An Intuitionistic Theory of Transfinite Types

(Draft 14 July,1990)

by

E.R. Griffor and Erik Palmgren
Department of Mathematics, Uppsala University
Uppsala, SWEDEN

Abstract

A sequence of successsively stronger universes or reflection principles U_0, U_1, ... was introduced for intuitionistic type theory by Martin-Löf, however the indexing was external to the theory. We introduce rules for a *universe operator* which, given A **Set** and a family of sets B(x) **Set**(x∈A) over A, produces a collection of sets which includes each B(x) for x in A, the set A itself and is closed under the operations Π, Σ, + and I (as well as W if desired). Thus this collection together with the sets it contains comprise a family of sets, which in turn can be taken as an input for this universe operator.

In this talk we describe the rules for the first of what can be called a *higher-order* universe or reflection principle V for intuitionistic type theory which, in addition to closure under Π, Σ, +, I (and W), is itself closed under the above universe operator. The strategy is completely general and an abstract theory of higher-order reflection principles follows entirely similar lines.

Among the rules for higher order universes we formulate *reflection rules* which can be seen as expressing the commutation of certain formal diagrams usually associated with an inductive system. Our use of the term *reflection* differs somewhat from its use by Martin-Löf in that set equalities, and not only set existence, via set constructors are reflected. However, as an example of a general *elimination rule* for these universes we give an elimination rule for our universe operator and remark that this rule, for any universe beyong the universe U_0, together with the reflection rules yields a theory where all sets are provably equal. In particular, N_0 and N_1 are provably equal and, hence, the theory is *inconsistent*. Thus, in general, reflection of all judgments of the forms C **Set** and C=C' is incompatible with the principle of transfinite induction embodied in the elimination rule for higher universes.

As an application of higher-order universes, we argue, via a naive set-theoretic interpretation of the elementary set construction principles, that they

provide a means of understanding *Mahlo numbers*. In particular that V, in this sense, naturally corresponds to $\pi_{2,1}$.

Introduction

In this article we outline an extension of Martin-Löf's intuitionistic theory of types (see Martin-Löf [1984]) to *transfinite universes*, socalled since, as the finite universes provide for internal iteration of the elementary set construction principles, transfinite universes provide a mechanism for internal transfinite iteration of universe operators(a similar operator was used by Palmgren [1989] in the context of partial type theory to give a construction of Type:Type). Our original interest was aroused by comments by Per Martin-Löf to the effect that, under a standard set-theoretic interpretation of type theory's elementary set constructors, the *finite universes* for type theory should correspond to the first ω *inaccessibles*. This together with seminal papers by Mahlo[1911] on cardinals in set theory exhibiting successively stronger closure properties, brought us to formulate an intuitionistic theory of transfinite universes which could give constructive meaning to *Mahlo numbers* .

We begin by giving a complete set of rules for Martin-Löf's finite universes in order to motivate our construction of a *higher-order* universe in the sections that follow. We define a sequence of successive universes $(U_k)_{k\in \mathbb{N}}$ and decoding functions $(T_k)_{k\in\mathbb{N}}$ by an *external* induction on k.

U_k-formation

$$U_k\ \mathbf{Set} \qquad \frac{a\in U_k}{T_k(a)\ \mathbf{Set}} \qquad \frac{a=b\in U_k}{T_k(a)=T_k(b)\ \mathbf{Set}}$$

U_k-introduction

$$\frac{a\in U_k \qquad \begin{matrix}(x\in T_k(a))\\ b\in U_k\end{matrix}}{\delta(a,(x)b)\in U_k} \qquad \frac{a\in U_k \qquad \begin{matrix}(x\in T_k(a))\\ b\in U_k\end{matrix}}{T_k(\delta(a,(x)b) = (\Delta x\in T_k(a))T_k(b)},$$

for δ each of π, σ and w and Δ each of the respective set construction principles Π, Σ and W.

$$\frac{a \in U_k \quad b \in U_k}{a+b \in U_k} \qquad \frac{a \in U_k \quad b \in U_k}{T_k(a+b) = T_k(a) + T_k(b)}$$

$$\frac{a \in U_k \quad b,c \in T_k(a)}{i(a,b,c) \in U_k} \qquad \frac{a \in U_k \quad b,c \in T_k(a)}{T_k(i(a,b,c)) = I(T_k(a),b,c)}$$

Remark: Each of the above forms of *canonical elements* in U_{k+1} has its corresponding *substitution rule* stating that equal inputs to a code for a type constructor give rise to equal codes for constructions, e.g., in the case of π we would have

$$\frac{\begin{array}{c}(x \in T_k(a)) \\ a=a' \in U_k \quad b=b' \in U_k\end{array}}{\pi(a,(x)b)=\pi(a',(x)b') \in U_k}.$$

Unless otherwise stated these rules are assumed in the sequel, although we refrain from giving them explicitely.

Ground Types (for k=0)

$n \in U_0 \quad T_0(n) = N$; $n_s \in U_0 \quad T_0(n_s) = N_s$, for s=0,1.

Remark: Our choice of ground types differs from the usual one in that we include only the first two finite sets N_0 and N_1. However, a standard argument shows that the remaining finite sets N_s (for s>1) together with their rules are derivable from these two with the help of +.

Embedding of U_k

$$u_k \in U_{k+1} \qquad T_{k+1}(u_k) = U_k$$

$$\frac{a \in U_k}{t_k(a) \in U_{k+1}} \qquad \frac{a \in U_k}{T_{k+1}(t_k(a)) = T_k(a)}.$$

We will refer to the mapping $(x)t_k(x)$ as the *inclusion* between U_k and U_{k+1}. While these last rules reflect judgments of the form C **Set** from U_k to U_{k+1}, the substitution rule

$$\frac{a = a' \in U_k}{t_k(a) = t_k(a') \in U_{k+1}}$$

ensures that the inclusion into U_{k+1} is well-defined and that judgments of the form C = C' **Set** *in* U_k are reflected in U_{k+1}. To reflect judgments C = C' in **Set,** which arise as a result of the interpretation of π, σ, +, i and w in **Set** as ∏, Σ, +, I and W (for example, by the rules for T_k, we have $T_k(a+b)=T_k(a)+T_k(b)$ in **Set**), we include the *reflection rules*

$$\frac{a \in U_k \quad b \in U_k}{t_k(a+b) = t_k(a) + t_k(b) \in U_{k+1}}$$

$$\frac{a \in U_k \qquad \begin{matrix}(x \in T_k(a))\\ b \in U_k\end{matrix}}{t_k(\delta(a,(x)b)) = \delta(t_k(a),(x)t_k(b)) \in U_{k+1}}$$

(for δ each of π, σ and w);

$$\frac{a \in U_k \quad b,c \in T_k(a)}{t_k(i(a,b,c)) = i(t_k(a),b,c) \in U_{k+1}}.$$

The Universe U_ω

To serve as motivation for the higher-order universe to be presented we exhibit some aspects of the pattern of definition by giving the rules for the *first* transfinite universe, which we denote U_ω.

