ATTI DEL CONGRESSO

LOGICA E FILOSOFIA DELLA SCIENZA, OGGI

organizzato dalla Società Italiana di Logica e Filosofia delle Scienze
(SILFS)

San Gimignano, 7-11 dicembre 1983

VOLUME I - LOGICA

a cura di

V. MICHELE ABRUSCI e ETTORE CASARI

Editrice CLUEB Bologna

INDICE

Relazioni

Comunicazioni

Atti del Congresso *Logica e Filosofia della Scienza, oggi.*
San Gimignano, 7-11 dicembre 1983. Vol. 1 - Logica.
CLUEB, Bologna (Italy) © 1986.

A NORMAL FORM FOR NATURAL DEDUCTIONS IN A TYPE THEORY WITH REALIZING TERMS

Helmut Schwichtenberg

Mathematisches Institut
Universität München
Theresienstraße 39, D-8000 München 2 – West Germany

We shall describe a formal system T which is well suited to formalize mathematical proofs and, in particular, to make explicit the algorithmic content of such proofs.

Formulas are allowed to contain, in addition to the classical logical operators, the constructive existential quantifier denoted by $\exists^*$; $\vee^*$ will be defined by

$$A \vee^* B :\equiv \exists^* x[(sg(x) = 0 \to A) \wedge (sg(x) = 1 \to B)] .$$

Looking at formulas like $(\exists^* xA \to \exists^* yB) \to \exists^* zC$ makes it clear that our system T will have to deal with functionals of finite types. Since we want to allow definitions by (primitive) recursion and proofs by induction, we have to work with the primitive recursive functionals from [Gödel 1958].

In T we shall derive not only formulas but formulas A together with terms $r_1, \ldots, r_n =: \underset{\sim}{r}$ realizing them in the sense of [Kreisel 1959]. Such realizing terms denote primitive recursive functionals. For " $\underset{\sim}{r}$ realizes A " we simply write " $\underset{\sim}{r} \in A$ " ; such expressions are called judgements. For formulas A without $\exists^*$ (so-called negative formulas) the sequence $\underset{\sim}{r}$ of realizing terms is empty and hence our judgements coincide with the usual formulas.

The logical rules of T are those of natural deduction in the sense of [Gentzen 1934]. We only use the rules for so-called minimal logic. $0=1$ can play the rôle of falsity; hence $\neg A$ can be defined by $A \to 0=1$. The presence of realizing terms makes it possible that the introduction and elimination rules for $\exists^*$ have the following simple formulation:

$$(\exists^{*+})\ \frac{\underset{\sim}{r} \in A[s]}{s,\underset{\sim}{r} \in \exists^* xA} \qquad (\exists^{*-})\ \frac{s,\underset{\sim}{r} \in \exists^* xA}{\underset{\sim}{r} \in A[s]} .$$

Note that, due to the presence of realizing terms, T is stronger than HA^{ω} : In T we can derive trivially any instance of the axiom of choice in the form

$$\forall x \exists^* y A[x,y] \quad \to \quad \exists^* f \forall x A[x,fx] ,$$

and also the so-called independence-of-premiss (IP) schema

$$(A \to \exists^* xB) \quad \to \quad \exists^* x(A \to B)$$

for negative A .

Functionals of T are intended to be extensional. Formally this will be expressed by the fact that we have equality rules

$$(\to^{=+})\ \frac{r=s}{\lambda xr = \lambda xs} \qquad (\to^{=-})\ \frac{t_1 = t_2 \qquad s_1 = s_2}{t_1 s_1 = t_2 s_2}$$

and that a term $\lambda x(rx)$ is identified with r .

It is easy to see that T (w.r.t. its negative fragment) is conservative over classical arithmetic Z , since within Z one can define the model HEO (cf. [Troelstra 1973]) of T . In this way many metamathematical results known for Z can be transferred to T . However, since it is our aim to make explicit the algorithmic content of proofs in a way as simple as possible, a direct proof-theoretic analysis of T seems to be appropriate.

The central point of such an enterprise is certainly Gentzen's Hauptsatz, saying that any derivation in T can be transformed into a normal form. The main property of such a normal form is that it does not make any "detours". We shall prove such a normal form theorem for T , by an adaption of a method of [Tait 1965].

Note that the existence of derivations in T without detours doesn't say much if proofs by induction are made possible in T by admitting induction axioms of the form

$$A[0] \wedge \forall x(A[x] \rightarrow A[x+1]) \rightarrow \forall y A[y] \ .$$

Then one doesn't have any control over the complexity of the induction formula A . Hence we shall formulate T with an ω-rule instead of induction axioms. Then derivations will be infinite objects, and by proper restrictions one has to take care that the resulting system is equivalent to the original one.

As an application of the normal form theorem we shall prove that from any derivation of a formula $\exists y(fxy=0)$ with the classical existential quantifier (and x,y of type 0) one can find a term $r[x]$ and a derivation of $fxr[x]=0$ and hence of $r[x] \in \exists^* y(fxy=0)$; this is (the appropriate formulation of) the Markov rule for T .

The idea of using formulas with realizing terms has been introduced into the literature independently by Martin-Löf and Minc (see [Martin-Löf 1984], [Minc 1975]). Minc used it as a tool in his proof of the conservativity of HA^{ω} + AC over HA , whereas Martin-Löf aims at a rather general theory of meaning.- From Martin-Löf I have taken the terminology $\underset{\sim}{r} \in A$ and the name "judgement" for such expressions.

Concerning the relation of T to intuitionistic mathematics it seems to me that T deals properly with one aspect of the intuitionistic approach, namely to make explicit the algorithmic content of proofs. However, other aspects, in particular the fact that a proof gives an insight into the validity of e.g. a Π^0_1-formula $\forall x(fx=0)$, are not made explicit.

