. But the strength of the principles used in this consistency argument will necessarily increase and our trust in theories weakens as we proceed along the proof theoretic scale.

References

[1] P. Aczel. An introduction to inductive definitions. In J. Barwise, editor, *Handbook of Mathematical Logic*, chapter C.7, pages 739–782. North-Holland, 1977.

[2] P. Aczel. The strength of Martin-Löf's intuitionistic type theory with one universe. In S. Miettinen and J. Väänänen, editors, *Proceedings of the symposium on mathematical logic (Oulu, 1974)*, University of Helsinki, 1977. Report no. 2, Dept. of philosophy.

[3] T. Arai. Introducing the hardline in proof theory. Draft, 1996.

[4] T. Arai. Proof theory of theories of ordinals I: Reflecting ordinals. Draft, 1996.

[5] T. Arai. Systems of ordinal diagrams. Draft, 1996.

[6] T. Arai. Proof theory of theories of ordinals II: Σ_1 stability. Draft, 1997.

[7] T. Arai. Proof theory of theories of ordinals III: Π_1 collection. Draft, 1997.

[8] T. Arai. A sneak preview of proof theory of ordinals. Revised version of a résumé for a talk at Kobe Seminar on Logic and Computer Science, Dec 1997, 1997.

[9] T. Arai. Ordinal diagrams for Π_3-reflection. *J. Symbolic Logic*, 65(3):1375 – 1394, 2000.

[10] T. Arai. Ordinal diagrams for recursively Mahlo universes. *Arch. Math. Logic*, 39(5):353 – 391, 2000.

[11] T. Arai. Wellfoundedness proofs by means of nonmonotonic inductive definitions II: first order operators. Submitted to Journal of Symbolic Logic, 2002.

[12] T. Arai. Proof theory for theories of ordinals. I. recursively Mahlo ordinals. *Ann. Pure Appl. Logic*, 122(1 – 3):1 – 85, 2003.

[13] T. Arai. Wellfoundedness proofs by means of nonmonotonic inductive definitions I: Π^0_2-operators. *Journal of Symbolic Logic*, 69(3):830 – 850, 2004.

[14] T. Arai. An introduction to ordinal analysis. Course notes. Available from http://kurt.scitec.kobe-u.ac.jp/∼arai/index.html, 2005.

[15] T. Arai. An introduction to proof theory. Course notes. Available from http://kurt.scitec.kobe-u.ac.jp/∼arai/index.html, 2005.

[16] T. Arai. Progress in the proof theory related to Hilbert's second problem. Submitted, September 2005.

[17] J. Barwise. *Admissible Sets and Structures. An Approach to Definability Theory.* Omega-series. Springer, Berlin, Heidelberg, New York, 1975.

[18] J. Barwise. On Moschovakis closure ordinals. *Journal of Symbolic Logic*, 42(2):292 – 296, June 1977.

[19] W. Buchholz. A new system of proof-theoretic ordinal functions. *Annals of Pure and Applied Logic*, 32:195 – 207, 1986.

[20] W. Buchholz. Wellordering proofs for systems of fundamental sequences. Draft, München, 1990.

[21] W. Buchholz. A simplified version of local predicativity. In P. Aczel, H. Simmons, and S. S. Wainer, editors, *Proof Theory. A selection of papers from the Leeds Proof Theory Programme 1990*, pages 115 – 147. Cambridge University Press, 1992.

[22] W. Buchholz. Explaining Gentzen's proof with infinitary proof theory. In G. Gottlob, A. Leitsch, and D. Mundici, editors, *Computational logic and proof theory. 5th Kurt Gödel Colloquium, KGC '97*, pages 4 – 17. Springer Lecture Notes in Computer Science, 1289, 1997.

[23] W. Buchholz, S. Feferman, W. Pohlers, and W. Sieg. *Iterated Inductive Definitions. Recent Prooftheoretical Studies*, volume 897 of *Springer Lecture Notes in Computer Science*. Springer, 1981.

[24] W. Buchholz and K. Schütte. *Proof Theory of Impredicative Subsystems of Analysis*. Bibliopolis, Naples, 1988.

[25] P. Dybjer. Inductive families. *Formal Aspects of Computing*, 6:440–465, 1994.

[26] P. Dybjer. A general formulation of simultaneous inductive-recursive definitions in type theory. *J.Sym.Log.*, 65(2):525 – 549, June 2000.