U_ω-formation

$$U_\omega \ \textbf{Set} \qquad \frac{a \in U_\omega}{T_\omega(a)\ \textbf{Set}} \qquad \frac{a = b \in U_\omega}{T_\omega(a) = T_\omega(b)\ \textbf{Set}}$$

U_ω-introduction

$$\frac{c \in U_k}{t_{k,\omega}(c) \in U_\omega} \qquad \frac{c \in U_k}{T_\omega(t_{k,\omega}(c)) = T_k(c)}, \text{ for } k=0,1,\ldots$$

$$\frac{\begin{matrix} & (x\in T_\omega(a)) \\ a\in U_\omega & b\in U_\omega\end{matrix}}{\delta(a,(x)b)\in U_\omega} \qquad \frac{\begin{matrix} & (x\in T_\omega(a)) \\ a\in U_\omega & b\in U_\omega\end{matrix}}{T_\omega(\delta(a,(x)b))=(\Delta x\in T_\omega(a))T_\omega(b)},$$

(for δ each of π, σ and w and Δ the corresponding set constructor Π, Σ or W);

$$\frac{a\in U_\omega \quad b\in U_\omega}{a+b\in U_\omega} \qquad \frac{a\in U_\omega \quad b\in U_\omega}{T_\omega(a+b) = T_\omega(a) + T_\omega(b)}$$

$$\frac{a\in U_\omega \quad b,c\in T_\omega(a)}{i(a,b,c)\in U_\omega} \qquad \frac{a\in U_\omega \quad b,c\in T_\omega(a)}{T_\omega(i(a,b,c)) = I(T_\omega(a),b,c)}.$$

U_ω-reflection

$$\frac{a,b\in U_\omega}{t_{k,\omega}(a+b)=t_{k,\omega}(a)+t_{k,\omega}(b)\in U_\omega}$$

$$\frac{\begin{matrix} & (x\in T_k(a)) \\ a\in U_k & b\in U_k\end{matrix}}{t_{k,\omega}(\delta(a,(x)b))=\delta(t_{k,\omega}(a),(x)t_{k,\omega}(b))\in U_\omega}$$

(for δ each of π, σ and w),

$$\frac{a\in U_k \quad b,c\in T_k(a)}{t_{k,\omega}(i(a,b,c))=i(t_{k,\omega}(a),b,c)\in U_\omega}$$

reflecting judgments of the form C=C' to U_ω. Finally, to reflect to U_ω judgments C=C' which arise as a result of the U_ω-introduction rule giving T_ω on elements inserted in U_ω via $t_{k,\omega}$ from U_k, we have

$$\frac{c\in U_k}{t_{k+1,\omega}(t_k(c)) = t_{k,\omega}(c)\in U_\omega}.$$

Formulation à la Russell

The formulation of rules for universes we have chosen here is refered to as the formulation à la Tarski due to the similarity between the family $T(x)(x\in U)$ and Tarski's truth definition. We now give examples of an alternate formulation, refered to as the formulation à la Russell, where Π, Σ,... are viewed both as set forming operations and as operations for forming canonical elements of the universe in question, i.e., the universe is viewed truly as a set of sets.

U_α-formation (for $\alpha=0,1, \ldots ,\omega$)

$$U_\alpha \textbf{ Set} \qquad \frac{A\in U_\alpha}{A \textbf{ Set}} \qquad \frac{A=B\in U_\alpha}{A=B \textbf{ Set}}$$

U_α-introduction (for $\alpha=0,1,\ldots,\omega$)

$$\frac{\begin{matrix} & (x\in A) \\ A\in U_\alpha & B\in U_\alpha \end{matrix}}{(\Delta x\in A)B\in U_\alpha}, \text{ for } \Delta \text{ each of } \Pi,\Sigma \text{ and W;}$$

$$\frac{A\in U_\alpha \quad B\in U_\alpha}{A+B\in U_\alpha} \qquad \frac{A\in U_\alpha \quad b,c\in A}{I(A,b,c)\in U_\alpha}$$

Ground Types(k=0)

$$N\in U_0 \qquad N_s\in U_0, \text{ for } s=0,1.$$

Embedding of U_k

$$U_k\in U_{k+1} \qquad \frac{C\in U_k}{C\in U_{k+1}}. \qquad \frac{C\in U_k}{C\in U_\omega}$$

Remarks: Again we assume the standard substitution rules without explicitely stating them.

Despite the relative simplicity of the formulation à la Russell, we have not chosen to use it for two reasons. First, the formulation à la Tarski ensures a monomorphic theory, i.e., one where terms carry complete typing information (presuming a monomorphic formulation of type theory itself; see Nordström, Petersson and Smith[1990]). Secondly it leaves open the choice whether to have or not to have the reflection rules (which, as we shall see, are incompatible with the natural elimination rule even for the universe U_1).

The Operator U(A,(x)B)

The operator U together with T(·) is given as a *collection of rules parameterized by a family of sets* which may or may not be included as a construction principle under which a universe is closed. These rules should be compared with those giving Martin-Löf's *finite* universes.

U-formation

$$\frac{A\ \mathbf{Set} \quad \overset{(x\in A)}{B\ \mathbf{Set}}}{U(A,(x)B)\ \mathbf{Set}}$$

together with

$$\frac{A=A' \quad \overset{(x\in A)}{B=B'}}{U(A,(x)B) = U(A',(x)B')}$$

and

$$\frac{c\in U(A,(x)B)}{T(A,(x)B,c)\ \mathbf{Set}} \qquad \frac{c = c'\in U(A,(x)B)}{T(A,(x)B,c) = T(A,(x)B,c')}.$$

U-introduction (Π,Σ and W-closure)

$$\frac{a\in U(A,(x)B) \quad \overset{(x\in T(A,(x)B,a))}{b\in U(A,(x)B)}}{\delta(a,(x)b)\in U(A,(x)B)}$$

with

$$\frac{a \in U(A,(x)B) \qquad \begin{array}{c}(x \in T(A,(x)B,a))\\ b \in U(A,(x)B)\end{array}}{T(A,(x)B,\delta(a,(x)b) = (\Delta x \in T(A,(x)B,a))\, T(A,(x)B,b)}$$

U-introduction (+-closure)

$$\frac{a \in U(A,(x)B) \quad b \in U(A,(x)B)}{a + b \in U(A,(x)B)}$$

with

$$\frac{a \in U(A,(x)B) \quad b \in U(A,(x)B)}{T(A,(x)B,a+b) = T(A,(x)B,a) + T(A,(x)B,b)}$$

U-introduction (I-closure)

$$\frac{a \in U(A,(x)B) \qquad b,c \in T(A,(x)B,a)}{i(a,b,c) \in U(A,(x)B)}$$

with

$$\frac{a \in U(A,(x)B) \qquad b,c \in T(A,(x)B,a)}{T(A,(x)B,i(a,b,c)) = I(T(A,(x)B,a),b,c)}$$

Rules of Reflection

We introduce rules to ensure that judgements of the forms C **set** and C = C' **set,** which *hold in the universe* (A,(x)B), i.e., in case A is U(A',(x)B') and (x)B is (x)T(A',(x)B',x) for some family of sets (A',(x)B'), are *reflected* in the universe (U(A,(x)B),(x)T(A,(x)B,x)):

$$\frac{c \in A}{l(A,(x)B,c) \in U(A,(x)B)};$$

and

$$\frac{c \in A}{T(A,(x)B,l(A,(x)B,c)) = B(c/x)}.$$

Aside from the above interpretation of these three rules in terms of the *reflection* of judgements, they (together with the suppressed substitution rules) can be viewed as stating that $(x)l(A,(x)B,x)$ is a well-defined function from A to U(A,(x)B), i.e., the *inclusion map* from A into U(A,(x)B). Finally, if (A,(x)B) is the result of applying the universe operator, we add rules to guarantee that equalities in **Set** are reflected in U(A,(x)B):(for the sake of readability we introduce the abbrevations