§ 1 The formal system T

The intended model of T consists of the primitive recursive functionals in the sense of [Gödel 1958], i.e. of all functionals which can be defined explicitely from recursion operators in all finite types. In order to understand properly the computation procedure given by the definition of such functionals it is advisable to use infinite terms instead of the recursion operators. Of course the formation of infinite terms has to be restricted properly in order to make sure that they define just the primitive recursive functionals. Here we follow this course and define our language accordingly.

Finite types are 0, the type of natural numbers, and with ρ,σ also $\rho\to\sigma$, the type of functionals mapping functionals of type ρ into those of type σ. We shall write $\rho_1,\dots,\rho_n\to\sigma$ for $\rho_1\to(\rho_2\to\dots(\rho_n\to\sigma)\dots)$. The level $L\rho$ of a type ρ is defined inductively by $L0=0$, $L(\rho\to\sigma)=\max(L\rho+1, L\sigma)$. Terms of type ρ are defined inductively, as follows.

(A) $z_0^\rho, z_1^\rho, z_2^\rho, \dots$ are terms of type ρ. They are called variables of type ρ and usually denoted by $x,y,\dots$.

(0^+) For any natural number j the numeral $\underline{j}$ is a term of type 0.

(0^-) If t is a term of type 0 and $s_0, s_1, s_2, \dots$ are terms of type ρ then $Et\langle s_i\rangle_{i<\omega}$ (E for evaluation) is a term of type ρ. This rule of term formation is called critical since $Et\langle s_i\rangle$ has the same type as its "minor parts" s_i .

($\to^+$) If x is a variable of type ρ and r is a term of type σ, then λxr is a term of type $\rho\to\sigma$.

($\to^-$) If t is a term of type $\rho\to\sigma$ and s is a term of type ρ, then ts is a term of type σ.

We shall only consider terms with only finitely many free variables. Then substitution $r_x[s]$ can be defined easily. <u>Formulas</u> of T are built from equations $r = s$ between terms of the same type ρ by means of $\wedge$, $\rightarrow$, $\forall$ and also $\exists^*$. $A \vee^* B$ is an abbreviation of $\exists^*x[(sg(x) = \underline{0} \rightarrow A) \wedge (sg(x) = \underline{1} \rightarrow B)]$; here sg is the term $\lambda x Ex\langle \underline{sg(i)} \rangle_{i<\omega}$. The classical operations $\vee$, $\exists$ are taken as defined, i.e. $A \vee B :\equiv \neg(\neg A \wedge \neg B)$, $\exists xA :\equiv \neg\forall x\neg A$; the negation $\neg A$ stands for $A \rightarrow \underline{0} = \underline{1}$. Note that variables of arbitrary types can be quantified.

As already mentioned, we want to derive in T <u>judgements</u> of the form $\underset{\sim}{r} \in A$, not just formulas A. Here $\underset{\sim}{r} \equiv r_1, \ldots, r_n$ is a finite sequence of terms of types $\rho_1, \ldots, \rho_n$; n and the types $\rho_1, \ldots, \rho_n =: \underset{\sim}{\rho}$ are determined by A, as follows. To an equation $r = s$ there belongs the empty sequence of types. If $\underset{\sim}{\rho}$ are the types of A and $\underset{\sim}{\sigma} \equiv \sigma_1, \ldots, \sigma_m$ are the types of B, then $\underset{\sim}{\rho}, \underset{\sim}{\sigma}$ are the types of $A \wedge B$, $\underset{\sim}{\rho} \rightarrow \sigma_1, \ldots, \underset{\sim}{\rho} \rightarrow \sigma_m$ are the types of $A \rightarrow B$, $\rho \rightarrow \sigma_1, \ldots, \rho \rightarrow \sigma_m$ are the types of $\forall x^\rho B$ and $\rho, \underset{\sim}{\sigma}$ are the types of $\exists^* x^\rho B$.

<u>Derivations</u> in T are defined inductively, by the following rules. It is helpful to think of them as tree-like figures made up from judgements. At the top nodes there are assumptions of the form $\underset{\sim}{u} \in A$ with different variables $\underset{\sim}{u} \equiv u_1, \ldots, u_n$, or axioms. An assumption can be bound at a later inference. We identify the equation $r = s$ with $s = r$. To form derivations we admit the following <u>rules of inference</u>. Again some of those rules are called critical, namely if their "minor parts" are of the same types as the conclusion.

(A) $\underset{\sim}{u} \in A$

(R) $x = x$

(T) $\dfrac{r = s \quad s = t}{r = t}$ critical

(O^-) $$\frac{\ldots\ \utilde{r} \in A\,(t = \underline{i})\ \ldots \text{ for all } i < \omega}{\utilde{r} \in A} \qquad \text{critical}$$

$(O^{=+})$ $$\underline{j} = \underline{j}$$

$(O^{=-})$ $$\frac{t_1 = t_2 \qquad \ldots\ s_{1i} = s_{2i}\ \ldots \text{ for all } i < \omega}{Et_1 \langle s_{1i} \rangle_{i<\omega} = Et_2 \langle s_{2i} \rangle_{i<\omega}} \qquad \text{critical}$$

$(\to^{=+})$ $$\frac{r = s}{\lambda xr = \lambda xs}$$

$(\to^{=-})$ $$\frac{t_1 = t_2 \qquad s_1 = s_2}{t_1 s_1 = t_2 s_2}$$

$(\wedge^+)$ $$\frac{\utilde{r} \in A \qquad \utilde{s} \in B}{\utilde{r}, \utilde{s} \in A \wedge B}$$

$(\wedge^-)$ $$\frac{\utilde{r}, \utilde{s} \in A \wedge B}{\utilde{r} \in A} \qquad \frac{\utilde{r}, \utilde{s} \in A \wedge B}{\utilde{s} \in B}$$

$(\to^+)$ $$\frac{\utilde{r} \in B \qquad (\utilde{u} \in A)}{\lambda \utilde{u}\utilde{r} \in A \to B}$$

$(\to^-)$ $$\frac{\utilde{t} \in A \to B \qquad \utilde{s} \in A}{\utilde{t}\utilde{s} \in B}$$

$(\forall^+)$ $$\frac{\utilde{r} \in A}{\lambda x \utilde{r} \in \forall x A}$$

$(\forall^-)$ $$\frac{\utilde{t} \in \forall x A}{\utilde{t} s \in A[s]}$$

$(\exists^{*+})$ $$\frac{\utilde{r} \in A[s]}{s, \utilde{r} \in \exists^* x A}$$

$$(\exists^{*-}) \qquad \frac{s, \underset{\sim}{r} \in \exists^{*} x A}{\underset{\sim}{r} \in A[s]}$$

In applications of $\to^{=+}$, $\to^{+}$ and $\forall^{+}$ the obvious conditions on variables are to be respected. Also we require that any derivation has only finitely many variables free in the derivation, and only finitely many different free assumptions. Then substitutions of terms for variables and of derivations for assumptions are possible, since necessary changes of bound variables can be carried out.