[27] P. Dybjer and A. Setzer. A finite axiomatization of inductive-recursive definitions. In J.-Y. Girard, editor, *Typed Lambda Calculi and Applications*, volume 1581 of *Lecture Notes in Computer Science*, pages 129–146, 1999.

[28] P. Dybjer and A. Setzer. Indexed induction-recursion. In R. Kahle, P. Schroeder-Heister, and R. Stärk, editors, *Proof Theory in Computer Science*, pages 93 – 113. LNCS 2183, 2001.

[29] P. Dybjer and A. Setzer. Induction-recursion and initial algebras. *Annals of Pure and Applied Logic*, 124:1 – 47, 2003.

[30] P. Dybjer and A. Setzer. Indexed induction-recursion. *Journal of Logic and Algebraic Programming*, 66:1 – 49, 2006.

[31] S. Feferman. A language and axioms for explicit mathematics. In J. Crossley, editor, *Algebra and Logic. Proc. 1974, Monash Univ Australia*, volume 450 of *Springer Lecture Notes in Mathematics*, pages 87 – 139, 1975.

[32] G. Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493 – 565, 1936.

[33] G. Gentzen. Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie. *Archiv für mathematische Logik und Grundlagenforschung*, 16(3 – 4):97 – 118, August 1974.

[34] J.-Y. Girard. *Proof Theory and Logical Complexity*. Bibliopolis, Napoli, 1987.

[35] K. Gödel. Über formal unentscheidbare Sätze der Principia mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173 – 198, 1931.

[36] K. Gödel. *The consistency of The Axiom of Choice and of the Generalized Continuum-Hypothesis with the Axioms of Set Theory*. Princeton University Press, 1940.

[37] E. Griffor and M. Rathjen. The strength of some Martin-Löf type theories. *Arch. math. Log.*, 33:347 – 385, 1994.

[38] J. v. Heijenoort. *From Frege to Gödel*. Harward University Press, 1967.

[39] D. Hilbert. Mathematische Probleme. *Nachrichten von der Königl. Gesellschaft der Wiss. zu Göttingen (short Göttinger Nachrichten)*, pages 253 – 297, 1900. As well in Archiv der Mathematik und Physik, 1 (3):44 - 63 and 213 – 237, 1901. English translation "Mathematical problems" in Bulletin of the American Mathematical Society 8:437 – 479, 1902.

[40] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, 95(1):161 – 190, December 1926.

[41] D. Hilbert. Die Grundlagen der Mathematik. *Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität*, 6:65 – 85, 1928. English translation "The foundations of mathematics" in [38], pp. 289 – 312.

[42] G. Jäger. *Theories for Admissible Sets: A Unifying Approach to Proof Theory.* Bibliopolis, Naples, 1986.

[43] G. Jäger. First order theories for nonmonotonic inductive definitions: Recursively inaccessible and Mahlo. *Journal of Symbolic Logic*, 66(3), September 2001.

[44] R. Kahle. Mathematical proof theory in the light of ordinal analysis. *Synthese*, 133(1 – 2):237 – 255, October 2002.

[45] P. Martin-Löf. An intuitionistic theory of types: predicative part. In H. Rose and J. Sheperdson, editors, *Logic Colloquium '73*, pages 73 – 118. North-Holland, 1975.

[46] P. Martin-Löf. Constructive mathematics and computer programming. In *Logic, Methodology and Philosophy of Science, VI, 1979*, pages 153–175. North-Holland, 1982.

[47] P. Martin-Löf. *Intuitionistic type theory.* Bibliopolis, Naples, 1984.

[48] P. Martin-Löf. An intuitionistic theory of types. In G. Sambin and J. Smith, editors, *Twenty-Five Years of Constructive Type Theory.* Oxford University Press, 1998. Reprinted version of an unpublished report from 1972.

[49] B. Nordström, K. Petersson, and J. Smith. *Programming in Martin-Löf's type theory. An Introduction.* Oxford University-Press, Oxford, 1990.

[50] B. Nordström, K. Petersson, and J. M. Smith. Martin-löf's type theory. In S. Abramsky, D. M. Gabbay, and T. S. E. Maibaum, editors, *Handbook of logic in computer science, Vol. 5*, pages 1 – 37. Oxford Univ. Press, 2000.