U' for U(A',(y)B') and (x)T' for (x)T(A',(y)B',x)

until we explicitely state otherwise)

$$\frac{A'\ \mathbf{Set} \quad \begin{matrix}(x \in A')\\ B'\ \mathbf{Set}\end{matrix} \quad a,b \in U'}{l(U',(x)T',a+b) = l(U',(x)T',a) + l(U',(x)T',b) \in U(U',(x)T')};$$

$$\frac{A'\ \mathbf{Set} \quad \begin{matrix}(y \in A')\\ B'\ \mathbf{Set}\end{matrix} \quad a \in U' \quad \begin{matrix}(z \in T'(a))\\ b \in U'\end{matrix}}{l(U',(x)T',\delta(a,(z)b)) = \delta(l(U',(x)T',a),(z)l(U',(x)T',b)) \in U(U',(x)T')},$$

for δ each of π, σ and w; and

$$\frac{A'\ \mathbf{Set} \quad \begin{matrix}(y \in A')\\ B'\ \mathbf{Set}\end{matrix} \quad a \in U' \quad b,c \in T'(a)}{l(U',(x)T',i(a,c,d)) = i(l(U',(x)T',a),c,d) \in U(U',(x)T')}.$$

Remark: The right hand side of the equality in this last rule is meaningful since, by the third of these reflection rules, we have that:

$$T(U',(x)T',l(U',(x)T',a)) = T'(a/x).$$

Iteration Rules

Thus far the rules for U(A,(x)B) do not introduce sets not already *in A*. The following are designed to insure that the iteration is *true* or *proper*, i.e., they introduce a name for A: (they can be viewed as U- introduction rules)

$$*(A,(x)B) \in U(A,(x)B) \qquad \text{and} \qquad T(A,(x)B,*(A,(x)B)) = A.$$

The logical framework, due to Martin-Löf (see, for example, Palmgren and Stoltenberg-Hansen [1990]), is a typed λ–calculus with dependent types. This section is not essential to those that follow, but is included to put to rest any worries the reader may have about the typing of those canonical constants new in the universe operator. Briefly, the logical framework has the judgment forms:

$$\Gamma \text{ context}$$
$$\Gamma \Rightarrow \alpha \text{ type}$$
$$\Gamma \Rightarrow \alpha = \beta$$
$$\Gamma \Rightarrow a \in \alpha$$

and

$$\Gamma \Rightarrow a = b \in \alpha;$$

where lower-case Roman letters are used for elements of a type and lower-case Greek letter for types. It has rules for contexts

$$\frac{\Gamma \Rightarrow \alpha \text{ type}}{\Gamma, x\in\alpha \text{ context}} \qquad \text{and} \qquad \frac{x_1\in\alpha_1, \ldots, x_n\in\alpha_n \text{ context}}{x_1\in\alpha_1, \ldots, x_n\in\alpha_n \Rightarrow x_i\in\alpha_i}$$

for i=1,...,n; as well as rules for the constant **Set**:

$$\Gamma \Rightarrow \textbf{Set} \text{ type} \qquad \text{and} \qquad \frac{\Gamma \Rightarrow x\in\textbf{Set}}{\Gamma \Rightarrow \hat{x} \text{ type}}$$

(with a rule stating that equal sets go to equal types under $\hat{\cdot}$). We have formation and application of families of types and their elements given by the following rules:

$$\frac{\Gamma \Rightarrow \alpha \text{ type} \qquad \Gamma, x\in\alpha \Rightarrow \beta \text{ type}}{\Gamma \Rightarrow (x\in\alpha)\beta \text{ type}}$$

$$\frac{\Gamma, x\in\alpha \Rightarrow b\in\beta}{\Gamma \Rightarrow (x)b\in(x\in\alpha)\beta}$$

and

$$\frac{\Gamma, x\in\alpha \Rightarrow b\in\beta \qquad \Gamma \Rightarrow a\in\alpha}{\Gamma \Rightarrow ((x)b)a = b(a/x)\in\beta(a/x)} \; ,$$

where $(x\in\alpha)\beta$ is abbreviated by $(\alpha)\beta$, if β does not depend on x.

We shall not plague the reader with derivations in the logical framework, but simply give the typings and equations for the constants involved in our formulation of the universe operator $U(A,(x)B)$ and $T(A,(x)B,\cdot)$, first U and T themselves

$$U\in(A\in\mathbf{Set})((\widehat{A})\mathbf{Set})\mathbf{Set}$$

and

$$T\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(\widehat{U(A,B)})\mathbf{Set}.$$

The typings for the operations on the universe are:

$$\pi,\sigma,w\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(a\in\widehat{U(A,B)})((\widehat{T(A,B,a)})\widehat{U(A,B)})\widehat{U(A,B)}$$

$$i\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(a\in\widehat{U(A,B)})(\widehat{T(A,B,a)})(\widehat{T(A,B,a)})\widehat{U(A,B)}$$

$$+\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(\widehat{U(A,B)})(\widehat{U(A,B)})\widehat{U(A,B)}.$$

Then for δ each of π, σ and w, respectively, Δ each of Π, Σ and W, we have the following equalities

$$\begin{gathered}(A)(B)(a)(b)T(A,B,\delta(A,B,a,b)) = (A)(B)(a)(b)\Delta(T(A,B,a),(x)T(A,B,b(x)))\\ \in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(a\in\widehat{U(A,B)})((\widehat{T(A,B,a)})\widehat{U(A,B)})\mathbf{Set}\end{gathered}$$

Analogous equalities are added for T's *commutation* with + and i. For the constants $*$ and l, we have

$$*\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})\widehat{U(A,B)}$$

with

$$(A)(B)T(A,B,*(A,B)) = (A)(B)A\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})\mathbf{Set};$$

and

$$l\in(A\in\mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(\widehat{A})\widehat{U(A,B)}$$

with

$$(A)(B)(a)T(A,B,l(A,B,a))=(A)(B)(a)B(a)\in(A\in \mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(\widehat{A})\mathbf{Set}.$$

The reflection rules for the universe operator, which state that the mapping $l(A,(x)B,\cdot)$ commutes with the elementary set constructors, are of particular interest. We give only the case of +:

$$\begin{aligned}&(A)(B)(a)(b)l(U(A,B),T(A,B),+(A,B,a,b))\\ &=(A)(B)(a)(b)+(U(A,B),T(A,B),l(U(A,B),T(A,B),a),l(U(A,B),T(A,B),b))\\ &\in(A\in \mathbf{Set})(B\in(\widehat{A})\mathbf{Set})(\widehat{U(A,B)})(\widehat{U(A,B)})U(\overline{U(A,B),T(A,B)}).\end{aligned}$$

Remark: Since a universe is at the same time the *type* of each set it contains, one can see here concretely how the formulation of universes à la Tarski preserves typing information, i.e., the inclusion maps l carries with it the information about *where* the set was constructed. Implicit *logical information* is, as we've seen above, made explicit through the theory's presentation in the logical framework.