A term $(\lambda xr)s$ obviously denotes the same functional as $r[s]$. Hence it seems to be appropriate to identify both terms. This is particularly advisable when later we want to eliminate detours in derivations and e.g. want to replace a derivation

$$\dfrac{\dfrac{\begin{array}{c}\underset{\sim}{u} \in A^{(1)} \\ | \\ \underset{\sim}{r} \in B\end{array}}{\lambda \underset{\sim}{u}\underset{\sim}{r} \in A \to B}(1) \qquad \underset{\sim}{s} \in A}{(\lambda \underset{\sim}{u}\underset{\sim}{r})\underset{\sim}{s} \in B}$$

by

$$\begin{array}{c}\underset{\sim}{s} \in A \\ | \\ \underset{\sim}{r}[\underset{\sim}{s}] \in B\end{array}$$

In order to be able to consider the second derivation as deriving the same judgement as the first one, we have to consider $(\lambda \underset{\sim}{u}\underset{\sim}{r})\underset{\sim}{s}$ and $\underset{\sim}{r}[\underset{\sim}{s}]$ as equal. For similar reasons we have to identify $E\underline{j}\langle s_i\rangle$ with s_j , $E(Et\langle s_i\rangle)\langle p_j\rangle$ with $Et\langle Es_i\langle p_j\rangle\rangle$ and $(Et\langle s_i\rangle)p$ with $Et\langle s_i p\rangle$. Since functionals are taken to be extensional we shall further identify $\lambda x(rx)$ with r .

Our formal procedure will be to show that any term can be reduced to a uniquely determined term in normal form not containing any subterm of the form $(\lambda xr)s$, $E\underline{j}\langle s_i\rangle$, $E(Et\langle s_i\rangle)\langle p_j\rangle$, $(Et\langle s_i\rangle)p$ or $\lambda x(rx)$. Then we identify two terms if they have the same normal form.

The existence of the normal form will be proved in § 3 by the method of [Tait 1965]. The proof of the uniqueness of the normal form will not be given here; it can be done by a method of [Maaß 1974] (cf. [Ruckert 1984]).

§ 2 Examples for derivations

Now we want to show that some expected derivations can be carried out in T .

Lemma 2.1 (Axiom of choice)

$$s,\underset{\sim}{r} \in \forall x \exists^* y A[x,y] \vdash s,\underset{\sim}{r} \in \exists^* f \forall x A[x,f]$$

Proof:
$$\cfrac{\cfrac{\cfrac{\cfrac{s,\underset{\sim}{r} \in \forall x \exists^* y A[x,y]}{sx,\underset{\sim}{r}x \in \exists^* y A[x,y]}}{\underset{\sim}{r}x \in A[x,sx]}}{\underset{\sim}{r} \in \forall x A[x,sx]}}{s,\underset{\sim}{r} \in \exists^* f \forall x A[x,fx]}$$

Lemma 2.2 (Independence of premiss) For negative A we have

$$s,\underset{\sim}{r} \in (A \to \exists^* yB) \vdash s,\underset{\sim}{r} \in \exists^* y(A \to B)$$

Proof:
$$\cfrac{\cfrac{\cfrac{\cfrac{s,\underset{\sim}{r} \in A \to \exists^* yB \qquad A^{(1)}}{s,\underset{\sim}{r} \in \exists^* yB}}{\underset{\sim}{r} \in B[s]}}{\underset{\sim}{r} \in A \to B[s]}\;(1)}{s,\underset{\sim}{r} \in \exists^* y(A \to B)}$$

Remark: Obviously Lemma 2.2 holds for all formulas A with an empty sequence of types (cf. § 1). These are exactly the so-called Harrop-formulas, i.e. formulas with $\exists^*$ only in premisses of implications.

Lemma 2.3 (Extensionality)

$$\forall \underset{\sim}{x}(r\underset{\sim}{x} = s\underset{\sim}{x}) \vdash tr = sr$$

Proof: We make use of the fact that $\lambda\underset{\sim}{x}(r\underset{\sim}{x})$ is identified with r.

$$\dfrac{t = t \qquad \dfrac{\dfrac{\forall \underset{\sim}{x}(r\underset{\sim}{x} = s\underset{\sim}{x})}{r\underset{\sim}{x} = s\underset{\sim}{x}}}{\underbrace{\lambda\underset{\sim}{x}(r\underset{\sim}{x})}_{r} = \underbrace{\lambda\underset{\sim}{x}(s\underset{\sim}{x})}_{s}}}{tr = sr} \qquad \square$$

The Peano axioms can be proved easily from the ω-rule 0^-. For the argument one needs

Lemma 2.4 (Equality)

a. $t_1 = t_2 \vdash r[t_1] = r[t_2]$

b. $t_1 = t_2$, $\underset{\sim}{r}[t_1] \in A[t_1] \vdash \underset{\sim}{r}[t_2] \in A[t_2]$

Proof: a can be seen easily by induction on r. b is proved by induction on A, using a.

Next we show that $\underline{0} = \underline{1}$ can play the rôle of falsity, i.e. that from the assumption $\underline{0} = \underline{1}$ one can derive an arbitrary judgement $\underset{\sim}{r} \in A$.