[51] E. Palmgren. *On Fixed Point Operators, Inductive Definitions and Universes in Martin-Löf's Type Theoies.* PhD thesis, University of Uppsala (Sweden), 1991. U.U.D.M. Report 1991:8.

[52] E. Palmgren. Type-theoretic interpretation of iterated, strictly positive inductive definitions. *Archive of Mathematical Logic*, 32:75–99, 1992.

[53] E. Palmgren. On universes in type theory. In G. Sambin and J. Smith, editors, *Twenty-Five Years of Constructive Type Theory.* Oxford University Press, 1998.

[54] C. Paulin-Mohring. Inductive definitions in the system Coq - rules and properties. In *Proceedings Typed λ-Calculus and Applications*, pages 328–245. Springer-Verlag, LNCS, March 1993.

[55] R. Platek. Foundations of recursion theory. Doctorial Dissertation and Supplement, Stanford, CA: Stanford University, 1966.

[56] W. Pohlers. *Proof Theory. An introduction*, volume 1407 of *Springer Lecture Notes in Mathematics.* Springer, 1989.

[57] W. Pohlers. Proof theory and ordinal anlysis. *Archive of Mathematical Logic*, 30:311 – 376, 1991.

[58] W. Pohlers. A short course in ordinal analysis. In P. Aczel, H. Simmons, and S. S. Wainer, editors, *Proof Theory. A selection of papers from the Leeds Proof Theory Programme 1990*, pages 27 – 78. Cambridge University Press, 1992.

[59] W. Pohlers. Pure proof theory. Aims, methods and results. *Bulletin of Symbolic Logic*, 2:159–188, 1996.

[60] W. Pohlers. Subsystems of set theory and second order number theory. In S. R. Buss, editor, *Handbook of Proof Theory*, pages 209 – 335. Elsevier, 1998.

[61] M. Rathjen. Ordinal notations based on a weakly Mahlo cardinal. *Archive of Mathematical Logic*, 29:249 – 263, 1990.

[62] M. Rathjen. Proof-theoretical analysis of KPM. *Archive for Mathematical Logic*, 30:377 – 403, 1991.

[63] M. Rathjen. Collapsing functions based on recursively large cardinals: A well-ordering proof for KPM. *Archive for Mathematical Logic*, 33:35–55, 1994.

[64] M. Rathjen. Proof theory of reflection. *Ann. Pure Appl. Logic*, 68:181 – 224, 1994.

[65] M. Rathjen. An ordinal representation system for Π^1_2-comprehension and related systems. Preprint, 1995.

[66] M. Rathjen. Recent advances in ordinal analysis: Π^1_2-CA and related systems. *Bulletin of Symbolic Logic*, 1:468 – 485, 1995.

[67] M. Rathjen. The realm of ordinal analysis. In S. Cooper and J. Truss, editors, *Sets and proofs*, pages 219 – 279, Cambridge, 1999. Cambridge University press.

[68] M. Rathjen. The strength of Martin-Löf type theory with a superuniverse. Part I. *Archive for Mathematical Logic*, 39(1):1 – 39, January 2000.

[69] M. Rathjen. The superjump in Martin-Löf type theory. In S. R. Buss, P. Hájek, and P. Pudlák, editors, *Logic Colloquium '98. Proceedings of the annual European summer meeting of the Association for Symbolic Logic, held in Prague, Czech Republic, August 9 – 15, 1998*, pages 363 – 386. Lecture Notes in Logic, no. 13, Association for Symbolic Logic, Urbana and A K Peters, Natick, Mass., 2000.

[70] M. Rathjen. The strength of Martin-Löf type theory with a super universe, part II. *Archive for Mathematical Logic*, 40, 2001.

[71] M. Rathjen. Realizing Mahlo set theory in type theory. *Archive for Mathematical Logic*, 42(1):89 – 101, January 2003.

[72] M. Rathjen. The constructive Hilbert program and the limits of Martin-Löf type theory. *Synthese*, 147:81 – 120, 2005.

[73] M. Rathjen. An ordinal analysis of parameter free Π^1_2-comprehension. *Arch. Math. Log.*, 44(3):263 – 362, April 2005.

[74] M. Rathjen. An ordinal analysis of stability. *Arch. Math. Logic*, 44(1):1 – 62, January 2005.

[75] M. Rathjen and S. Tupailo. Characterizing the interpretation of set theory in Martin-Löf type theory. *Annals of Pure and Applied Logic*, 141(3):442 – 471, September 2006.