Elimination Rules for Universes

Sets in intuitionistic type theory are generally given by rules giving their canonical elements, i.e., elements as they are given by the definition of the set and not as they may be obtained as the value of some function, for example, as 1+1 *gives the numeral* 2(introduction rules); rules specifying that canonical constants are well-defined operations with respect to judgmental equality(substitution rules); and rules giving a schema for defining a function on a set(elimination and equality rules).

In the case that the set A is an *inductively defined set*, i.e., one or more of its introduction rules itself has an hypothesis of the form $t\in A$, the elimination rule can be viewed as a schema for proof by induction or definition by recursion with a case corresponding to each of A's introduction rules. Two examples are the sets N and $(Wx\in A)B(x)$, the former an example of an elementary inductive definition and the latter of a generalized inductive definition.

The universes we consider here are clearly inductively defined sets, although we have refrained until now from giving elimination and equality rules for them. We begin by giving these rules for type theory with a single universe. Our notation for the assumptions of a rule differs for typographical reasons in that *sets of assumptions* are given horizontally with the help of $\Rightarrow$ denoting derivability. It should be noted that subsequent application of the elimination rule entails the *discharge* of those hypotheses occuring to the left of $\Rightarrow$. We denote this universe by U and (x)T and suppress the assumption that one has given a family of sets $C(z)(z\in U)$:

U-elimination

Let Γ denote the following sequence of sets of assumptions:

$$x\in U,\ y\in T(x)\rightarrow U,\ u\in C[x],\ v\in (\Pi w\in T(x))C[App(y,w)] \Rightarrow \\ d_{\delta}(x,y,u,v)\in C[\delta(x,(w)App(y,w))],$$

for δ each of π, σ and w;

$$x,y\in U,\ u\in C[x],\ v\in C[y] \Rightarrow d_{+}(x,y,u,v)\in C[x+y];$$

$$x\in U,\ y\in T(x),\ u\in T(x),\ v\in C[x] \Rightarrow d_{i}(x,y,u,v)\in C[i(x,y,u)];$$

$$d_{n}\in C[n]$$

$d_{n_i}\in C[n_i]$, for i=0,1.

Then the elimination rule is

$$\frac{s\in U \quad \Gamma}{H(s,d_{n_0},d_{n_1},d_{n}\ ,(x,y,u,v)d_{\pi},(a,y,u,v)d_{\sigma},(x,y,u,v)d_{w},(x,y,u,v)d_{+},(x,y,u,v)d_{i})\in C[s]}$$

the conclusion of which we abbreviate by $H(s;\ \overline{\tau})$.

U-equality

Under the assumptions in Γ the following judgments are immediately derivable:

$$H(n_i;\overline{\tau}) = d_{n_i}\in C[n_i],\text{for } i=0,1;$$

$$H(\delta(a,(x)b);\ \overline{\tau}) = d_\delta[a,(\lambda x)b,H(a;\ \overline{\tau}),(\lambda x)H(b;\ \overline{\tau})] \in C[\delta(a,(x)b)],$$ for δ each of π,σ and w;

$$H(i(a,b,c);\ \overline{\tau}) = d_i[a,b,c,H(a;\ \overline{\tau})] \in C[i(a,b,c)];$$

$$H(a+b;\ \overline{\tau}) = d_+[a,b,H(a;\ \overline{\tau}),H(b;\ \overline{\tau})] \in C[a+b].$$

The elimination rule for the universe operator $U(A,(x)B)$ and $(y)T(A,(x)B,y)$ will differ from that for a single universe in that:

(1) the assumptions for the sets N_0, N_1 and N are replaced by

$$d_* \in C[*(A,(x)B)]$$

and

$$x \in A \Rightarrow d_l(x) \in C[l(A,(x)B,x)];$$

while (2) the function H will also depend on the family $B(x)(x \in A)$.

Using Γ' to denote this new sequence of sets of assumptions and $\overline{\tau}'$ to denote the corresponding sequence of arguments to H, we have

U(A,(x)B)-elimination

$$\frac{s \in U(A,(x)B) \qquad \Gamma'}{H(A,(x)B,s;\overline{\tau}') = C[s]}$$

U(A,(x)B)-equality

With the above mentioned notational changes in H, we have the equalities for π, σ, w, and i as well as:

$$H(A,(x)B,*(A,(x)B);\ \overline{\tau}') = d_* \in C[*(A,(x)B)]$$

and

$$H(A,(x)B,l(A,(x)B,a);\ \overline{\tau}') = d_l[a] \in C[l(A,(x)B,a)].$$

Universe Elimination and Reflection Rules

The elimination rule for U_0 and $(x)T_0$ is that we gave for a single universe and that for U_1 and $(x)T_1$ is easily seen from that for the universe operator with the notational changes: $*_0$ for $*(A,(x)B)$, d_{t_0} for d_l and $(x)t_0(x)$ for $(y)l(A,(x)B,y)$. By the reflection rules we have for $a,b \in U_0$ that

$$t_0(a+b) = t_0(a) + t_0(b) \in U_1.$$

However the elimination rule for U_1 and $(x)T_1$ allows us to define a function from U_1 into, say, U_0 such that d_+ is constant n_0 and d_{t_0} is constant n_1. Hence, by the substitution, elimination and corresponding equality rules, we derive the judgment $n_0 = n_1 \in U_0$ and consequently that $N_0 = N_1$. Thus, in general, the reflection rules together with elimination rules for higher universes (beyond that for U_0 and $(x)T_0$) yield a *logically inconsistent* theory where all sets are equal. The embedding of one universe into the *next* together with the reflection rules give rise to a non-deterministic (or, in the terminology of Kleene[1952], non-fundamental) inductive definition, hence the difficulty in defining a function on that set by the corresponding principle of recursion.

A *Second-order* Universe

Because we have parameterized the *process* of constructing a universe à la Tarski closed under Π, Σ, + and I over a given family of ground sets (which may or may not itself be a universe), we refer to U(A,(x)B) and (z)T(A,(x)B,z) as a *universe operator*. Notice that if A **Set** and $x \in A \Rightarrow B$ **Set**, i.e., (A,(x)B) is a family of sets, then U(A,(x)B) **Set** and $x \in U(A,(x)B) \Rightarrow T(A,(x)B,x)$ **Set**, i.e., (U(A,(x)B),(x)T(A,(x)B,x)) is a family of sets. Thus the *value* of this operator on a family of sets is also a family of sets and we have set the stage for an *internal iteration* of the process of giving the *next universe*. We shall now give a reflection principle or universe V which is closed under the operator U. Distinguishing the set construction principles Π, Σ, +, I and W from reflection principles, it is reasonable and, as we shall see, quite precise to call V a *second-order* reflection principle.