Lemma 2.5 $\quad \underline{0} = \underline{1} \vdash \underset{\sim}{r} \in A$

Proof: Step 1: $\underline{0} = \underline{1} \vdash \underline{i} = \underline{j}$. Obviously it suffices to define a term $+$ of type $0,0 \to 0$ such that $\underline{i} + \underline{j}$ (more precisely:

$(+\underline{i})\underline{j}$) has $\underline{i+j}$ as its normal form. Take

$$+ :\equiv \lambda xy.\mathrm{Ex}\langle \mathrm{Ey}\langle \underline{i+j}\rangle_{j<\omega}\rangle_{i<\omega}$$

Step 2: $\underline{0}=\underline{1} \vdash r=s$ for arbitrary terms r,s of type 0. This can be obtained from step 1 using the ω-rule 0^{-}:

$$\dfrac{\dfrac{\dots\quad \dfrac{\begin{array}{c}\underline{0}=\underline{1}\\ |\\ \underline{i}=\underline{j}\end{array}\quad r=\underline{i}^{(1)}}{r=\underline{j}}\quad \dots \text{ for all } i<\omega}{r=\underline{j}}(1)\qquad s=\underline{j}^{(2)}}{\dots\qquad r=s\qquad \dots \text{ for all } j<\omega}$$

$$\dfrac{\dots\qquad r=s\qquad \dots \text{ for all } j<\omega}{r=s}(2)$$

Step 3: $\underline{0}=\underline{1} \vdash r=s$ for arbitrary terms r,s of type ρ. We use induction on the type ρ. The case $\rho\equiv 0$ has been dealt with in step 2. For the general case $\rho\to\sigma$ we make use of the fact that functionals are taken in T to be extensional:

$$\dfrac{\begin{array}{c}\underline{0}=\underline{1}\\ |\\ rx=sx\end{array}}{\lambda x(rx)=\lambda x(sx)}\quad ,$$

and $\lambda x(rx)$, $\lambda x(sx)$ are identified with r,s.

Step 4: $\underline{0}=\underline{1} \vdash \underset{\sim}{r}\in A$. We use induction on the formula A. The case of an equation has been dealt with in step 3. The case $\wedge$ is trivial. In the case $\to$ we have

$$\dfrac{\begin{array}{c}\underline{0}=\underline{1}\\ |\\ \underset{\sim}{ru}\in B\end{array}}{\lambda\underset{\sim}{u}(\underset{\sim}{ru})\in A\to B}\quad ,$$

and $\lambda\underset{\sim}{u}(\underset{\sim}{ru})$ is identified with $\underset{\sim}{r}$. The case $\forall$ is treated similarly, and the case $\exists^*$ is trivial again. □

As is to be expected we have the usual rules for the defined classical operations $\vee$, $\exists$:

<u>Lemma 2.6</u> For negative formulas A,B,C we have

a. $A \vdash A \vee B$ $\qquad B \vdash A \vee B$

b. $A \vee B \ , \ A \to C \ , \ B \to C \vdash C$

c. $A[s] \vdash \exists x A$

d. $\exists x A \ , \ A \to C \vdash C$

<u>Remark</u>: This lemma can be strengthened somewhat: A,B may be arbitrary formulas, and for C it suffices that C is a Harrop formula (i.e. with $\exists^*$ only in premisses of implications); for such C we have $\neg\neg C \vdash C$.

Proof: We shall prove the lemma in the strengthened form of the remark above. It suffices to deal with c and d ; a and b are treated similarly. Note that $\exists x A$ abbreviates $\neg \forall x \neg A$.
For c we have

$$\dfrac{\dfrac{\dfrac{\forall x \neg A^{(1)}}{\neg A[s]} \qquad \underset{\sim}{r} \in A[s]}{\underline{0} = \underline{1}}}{\exists x A}\,(1)$$

For d we have to use $\neg\neg C \vdash C$; then the required derivation can be obtained easily:

$$\dfrac{\dfrac{\dfrac{\exists x A \qquad \dfrac{\dfrac{\neg C^{(1)} \qquad \begin{array}{c} A \to C \\ | \\ \neg C \to \neg A \end{array}}{\neg A}}{\forall x \neg A}}{\underline{0} = \underline{1}}}{\neg\neg C}\,(1) \qquad \neg\neg C \to C}{C}$$

So it remains to prove $\neg\neg C \vdash C$ for Harrop formulas C. We show this by induction on the formula C.

Step 1: $\neg\neg(\underline{i} = \underline{j}) \vdash \underline{i} = \underline{j}$. This is clear if i and j are equal numbers. If they are unequal, we have

$$\dfrac{\neg\neg(\underline{i} = \underline{j}) \qquad \dfrac{\begin{array}{c} \underline{i} = \underline{j}^{\,(1)} \\ | \\ \underline{0} = \underline{1} \end{array}}{\underline{i} = \underline{j} \to \underline{0} = \underline{1}}\,(1)}{\begin{array}{c} \underline{0} = \underline{1} \\ | \\ \underline{i} = \underline{j} \end{array}}$$

Step 2: $\neg\neg(r = s) \vdash r = s$ for arbitrary terms r, s of type 0. This can be obtained easily from step 1 above using the ω-rule 0^-; cf. step 2 of the proof of Lemma 2.5.