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

[77] W. Richter and P. Aczel. Inductive definitions and reflecting properties of admissible ordinals. In J. E. Fenstad and P. G. Hinman, editors, *Generalized recursion theory*, pages 301 – 381, Amsterdam, 1973. North-Holland.

[78] B. Russell. Mathematical logic as based on the theory of types. *American Journal of Mathematics*, 30(3):222 – 262, July 1908.

[79] A. Schlüter. On provability in set theories with reflection. Preprint.

[80] K. Schütte. *Proof Theory*. Springer, 1977.

[81] A. Setzer. *Proof theoretical strength of Martin-Löf Type Theory with W-type and one universe*. PhD thesis, Universität München, available via http://www.cs.swan.ac.uk/∼csetzer, 1993.

[82] A. Setzer. A model for a type theory with Mahlo Universe. 10pp. Preprint, available via http://www.cs.swan.ac.uk/∼csetzer/articles/uppermahlo.dvi or .ps or .pdf, 1996.

[83] A. Setzer. An introduction to well-ordering proofs in Martin-Löf's type theory. In G. Sambin and J. Smith, editors, *Twenty-five years of constructive type theory*, pages 245 – 263, Oxford, 1998. Clarendon Press.

[84] A. Setzer. Well-ordering proofs for Martin-Löf's type theory with W-type and one universe. *Annals of Pure and Applied Logic*, 92:113 – 159, 1998.

[85] A. Setzer. Ordinal systems. In C. B. and J. Truss, editors, *Sets and Proofs*, pages 301 – 331, Cambridge, 1999. Cambridge University Press.

[86] A. Setzer. Extending Martin-Löf type theory by one Mahlo-universe. *Arch. Math. Log.*, 39:155 – 181, 2000.

[87] A. Setzer. Ordinal systems part 2: One inaccessible. In S. Buss, P. Hajek, and P. Pudlak, editors, *Logic Colloquium '98*, ASL Lecture Notes in Logic 13, pages 426 – 448, Massachusetts, 2000. Peters Ltd.

[88] A. Setzer. Proof theory of Martin-Löf Type Theory – An overview. *Mathematiques et Sciences Humaines*, 42 année, n^{o}165:59 – 99, 2004.

[89] A. Setzer. Universes in type theory part II: Autonomous Mahlo and Π_3-reflection. 32 pages. Submitted. Available via http://www.cs.swan.ac.uk/∼csetzer/, 2005.

[90] A. Setzer. Review of [72]. *Mathematical Reviews*, MR2182643 (2006m:03023), 2006.

[91] A. Setzer. Universes in type theory part I – Inaccessibles and Mahlo. In A. Andretta, K. Kearnes, and D. Zambella, editors, *Logic Colloquium '04*, pages 123 – 156. Association of Symbolic Logic, Lecture Notes in Logic 29, 2008.

[92] W. Sieg. Hilbert's program sixty years later. *Journal of Symbolic Logic*, 53(2):338 – 348, June 1988.

[93] W. Sieg. Hilbert's Programs: 1917 – 1922. *Bulletin of Symbolic Logic*, 5(1):1 – 44, March 1999.

[94] S. G. Simpson. *Subsystems of second-order arithmetic*. Springer, 1999.

[95] S. G. Simpson, editor. *Reverse Mathematics 2001.* ASL lecture notes in logic 21, Association for Symbolic Logic and A. K. Peters, 2005.

[96] T. Skolem. Begründung der elementaren Arithmetik durch die rekurrierende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich. *Videnskapsselskapets skripfter, I. Matematisk-naturvidenskabelig klasse*, 6, 1923. See as well [38], pages 302 – 333.

[97] W. Tait. Normal derivability in classical logic. In J. Barwise, editor, *The syntax and semantics of infinitary languages*, pages 204 – 236. Springer Lecture Notes in Mathematics 72, 1968.

[98] G. Takeuti. *Proof Theory.* North–Holland Publishing Company, Amsterdam, second edition, 1987.

[99] A. Troelstra and D. v. Dalen. *Constructivism in Mathematics. An Introduction, Vol. I.* North-Holland, 1988.

[100] A. Troelstra and D. v. Dalen. *Constructivism in Mathematics. An Introduction, Vol. II.* North-Holland, 1988.

[101] A. Troelstra and H. Schwichtenberg. *Baisc Proof Theory.* Cambridge University Press, 1996.