V-formation

$$V\ \textbf{Set} \qquad \frac{a \in V}{S(a)\ \textbf{Set}} \qquad \frac{a = a' \in V}{S(a) = S(a')\ \textbf{Set}}.$$

V-introduction

$n \in V \qquad S(n) = N$

$n_k \in V \qquad S(n_k) = N_k$, for $k = 0,1,\ldots$

$$\frac{a\in V \qquad \overset{(x\in S(a))}{b\in V}}{\delta(a,(x)b)\in V} \qquad \frac{a\in V \qquad \overset{(x\in S(a))}{b\in V}}{S(\delta(a,(x)b)) = (\Delta x\in S(a))S(b)},$$

for δ each of π, σ or w and Δ the corresponding set constructor $\prod$, Σ or W;

$$\frac{a\in V \quad b\in V}{a + b\in V} \qquad \frac{a\in V \quad b\in V}{S(a + b) = S(a) + S(b)}$$

$$\frac{a\in V \quad b\in S(a) \quad c\in S(a)}{i(a,b,c)\in V} \qquad \frac{a\in V \quad b\in S(a) \quad c\in S(a)}{S(i(a,b,c)) = I(S(a),b,c)}.$$

In addition to the above introduction rules for V we now add those giving the operator U(A,(x)B). As before we continue to suppress the substitution rules.

V-introduction (U-closure)

$$\frac{a\in V \qquad \overset{(x\in S(a))}{b\in V}}{u(a,(x)b)\in V} \qquad \frac{a\in V \qquad \overset{(x\in S(a))}{b\in V}}{S(u(a,(x)b)) = U(S(a),(x)S(b))}$$

$$(*) \quad \frac{a\in V \qquad \overset{(x\in S(a))}{b\in V} \qquad c\in U(S(a),(x)S(b))}{t(a,(x)b,c)\in V}$$

and

$$(**) \quad \frac{a\in V \qquad \overset{(x\in S(a))}{b\in V} \qquad c\in U(S(a),(x)S(b))}{S(t(a,(x)b,c)) = T(S(a),(x)S(b),c)}.$$

(*) gives the insertion into V of codes from some *local* U and (**) that the inserted codes are interpreted as the same sets as they were *locally*. Thus these two rules guarantee that judgements of the form C **Set** in U(A,(x)B) for (A,(x)B) a family of sets *over V* are reflected in V. Finally, to ensure that judgements of the form C = C' **Set** are reflected in the same manner:

$$\frac{a \in V \qquad \begin{matrix}(x \in S(a))\\ b \in V\end{matrix} \qquad c,d \in U(S(a),(x)S(b))}{t(a,(x)b,c + d) = t(a,(x)b,c) + t(a,(x)b,d) \in V};$$

for δ each of π, σ and w;

$$\frac{a \in V \qquad \begin{matrix}(x \in S(a))\\ b \in V\end{matrix} \qquad \begin{matrix}c \in U(S(a),(x)S(b))\\ d,e \in T(S(a),(x)S(b),c)\end{matrix}}{t(a,(x)b,i(c,d,e)) = i(t(a,(x)b,c),d,e) \in V}.$$

Once again the right hand side of the equality in the last rule is meaningful since, by a previous rule, we have that

$$S(t(a,(x)b,c)) = T(S(a),(x)S(b),c),$$

while

$$\begin{aligned}S(t(a,(x)b,i(c,d,e))) &= T(S(a),(x)S(b),i(c,d,e))\\ &= I(T(S(a),(x)S(b),c),d,e).\end{aligned}$$

Finally, we add

$$\frac{a \in V \qquad \begin{matrix}(x \in S(a))\\ b \in V\end{matrix} \qquad c \in U(S(a),(x)S(b))}{t(u(a,(x)b),(y)t(a,(x)b,y),l(S(u(a,(x)b)),(y)S(t(a,(x)b,y)),c) = t(a,(x)b,c) \in V}$$

and

$$\frac{a \in V \qquad \begin{matrix}(x \in S(a))\\ b \in V\end{matrix}}{t(a,(x)b,*(S(a),(x)S(b))) = a \in V}.$$

The Universe Operator and V à la Russell

U-formation

$$\frac{\begin{matrix} & (y\in A) \\ A\ \mathbf{Set} & B(y)\ \mathbf{Set}\end{matrix}}{U(A,(y)B)\ \mathbf{Set}} \qquad\qquad \frac{C\in U(A,(y)B)}{C\ \mathbf{Set}}$$

U-introduction

$$A\in U(A,(y)B) \qquad\qquad \frac{a\in A}{B(a/y)\in U(A,(y)B)}$$

$$\frac{\begin{matrix} & (x\in C) \\ C\in U(A,(y)B) & D\in U(A,(y)B)\end{matrix}}{(\Delta x\in C)D\in U(A,(y)B)} \text{ for } \Delta \text{ each of } \Pi, \Sigma \text{ and } W$$

$$\frac{C\in U(A,(y)B) \quad D\in U(A,(y)B)}{C+D\in U(A,(y)B)}$$

$$\frac{C\in U(A,(y)B) \quad a\in C \quad b\in C}{I(C,a,b)\in U(A,(y)B)}.$$

V-formation

$$V\ \mathbf{Set} \qquad\qquad \frac{A\in V}{A\ \mathbf{Set}}$$

V-introduction

$$N\in V \qquad\qquad N_k\in V, \text{ for } k=0,1,\ldots$$

$$\frac{\begin{matrix} & (x\in A) \\ A\in V & B\in V\end{matrix}}{(\Delta x\in A)B\in V} \text{ for } \Delta \text{ each of } \Pi,\ \Sigma \text{ and } W$$

$$\frac{\begin{matrix} & (x\in A) \\ A\in V & B(x)\in V\end{matrix}}{U(A,(x)B)\in V} \quad \text{and} \quad \frac{A\in V \qquad \begin{matrix}(x\in A)\\ B(x)\in V\end{matrix} \qquad C\in U(A,(x)B)}{C\in V}.$$

Two Applications of V

In this section we use the universe V to give Martin-Löf's external sequence of finite universes as an *internal* sequence and then to interpret the universe U_ω.

The Sequence $(U_k)_{k\in N}$

In the introduction we defined the sequence of finite universes by a recursion on natural numbers *external* to type theory. We shall now construct this sequence as an *internal* sequence with the help of V. Since V **Set** we have that $(\Sigma x\in V)(S(x)\to V)$ **Set**, which we will denote by **PAR.** For $z\in N$ let C be the constant family of sets given by $C(z)\overset{def}{=}$**PAR**. Thus $(\Pi z\in N)C(z)$ **Set** and gives a plausible typing of our sequence of universes, since each consists of a set (of codes) together with a function taking these codes to sets.

Consider first the case of obtaining a pair of codes for U_0 and T_0. We have that $u(n_3,(x)R_3(x,n,n_0,n_1))\in V$ and for $z\in S(u(n_3,(x)R_3(x,n,n_0,n_1)))$ we have that $t(n_3,(x)R_3(x,n,n_0,n_1),z)\in V$ and hence that $(\lambda z)t(n_3,(x)R_3(x,n,n_0,n_1),z)\in (S(u(n_3,(x)R_3(x,n,n_0,n_1)))\to V)$. Thus, we have that

$$(u(n_3,(x)R_3(x,n,n_0,n_1)),(\lambda z)t(n_3,(x)R_3(x,n,n_0,n_1),z))\in C(0),$$

which we denote by $(u_0,(\lambda z)t_0)$.