Step 3: $\neg\neg(r = s) \vdash r = s$ for arbitrary terms r, s of type ρ. We use induction on the type ρ. The case $\rho = 0$ has been dealt with in step 2. For the general case we again make use of the fact that functionals are taken in T to be extensional:

$$\begin{array}{ll} r = s \vdash rx = sx & \\ \neg\neg(r = s) \vdash \neg\neg(rx = sx) \vdash rx = sx & \text{by ind. hyp.} \\ \neg\neg(r = s) \vdash \underbrace{\lambda x(rx)}_{r} = \underbrace{\lambda x(sx)}_{s} & \end{array}$$

Step 4: $\neg\neg C \vdash C$ for Harrop formulas C. We use induction on C. The case of an equation has been dealt with in step 3. The case $\wedge$ is similar to the case $\forall$:

$$\begin{array}{ll} \forall x C \vdash C & \\ \neg\neg\forall x C \vdash \neg\neg C \vdash C & \text{by ind. hyp.} \\ \neg\neg\forall x C \vdash \forall x C & \end{array}$$

Case $\to$:

$$C\,,\ C \to D \vdash D$$
$$C\,,\ \neg\neg(C \to D) \vdash \neg\neg D \vdash D \qquad \text{by ind. hyp.}$$
$$\neg\neg(C \to D) \vdash C \to D \qquad \square$$

Finally we show that for the operation $\vee^*$ defined above by $A \vee^* B :\equiv \exists^* x[(sg(x) = \underline{0} \to A) \wedge (sg(x) = \underline{1} \to B)]$ we have the following rules:

Lemma 2.7

a. $\underset{\sim}{r} \in A \vdash \underline{0}\,,\ \underset{\sim}{r}\,,\ \underset{\sim}{s} \in A \vee^* B$

$\quad\ \underset{\sim}{s} \in B \vdash \underline{i+1}\,,\ \underset{\sim}{r}\,,\ \underset{\sim}{s} \in A \vee^* B$

b. $t\,,\ \underset{\sim}{r}\,,\ \underset{\sim}{s} \in A \vee^* B\ ,\ \underset{\sim}{p} \in A \to C\ ,\ \underset{\sim}{q} \in B \to C \vdash \underset{\sim}{d} \in C$

The proof of a is obvious. For b we have

$$\dfrac{\dfrac{\underset{\sim}{p} \in A \to C \qquad \dfrac{\dfrac{t,\underset{\sim}{r},\underset{\sim}{s} \in A \vee^* B}{\underset{\sim}{r} \in (sg(t) = 0 \to A)} \quad \dfrac{t = \underline{0}\ ^{(1)}}{sg(t) = \underline{0}}}{\underset{\sim}{r} \in A}}{\underbrace{\underset{\sim}{p}\underset{\sim}{r}}_{\underset{\sim}{d}[\underline{0}]} \in C} \quad t = \underline{0}\ ^{(1)} \;\Big/\; \underset{\sim}{d}[t] \in C \qquad \cdots\ \dfrac{\underset{\sim}{q} \in B \to C \qquad \dfrac{\dfrac{t,\underset{\sim}{r},\underset{\sim}{s} \in A \vee^* B}{s \in (sg(t) = \underline{1} \to B)} \quad \dfrac{t = \underline{i+1}\ ^{(1)}}{sg(t) = \underline{1}}}{\underset{\sim}{s} \in B}}{\underbrace{\underset{\sim}{q}\underset{\sim}{s}}_{\underset{\sim}{d}[\underline{i+1}]} \in C} \quad t = \underline{i+1}\ ^{(1)} \;\Big/\; \underset{\sim}{d}[t] \in C\ \cdots}{\underset{\sim}{d}[t] \in C}\ (1)$$

where $\underset{\sim}{d} :\equiv Ex\langle \underset{\sim}{p}\underset{\sim}{r}, \underset{\sim}{q}\underset{\sim}{s}, \underset{\sim}{q}\underset{\sim}{s}, \ldots \rangle$.

§ 3 A normal form for terms

As already mentioned we want to prove now that any term can be brought into a normal form not containing any subterm of the form $(\lambda xr)s$, $E\underline{j}\langle s_i\rangle$, $E(Et\langle s_i\rangle)\langle p_j\rangle$, $(Et\langle s_i\rangle)p$ or $\lambda x(rx)$. The easiest way to achieve that is by applying a method due to [Tait 1965]. We present the proof in such a way that it can be generalized easily to the case of derivations to be treated in § 5 .

The rules of term formation have been described in § 1 ; they are

(A)	x	
(O^+)	$\underline{j}$	
(O^-)	$Et\langle s_i\rangle_{i<\omega}$	critical
$(\to^+)$	λxr	
$(\to^-)$	ts	

Rules marked with + (-) are called <u>introduction</u> (<u>elimination</u>) <u>rules</u>. In an elimination rule we call the term denoted by t its <u>main part</u> and the terms denoted by s_i its <u>minor parts</u>. A term is called <u>convertible by evaluation</u> if it ends with an elimination rule whose main part is obtained by an introduction rule for numerals, i.e. O^+ . The relation " p is convertible by evaluation into q " is defined by

1. $E\underline{j}\langle s_i\rangle \mapsto s_j$.

A term is called <u>properly convertible</u> if it ends with an elimination rule whose main part is obtained by a non-numerical introduction rule, i.e. $\to^+$. The relation " p is properly convertible into q " is defined by

2. $(\lambda xr)s \mapsto r_x[s]$.

A term is called <u>permutatively convertible</u> if it ends with an elimination rule whose main part is obtained by a critical rule, i.e. O^- . The relation " p is permutatively convertible into q " is defined by

3. $E(Et\langle s_i\rangle)\langle p_j\rangle \quad \mapsto \quad Et\langle Es_i\langle p_j\rangle\rangle$

4. $(Et\langle s_i\rangle)p \quad \mapsto \quad Et\langle s_i p\rangle$.

Finally, a term is called η - <u>convertible</u> if it is of the form $\lambda x(rx)$. The relation " p is η - convertible into q " is defined by

5. $\lambda x(rx) \quad \mapsto \quad r$.

The relation of <u>reducibility</u> between terms is the smallest reflexive and transitive relation containing the relation of convertibility and closed against the rules of term formation. A term is said to be in <u>normal form</u> if it contains no properly, permutatively or η - convertible subterm and also no subterm convertible by evaluation.