In general, using N-elimination, one shows analogously that the desired sequence is

$$(\lambda k)R(k,(u_0,(\lambda z)t_0),(k,c)(u(p(c),(z)Ap(q(c),z)),(\lambda w)t(u(p(c),(z)Ap(q(c),z),w))\in (\Pi k\in N)C(k),$$

which we denote by $(\lambda k)d(k)$ in the application of V that follows. In particular we take: $U_k \overset{\text{def}}{=} S(p(d(k)))$ and $T_k(a) \overset{\text{def}}{=} S(Ap(q(d(k)),a))$. To verify that these give an interpretation of the finite universes in V, one must interpret the rules of U_k as rules in V.

Interpretation of the ω-hierarchy of universes in V

We define an interpretation $(\cdot)^*$ of terms and types from the universes U_α, for $\alpha=0,1,\ldots,\omega$, as terms and types in V. Let U_k^* be $S(p(d(k)))$ and $T_k^*(x)$ be $S(Ap(q(d(k)),x))$ so that $T_k(a)^* = T_k^*(a^*)$. We have that $U_0^* = U(N_3,(x)S(R_3(x,n_0,n_1,n)))$ and $T_0^*(a) = T(N_3,(x)S(R_3(x,n_0,n_1,n)),a)$. A simple computation shows that

$$\begin{aligned} U_{k+1}^* &= S(p(d(k+1))) = S(u(p(d(k)),(z)Ap(q(d(k)),z)) = \\ &\quad U(S(p(d(k)),(z)S(Ap(q(d(k)),z)) = U(U_k^*,(z)T_k^*(z)) \end{aligned}$$

and

$$\begin{aligned} T_{k+1}^*(a) &= S(Ap(q(d(k+1)),a)) = S(t(p(d(k)),(z)Ap(q(d(k)),z),a)) = \\ &\quad T(S(p(d(k))),(z)S(Ap(q(d(k)),z)),a) = T(U_k^*,(z)T_k^*(z),a). \end{aligned}$$

Now let u_k^* be $*(U_k^*,(x)T_k^*(x))$ and notice that $u_k^* \in U_{k+1}^*$. Then we have that

$$\begin{aligned} T_{k+1}(u_k)^* &= T_{k+1}^*(u_k^*) = T(U_k^*,(z)T_k^*(z),*(U_k^*,(x)T_k^*(x)) \\ &= U_k^*. \end{aligned}$$

Furthermore, if we let n_i^* be $l(N_3,(x)S(R_3(x,n_0,n_1,n),i_3)$, for i=0 and 1, and let n^* be $l(N_3,(x)S(R_3(x,n_0,n_1,n),2_3)$, then $n_i^*,n^* \in U_0^*$ and we have that

$$\begin{aligned} T_0(n_i)^* &= T_0^*(n_i^*) \\ &= T(N_3,(x)S(R_3(x,n_0,n_1,n),l(N_3,(x)S(R_3(x,n_0,n_1,n),i_3) \\ &= S(R_3(i_3,n_0,n_1,n)) = S(n_i) = N_i. \end{aligned}$$

If we let $t_k^*(a)$ be $l(U_k^*,(z)T_k^*(z),a)$, then we have that

$$\begin{aligned} T_{k+1}(t_k(a))^* &= T_{k+1}^*(t_k^*(a^*)) = T(U_k^*,(z)T_k^*(z),l(U_k^*,(z)T_k^*(z),a^*)) \\ &= T_k^*(a^*) = T_k(a)^*. \end{aligned}$$

For δ each of π, σ and w, we let $\delta(a,(x)b)^*$ be $\delta(a^*,(x)b^*)$ as well as $(a+b)^*$ be a^*+b^* and $i(a,b,c)^*$ be $i(a^*,b^*,c^*)$, then we have for example that

$$\begin{aligned} t_k(a+b)^* &= t_k^*(a^*+b^*) = l(U_k^*,(z)T_k^*(z),a^*+b^*) \\ &= l(U_k^*,(z)T_k^*(z),a^*)+l(U_k^*,(z)T_k^*(z),b^*) \\ &= t_k^*(a^*)+t_k^*(b^*) = (t_k(a)+t_k(b))^*. \end{aligned}$$

Now let $t_{k,\omega}^*(c)$ be $Ap(q(d(k)),c)$ and U_ω^* be V, then we have that $t_{k+1,\omega}(u_k)^* = t_{k+1,\omega}^*(u_k^*)$ and one shows by induction on k that if $c \in U_k^*$, then $t_{k,\omega}^*(c) \in U_\omega^*$. Let $T_\omega^*(c)$ be S(c), then obviously we have that

$$T_\omega(t_{k,\omega}(c))^* = S(Ap(q(d(k)),c^*) = T_k^*(c^*) = T_k(c)^*.$$

Finally, the interaction between inclusions is verified by the computation:

$$t_{k+1,\omega}(t_k(c))^* = Ap(q(d(k+1)),l(U_k^*,(z)T_k^*(z),c)) =$$

$$t(p(d(k)),(z)Ap(q(d(k)),z),l(S(p(d(k))),(z)S(Ap(q(d(k)),z),c^*) =$$

$$t(p(d(k)),(z)Ap(q(d(k)),z),c^*) = t_{k,\omega}(c)^*.$$

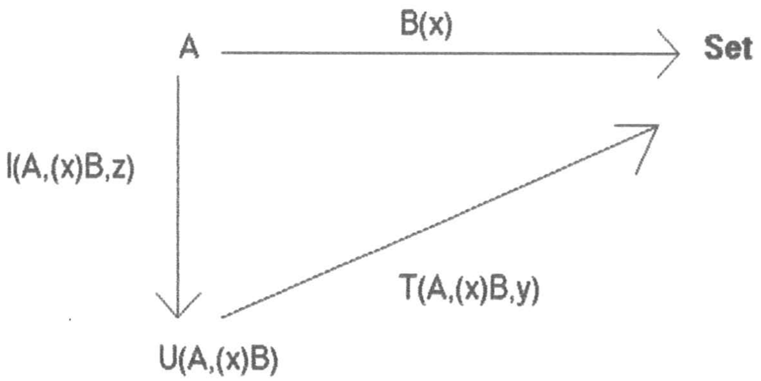
Reflection Rules and Commutative Diagrams

What we have called reflection rules are of distinct sorts, those expressing the fact that mappings defined in the formulation of a universe à la Tarski are *inclusions* between the collections of sets defined, those expressing the fact that these mappings are *homomorphisms* in the sense that they commute with elementary set constructors and those stating that our global mappings cohere with the local mappings between universes.

As an example of a rule of the first sort, consider the rule for the universe operator:

$$\frac{c \in A}{T(A,(x)B,l(A,(x)B,c)) = B(c/x)}.$$

This rule can be graphically interpreted as the commutation of the following diagram:

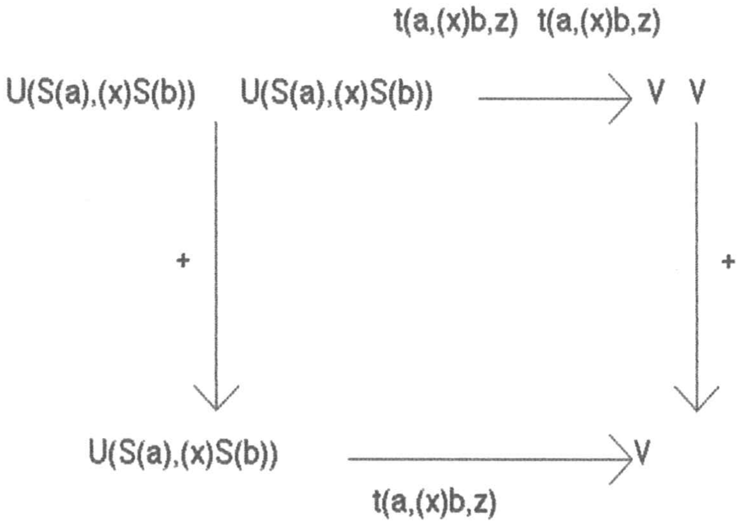


.

As an example of a rule of the second sort, consider the following rule of V stating that global mappings commute with sums:

$$\frac{a \in V \qquad \begin{matrix}(x \in S(a)) \\ b \in V\end{matrix} \qquad c,d \in U(S(a),(x)S(b))}{t(a,(x)b,c+d) = t(a,(x)b,c) + t(a,(x)b,d) \in V};$$

and the *formal diagram* it can be seen as stating is commutative:

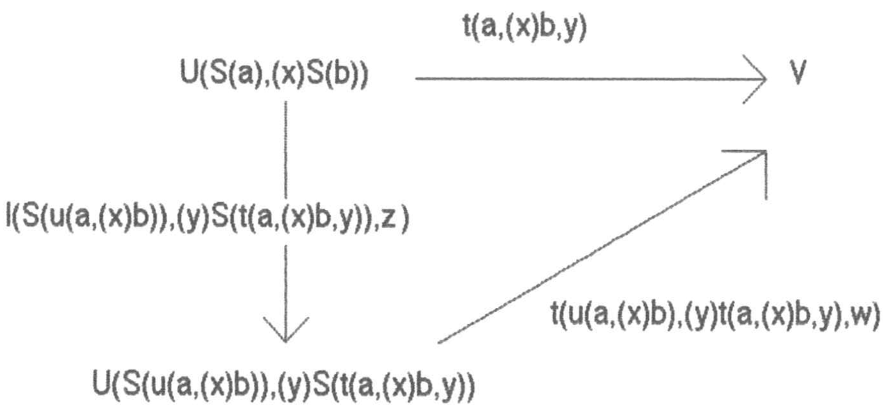


.

As an example of a reflection rule of the third sort, consider the following rule for V stating that global mappings commute with local mappings:

$$\frac{a \in V \qquad \begin{matrix}(x \in S(a))\\ b \in V\end{matrix} \qquad c \in U(S(a),(x)S(b))}{t(u(a,(x)b),(y)t(a,(x)b,y),l(S(u(a,(x)b)),(y)S(t(a,(x)b,y)),c) = t(a,(x)b,c) \in V}.$$

This also can be viewed as asserting that the formal diagram



is commutative.

These diagrams are purely formal, however the similarity with the commutative diagrams and homomorphisms of an inductive system in a category-theoretic setting is striking. They do nonetheless provide a pattern for giving the rules for such *limit universes*. In the sequel to this paper we shall see that there is more than a formal analogy.

Cardinality and Mahlo Numbers

A Hierarchy of Universes

Just as with the universe operator, the universe V can itself be given as a *collection of rules parameterized by a family of sets*, i.e., a *second-order* operator: V(A,(x)B) together with S(A,(x)B,·). Then, as with V, one can give a universe V' which, aside from closure under elementary set constructors, is closed under V(A,(x)B) with S(A,(x)B,·). Repetition of this sort of diagonalization gives rise to a *hierarchy of universes*. Suppose we let $U_{k,0}$, for k=0,1, ... , enumerate externally the finite universes. A similar sequence of universes can easily be given $U_{0,1}$, $U_{1,1}$, ... starting with V and proceeding as with $U_{k,0}$. In general, assume that we have defined a universe operator of *order n* and let $U_{k,n}$, for

$k=0,1, \ldots$, enumerate successive universes closed under that operator. In this way we obtain a doubly indexed hierarchy of universes $U_{k,l}$, for $k,l=0,1, \ldots$.

Remark: The theory of universe operators is actually far more subtle and will be studied in a sequel to this article, for example, carrying out these constructions *internally*, formulating correctly the corresponding reflection rules and indexing the resulting hierarchy into the transfinite.

Mahlo Numbers

In Mahlo [1911], a hierarchy of successively more strongly closed cardinals in set theory is defined: $\pi_{\mu,\nu}$, for ordinals μ and ν. The sequence $\pi_{0,\nu}, \pi_{1,\nu}, \ldots, \pi_{\gamma,\nu}, \ldots$, for all ordinals γ are called *π_ν-numbers.* The π_0-numbers are just the regular cardinals and, in general, α is a *π_ν-number*, if for all $\sigma<\nu$ α is a π_σ-number, $\pi_{\gamma,\sigma}$, such that $\pi_{\gamma,\sigma}= \gamma$, i.e., a fixed point in the enumeration of π_σ-numbers (for all ν, $\pi_{0,\nu} = 0$). For example, $\pi_{\gamma,1}$ enumerates the *weakly inaccessible* cardinals (a regular cardinal $\kappa > \omega$ such that $\aleph(\tau) < \kappa$, for all $\tau < \kappa$; where $\aleph(x)$ is the Hartog function).

Cardinality in Type Theory

One way of investigating the *size* of a set-theoretic universe is to study those cardinals which it contains. In ZFC there are two ways of constructing a cardinal larger that a given cardinal κ: exponientiation 2^κ and the Hartog number κ^+(the least μ such that there is no surjection of κ onto μ).

In order to draw a parallel between Mahlo's hierarchy of cardinals and the above hierarchy of universes, we intend to study the closure of a universe for type theory under constructions of new *cardinals.* To do so we require a means of comparing the *size* of sets. Throughout this section we abbreviate I(A,a,b) by $a =_A b$, which denotes intentional equality ($a = b \in A$ is reserved for *definitional equality*).

<u>Definition</u> (i) $f \in A \to B$ is *injective*, if

$$(\forall a,b\in A)[f(a) =_B f(b) \supset a =_A b];$$

(ii) $f \in A \to B$ is *surjective*, if

$$(\forall b\in B)(\exists a\in A)[f(a) =_B b].$$

If $f \in A \to B$ is surjective, then we can produce a *right inverse* for f, i.e., a $g \in B \to A$ such that $(\forall b \in B)[f(g(b)) =_B b]$, since the axiom of choice holds. Note that if f is both surjective and injective, then g is an inverse: if $a \in A$, then $f(g(f(a))) =_B f(a)$ and hence, by injectivity, $g(f(a)) =_A a$.

We shall use the notation $A \geq B$ for the claim that there exists a surjection of A onto B. Now there is no reason to believe that we can construct a surjection in $A \to B$ given an injection in $B \to A$ or that the Schroeder-Bernstein theorem should hold in type theory. In contrast to ZF we cannot show from the axiom of choice that all sets can be "well-ordered", e.g., $N \to N$ (a proof here would be of considerable interest). We shall describe a *diagonalization principle* for sets and try to construct well-ordered families of cardinals.

The well-ordering type $(Wx \in A)B(x)$ provides a diagonalization principle in type theory. First a lemma stating that a *tree* in $(Wx \in A)B(x)$ cannot be equal to any of its immediate subtrees.