We want to show that any term can be reduced to a term in normal form. To achieve this it suffices to produce a normal form w.r.t. proper and permutative conversions; this is carried out in Theorem 3.4 below. For after this is done one can easily eliminate all η - convertible subterms by replacing each subterm $\lambda x(rx)$ by r ; doing that no new convertible subterms will arise. Finally one can also eliminate all subterms convertible by evaluation: replace each subterm $E\underline{j}\langle s_i\rangle$ by s_j . (We shall show in § 4 that the corresponding function on code numbers for terms is the only one that can not be choosen to be primitive recursive). Note that doing this at most new η - convertible subterms will arise; they can again be eliminated as above.

We also want to note the effect of the construction of the normal form to the height of the terms involved. (Here the word "height" refers to the helpful picture of a term as given by a labelled tree). The <u>height</u> $|r|$ of a term r is defined as follows.

$$|x| = |\underline{j}| = 1$$

$$|Et\langle s_i\rangle| = \max(|t|, \sup_i |s_i|) + 1$$

$$|\lambda xr| = |r|$$

$$|ts| = \max(|t|, |s|) + 1 .$$

One can see easily that with this definition we have

<u>Lemma 3.1</u> $|r[s]| \leq |s| + |r|$.

Furthermore it is clear that η - conversions and conversions by evaluation can only lower the height of a term.

The <u>level</u> Lr of a term r is the level of the type of r . By the <u>rank</u> Rr of a term r we mean the least number greater than all levels of main parts of properly or permutatively convertible subterms of r .

<u>Lemma 3.2</u> (Reduction Lemma) Let p be a term ending with an elimination with main part t and minor parts s_i , so $p \equiv Et\langle s_i\rangle$ or $p \equiv ts_o$. If $Lt \leq k$ and $Rt, Rs_i \leq k$, then t can be reduced to a term q with $Rq \leq k$ and

$$|q| \leq \max(|p|, (\sup_i |s_i|) + |t|) .$$

Proof: By induction on t . We may assume that t ends with a non-numerical introduction (i.e. $\to^+$) or with a critical rule (i.e. O^-), for otherwise we already would have $Rp \leq k$.

Case $\to^+$: Then $t \equiv \lambda xr$, hence $p \equiv (\lambda xr)s_o$. Since $L(\lambda xr) \leq k$ we have $Ls_o < k$ and hence $Rr[s_o] \leq k$. Furthermore we know by Lemma 3.1 that $|r[s_o]| \leq |s_o| + |r|$. So we can take $q \equiv r[s_o]$.

Case O^- : Then $t \equiv Er\langle p_j\rangle$, hence $p \equiv E(Er\langle p_j\rangle)\langle s_i\rangle$ or $p \equiv (Er\langle p_j\rangle)s_o$. Each p_j is a proper subterm of t , and $Lp_j =$ $= Lt \leq k$. Hence, by induction hypothesis, $Ep_j\langle s_i\rangle$ or p_js_o can be reduced to a term q_j with $Rq_j \leq k$ and

$$|q_j| \leq \max(\max(|p_j|, \sup_i |s_i|) + 1 , (\sup_i |s_i|) + |p_j|) .$$

We can take $q \equiv Er\langle q_j\rangle$, since $REr\langle q_j\rangle \leq k$ and

$$\begin{aligned} |Er\langle q_j\rangle| &= \max(|r| , \sup_j |q_j|) + 1 \\ &\leq \max(|r| , \sup_j(|p_j| + 1) , (\sup_i |s_i|) + \sup_j |p_j|) + 1 \\ &\leq \max((\sup_j |p_j|) + 2 , (\sup_i |s_i|) + |t|) \\ &\leq \max(|p| , (\sup_i |s_i|) + |t|) \end{aligned}$$

<u>Lemma 3.3</u> (Structural Lemma) Consider a term of rank $\leq k+1$ ending with an elimination with main part t of a level $Lt \geq k+1$. Then t is obtained by in general multiple uncritical eliminations (i.e. applications of $\to^-$) from a variable.

Proof: If $t \equiv: t_o$ is obtained by an uncritical elimination, take its main part t_1 . If t_1 again is obtained by an uncritical elimination, take again its main part t_2 , and so on. Note that all t_1 have levels $\geq k+1$. The procedure terminates with a main part t_n . By construction t_n is not obtained by an uncritical elimination. Since by assumption the whole term has a rank $\leq k+1$, we also know that t_n can neither be obtained by an introduction nor by a critical rule. Hence t_n is a variable.

<u>Theorem 3.4</u> (Existence of the Normal Form)

a. Any term r of rank $\leq k+1$ can be reduced to a term r' of rank $\leq k$, such that $|r'| \leq 2^{|r|}$.

b. Any term r (of arbitrary rank $\leq \omega$) can be reduced to a term r^* of rank 0 , such that $|r^*| \leq \varepsilon_{|r|}$.

Proof of a: By induction on r . If r doesn't end with an elimination the claim follows immediately from the induction hypothesis. So assume that r ends with an elimination, i.e.

$r \equiv Et\langle s_i\rangle$ or $r \equiv ts_o$. Then, by induction hypothesis, we have t', s_i' such that Rt' , $Rs_i' \leq k$ and $|t'| \leq 2^{|t|}$, $|s_i'| \leq 2^{|s_i|}$.

Case 1. The main part t has a level $Lt \geq k+1$. Then by the Structural Lemma 3.3 we know that t is obtained by multiple uncritical eliminations from a variable, i.e. $t \equiv xr_1 \ldots . r_n$. Then we have $t' \equiv xr_1' \ldots . r_n'$ (by induction hypothesis w.r.t. a claim to be proved simultaneously, saying that

$$(xr_1 \ldots . r_n)' \equiv xr_1' \ldots . r_n' \quad) .$$

Take $r' \equiv Et'\langle s'\rangle$ or $r' \equiv t's_o'$, respectively. Then we know from the form of t' that $Rr' \leq k$. For the height we have

$$\begin{aligned} |r'| &= \max(|t'| , \sup_i |s_i'|) + 1 \\ &\leq \max(2^{|t|} , \sup_i 2^{|s_i|}) + 1 \\ &< 2^{\max(|t| , \sup_i |s_i|) + 1} \\ &= 2^{|r|} \end{aligned}$$

Case 2. The main part t has a level $Lt \leq k$. Then we can apply the Reduction Lemma 3.2 to $Et'\langle s_i'\rangle$ or $t's_o'$, respectively. We obtain the required term r' with rank $Rr' \leq k$. For the height we have

$$\begin{aligned} |r'| &\leq \max(\max(|t'| , \sup_i |s_i'|) + 1 , (\sup_i |s_i'|) + |t'|) \\ &= \max(|t'| + 1 , (\sup_i |s_i'|) + |t'|) \\ &\leq \max(2^{|t|} + 1 , (\sup_i 2^{|s_i|}) + 2^{|t|}) \\ &\leq \max(2^{|t|+1} , 2^{\max(|t| , \sup_i |s_i|) + 1}) \\ &= 2^{|r|} \end{aligned}$$

Proof of b: Again by induction on r . If r doesn't end with an elimination the claim again follows immediately from the induction hypothesis. So assume that r ends with an elimination, i.e. $r \equiv Et\langle s_i\rangle$ or $r \equiv ts_o$. Then, by induction hypothesis, we have t^*, s_i^* such that $Rt^*, Rs_i^* = 0$ and $|t^*| \leq \varepsilon_{|t|}$, $|s_i^*| \leq \varepsilon_{|s_i|}$. Let $k < \omega$ be the level of the main part t. Then we construct the required term r^* by a k-fold application of part a to $Et^*\langle s_i^*\rangle$ or $t^*s_0^*$, respectively. For then we have $Rr^* = 0$ and

$$\begin{aligned}
|r^*| &\leq \exp^k(\max(|t^*|, \sup_i |s_i^*|) + 1) \quad \text{with } \exp\alpha := 2^\alpha \\
&\leq \exp^k(\max(\varepsilon_{|t|}, \sup_i \varepsilon_{|s_i|}) + 1) \\
&= \exp^k(\varepsilon_{\max(|t|, \sup_i |s_i|)} + 1) \\
&< \varepsilon_{\max(|t|, \sup_i |s_i|) + 1} \\
&= \varepsilon_{|r|}
\end{aligned}$$

□

§ 4 Restrictions on infinite sequences in terms

Up to now our terms have been allowed to be rather arbitrary; in particular, in the case of the rule $Et\langle s_i\rangle_{i<\omega}$ there has not been any restriction on how the terms s_i should be choosen. However, such a restriction seems to be advisable when we want to use terms to represent the primitive recursive functionals. For otherwise e.g. any function f of type $0 \rightarrow 0$ could be represented, namely by $\lambda x E x \langle \underline{f(i)}\rangle_{i<\omega}$.

Now what requirements should be satisfied by a properly restricted class of terms? First of all, we certainly want that the functionals represented by such terms are exactly the primitive recursive functionals. Second, it seems advisable that the formation of normal forms described above does not lead us out of our class, for then it is possible to give a detailed analysis of the trade-off of higher type primitive recursion against lower type α-recursion, for some $\alpha < \varepsilon_0$. We do not go into the details of this matter here but just refer to [Schwichtenberg 1973] or [Schwichtenberg 1975] and the literature cited here.

To fulfill those requirements we certainly have to bound the rank of the terms s_i in $\mathrm{Et}\langle s_i\rangle_{i<\omega}$, and further have to bound their heights by an ordinal $\alpha < \varepsilon_0$. In order to exclude the possibility of representing arbitrary functions as above, we only allow terms that can be "coded" by natural numbers. This is a rather standard procedure; for later reference we describe the main points of such a coding, following [Schwichtenberg 1977].

Code numbers of terms are defined inductively, corresponding to the inductive build-up of terms. The inductive definition is trivial for the finite term formation rules A, O^+, $\to^+$ and $\to^-$. However, in the case of the rule $O^-:\mathrm{Et}\langle s_i\rangle_{i<\omega}$ we have a difficulty since then we have infinitely many premisses s_i. The idea now is to assume that the codes for the premisses s_i can be enumerated by a primitive recursive function, and to use a code (or primitive recursive index) of such an enumeration function to construct a code number of the whole term. Another essential point is that our code numbers should contain sufficient information about the coded term. In particular, if a number u codes a term r, then we want to be able to read off primitive recursively from u

1. the name of last rule of term formation applied in r and the type of r,
2. an ordinal $<\varepsilon_0$ which is a bound for the height $|r|$,
3. a bound for the rank Rr, and
4. a bound for the (finite) set of variables free in r.

The corresponding primitive recursive functions will be denoted by $\mathrm{Rule}(u)$, $\mathrm{Type}(u)$, $|u|$, $\mathrm{Rank}(u)$ and $\mathrm{Var}(u)$, respectively. The ordinals $<\varepsilon_0$ from 2. are represented by natural numbers; for this we refer to a fixed canonical well-ordering $\prec$ of the natural numbers of order type ε_0.

We do not write out all cases of the inductive definition of the predicate $u \in \mathrm{Code}$ (u is a code number for a term), but rather give two examples corresponding to the rules $\rightarrow^-$ and 0^- :

$\rightarrow^-$: If $u,v \in \mathrm{Code}$, $\mathrm{Type}(u) = \ulcorner\rho\rightarrow\sigma\urcorner$ and $\mathrm{Type}(v) = \ulcorner\rho\urcorner$, then $\langle\ulcorner\rightarrow^-\urcorner, u, v\rangle \in \mathrm{Code}$.