Lemma Let W denote $(Wx \in A)B(x)$, then

$$(\forall c \in W)(\forall x \in B(p(c)))[\neg\, c =_W Ap(q(c),x)],$$

(where p(c) denotes T(c,(x,y,z)x) and q(c) denotes T(c,(x,y,z)y), i.e., the two projection functions on W).

proof: We show the proposition C(z), which is defined to be $(\forall x \in B(p(z)))\neg z =_W Ap(q(z),x)$ by induction on z. Suppose that $c \in W$, $a \in A$ and that $b \in B(a) \to W$, as well at that $d \in (\prod v \in B(a))C(Ap(b,v))$. Assume further that $v \in B(p(sup(a,b)))$, $u \in sup(a,b) =_W Ap(q(sup(a,b)),v)$. But $p(sup(a,b)) = a$, so

$$d(v) \in (\forall x \in B(p(p(Ap(b,v))))\neg\, Ap(b,v) =_W Ap(q(Ap(b,v)),x).$$

The construction u gives

$$J(u,(x')d(v)) \in (\forall x \in B(p(Ap(b,v)))\neg\, Ap(b,v) =_W Ap(b,v),$$

which is a contradiction. Thus $\neg\, sup(a,b) =_W Ap(q(sup(a,b),v)$ and, hence, C(sup(a,b)) true. By W-elimination, we have C(c) true. □

Per Martin-Löf has pointed out that $(Wx\in A)B(x)$ is a set *larger* than all $B(x)(x\in A)$ in the following sense: if we try to define a surjection $f\in B(a)\rightarrow(Wx\in A)B(x)$, then there is always an element not in the range of f, namely sup(a,f).

<u>Theorem</u> $(\forall a\in A)\neg B(a)\geq(Wx\in A)B(x)$

<u>proof</u>: If $f\in B(a)\rightarrow W$ is a surjection with the corresponding right inverse $g\in W\rightarrow B(a)$, then let b be $sup(a,f)\in W$. Then $g(b)\in B(a)$ and hence $f(g(b)) =_W sup(a,f)$, which contradicts the lemma. □

Given a set A, define

$$A^+ \stackrel{def}{=} (Wx\in N_2)R_2(x,N_0,A).$$

Then by the above theorem, we have that $\neg(A\geq A^+)$. More generally, $(\cdot)^+$ is monotone with respect to $\leq$.

<u>Theorem</u> If $A\leq B$, then $A^+\leq B^+$.

<u>proof</u>: omitted.

Thus, for every set A in a universe U, we have that A^+ is also an element of U. The following theorem states that a universe is larger than any of its sets. First, a technical lemma whose proof we omit. We note, however, that the proof of the lemma makes use of U-elimination.

<u>Lemma</u> $(\forall a\in U)(\forall f\in T(a)\rightarrow U)(\forall x\in T(a))\ \neg\ \sigma(a,(x)Ap(f,x)) =_U f(x)$.

<u>Theorem</u> $(\forall a\in A)\neg T(a)\geq U$.

<u>proof</u>: Suppose that $f\in T(a)\rightarrow U$ is a surjection with right inverse $g\in U\rightarrow T(a)$. Let b denote $\sigma(a,(x)Ap(f,x))$, then $f(g(b)) =_U b$, contradicting the lemma.

□

We mentioned in the introduction the naive set-theoretic interpretation of all of the elementary set constructors with the exception of the well-ordering type. It can also be interpreted as a set in the classical sense naively by given by a generalized inductive definition (see Salvesen [] for details). Hence, viewing universes as *limit cardinals*, we have a rough interpretation of the Mahlo number of finite level, $\pi_{n+1,m+1}$, as the universe $U_{n,m}$. In particular, the universe V "corresponds" to the Mahlo number $\pi_{1,2}$. It should be possible to formulate the formal theory of Mahlo numbers and interpret it in intuitionistic type theory with universes along these lines.

Concluding Remarks

Very little is known about the proof-theoretic strength of these higher-order universes. By Aczel [1977]:

$$|ML_1| = |ID_1^-|,$$

where ML_1 denotes type theory with one universe and without the wellordering type and $|ID_1^-|$ denotes the theory of a single inductive definition where the axioms for the fixpoint do not state that it is the least such. This result was later extended by Feferman [1982] to $ML_{<\omega}$:

$$|ML_{<\omega}| = |ID_{<\omega}^-| = \Gamma_0,$$

where $ML_{<\omega}$ is type theory with all the finite universes but still without the well-ordering type and $ID_{<\omega}^-$ is the theory of iterated inductive definitions with the *weak* fixpoint axioms mentioned above. The second author has recently interpreted an intuitionistic theory of iterated, strictly positive inductive definitions, s.p.-$ID_{<\omega}^i$, in type theory with the well-ordering type and all the finite universes, $MLW_{<\omega}$. On the other hand, W. Sieg ([], p.184) has shown that the theory of finite constructive tree classes, $ID_{<\omega}^i(O)$, is included in s.p.-$ID_{<\omega}^i$, i.e.,

$$|ID_{<\omega}^i(O)| = |ID_{<\omega}|.$$

Hence we have that: $|MLW_{<\omega}| \geq |ID_{<\omega}|$. In point of fact, no interesting upper bound even for MLW_1 is known (other than the stated *analogy* with the Mahlo number $\pi_{1,1}$), but we conjecture that

$$|MLW_1| \geq |ID_{<\varepsilon_0}|,$$

which by $|ID_{<\varepsilon_0}| = |\Sigma^1_2\text{-AC}|$ and the result that $|\Sigma^1_2\text{-AC}| = |\Delta^1_2\text{-CA}|$, would give

$$|MLW_1| \geq |\Delta^1_2\text{-CA}|,$$

by viewing iterations of constructive tree classes as iterations of the well-ordering type.

These considerations would seem to suggest that intuitionistic type theory with higher-order universes itself can provide a more uniform setting for the study of the proof-theoretic strength of a large variety of theories.

References

Acxel, P. [1977], The strength of Martin-Löf's intuitionistic type theory with one universe, in: S. Miettinen, S. Väänänen (eds.) *Proceedings of the symposium on mathematical logic (Oulu 1974)*, Report No. 2, Dept. of Philosophy, University of Helsinki.

Feferman, S. [1982], Iterated inductive fixed-point theories: applications to Hancock's conjecture, in: G. Metakides (ed.), *Patras logic symposium '80*, North-Holland Publishing Co., Amsterdam.

Kleene, S.C. [1952], *Introduction to Metamathematics*, North-Holland Publishing Co., Amsterdam.

Mahlo, P. [1911], Über lineare transfinite Mengen, Leipziger Berichte, Math.-Phys. Klasse 63, pp. 187-225.

Martin-Löf, P. [1984], *Intuitionistic Type Theory*, Bibliopolis, Napoli.

Nordström, Petersson and Smith [1990], *Programming in Martin-Löf's Type Theory: An Introduction*, Oxford University Press.

Palmgren, E. [1989], A construction of Type:Type in Martin-Löf's partial type theory, manuscript.

Palmgren, E. [1990], A type-theoretic interpretation of iterated, strictly positive inductive definitions, manuscript.

Palmgren, E. and V. Stoltenberg-Hansen [1990], Martin-Löf's Partial Type Theory, Uppsala Univ. Preprint 1990:3.