0^- : If, for any i , $[e](i) =: u_i \in \mathrm{Code}$, $\mathrm{Type}(u_i) = \ulcorner\rho\urcorner$, $|u_i| \preceq \ulcorner\alpha\urcorner \prec \ulcorner\beta\urcorner$, $\mathrm{Rank}(u_i) \leq k$ and $\mathrm{Var}(u_i) \subseteq^{\#} \ulcorner\{x_1,\ldots,x_n\}\urcorner$, and furthermore $v \in \mathrm{Code}$, $\mathrm{Type}(v) = \ulcorner 0\urcorner$, $|v| \prec \ulcorner\beta\urcorner$, $\mathrm{Rank}(v) \leq k$ and $\mathrm{Var}(v) \subseteq^{\#} \ulcorner\{x_1,\ldots,x_n\}\urcorner$, then

$$\langle \ulcorner 0^-\urcorner, \ulcorner\rho\urcorner, \ulcorner\beta\urcorner, k, \ulcorner\{x_1,\ldots,x_n\}\urcorner, v, e\rangle \in \mathrm{Code} .$$

Here $[e]$ denotes the primitive recursive function coded by e . $\ulcorner\ldots\urcorner$ denotes as usual a natural code number for the finite object ... ; $\subseteq^{\#}$ corresponds (under the relevant coding of finite sets of variables) to $\subseteq$; $\langle x_0,\ldots,x_{n-1}\rangle$ is a primitive recursive coding of finite sequences of natural numbers with primitive recursive inverses $(x)_i$, i.e. $(\langle x_0,\ldots,x_{n-1}\rangle)_i = x_i$ for $i<n$. We also skip the (trivial) primitive recursive definitions of the functions $\mathrm{Rule}(u),\ldots$ mentioned above.

It is easy to see that all primitive recursive functionals in the sense of [Gödel 1958], i.e. all functionals explicitely definable from the recursion operators of finite types, can be represented by coded terms; this can be done in such a way that the code numbers have heights $|u| \prec \ulcorner\omega.2\urcorner$. Conversely it is easy to see too that any functional represented by a term with code number is primitive recursive.

Furthermore one can show that to the operations on terms defined above (substitution, in the Reduction Lemma 3.2 and in the Theorem 3.4 on the existence of the normal form, part a) there correspond primitive recursive functions on the code numbers (cf. [Schwichtenberg 1977]). However, to the elimination of subterms convertible by evaluation there does not correspond a primitive recursive function any more, but an α - recursive function, where $\alpha < \varepsilon_0$ is the bound on the height of the given term of rank 0 .

§ 5 A normal form for derivations

We now transfer the arguments of § 3 from terms to derivations. More precisely, we want to show that any derivation can be transformed into a normal form with the same assumptions and the same final judgement, where the normal form has the property that it does not contain any subderivation ending with an elimination rule whose main part is obtained by an introduction rule or a critical rule. Such derivations in normal form have many desirable properties, in particular they do not make any "detours", in a sense which can easily be made precise. We shall give some applications of the existence of a normal form for derivations in § 6 (equivalence of a restricted form of our infinitary system T to a version of T without infinite rules) and in § 7 (closure of the restricted form of T against the Markov rule).

The rules of the inductive definition of a derivation have been described in § 1 . Remember that we now identify equivalent terms, i.e. terms with the same normal form. Rules marked with + (-) are called introduction (elimination) rules. In an elimination rule other than

$$(0^-) \qquad \frac{\ldots\ldots \ \underset{\sim}{r} \in A(t=\underline{i}) \ \ldots\ldots \ \text{for all} \ \ i<\omega}{\underset{\sim}{r} \in A}$$

we call its left hand subderivation its main part; in 0^- the term t is considered as the main part of this rule. In an elimination rule other than

$$(\forall^-) \qquad \frac{\underset{\sim}{r} \in \forall x A}{\underset{\sim}{r} s \in A[s]}$$

all subderivations except the main part are called minor parts; in $\forall^-$ the term s is considered as the minor part of this rule.

A derivation is called convertible by evaluation if it ends with an elimination rule whose main part is obtained by an introduction rule for numerals, i.e. 0^+ or $0^{=+}$. The relation " d is convertible by evaluation into e " is defined by

1.
$$\frac{\begin{array}{c}\underline{j}=\underline{i}^{\,(1)}\\ \vert\\ \ldots\ \underset{\sim}{r}\in A\ \ldots\ \text{for all}\quad i<\omega\end{array}}{\underset{\sim}{r}\in A}\,(1) \quad\mapsto\quad \begin{array}{c}\underline{j}=\underline{j}\\ \vert\\ \underset{\sim}{r}\in A\end{array}$$

2.
$$\frac{\underline{j}=\underline{j}\quad \ldots\ \begin{array}{c}\vert\\ s_{1i}=s_{2i}\end{array}\ \ldots\ \text{for all}\quad i<\omega}{E\underline{j}\langle s_{1i}\rangle = E\underline{j}\langle s_{2i}\rangle} \quad\mapsto\quad \begin{array}{c}\vert\\ s_{1j}=s_{2j}\end{array}$$

A derivation is called <u>properly</u> <u>convertible</u> if it ends with an elimination rule whose main part is obtained by a non-numerical introduction rule, i.e. $\to^{=+}$, $\wedge^{+}$, $\to^{+}$ for derivations, $\forall^{+}$ or $\exists^{*+}$. The relation " d is properly convertible into e " is defined by

3.
$$\frac{\dfrac{\begin{array}{c}\vert\\ r_1=r_2\end{array}}{\lambda x r_1=\lambda x r_2}\qquad \begin{array}{c}\vert\\ s_1=s_2\end{array}}{(\lambda x r_1)s_1=(\lambda x r_2)s_2} \quad\mapsto\quad \frac{\begin{array}{c}\vert\\ r_1[s_1]=r_2[s_1]\end{array}\qquad \begin{array}{c}\vert\\ r_2[s_1]=r_2[s_2]\end{array}}{r_1[s_1]=r_2[s_2]}\ .$$

Here the derivation of $r_1[s_1]=r_2[s_1]$ is obtained from the given derivation of $r_1=r_2$ by substituting s_1 for the free occurrences of x in this derivation; axioms $x=x$ are to be replaced by derivations of $s_1=s_1$ built up according to the formation of the term s_1. The derivation of $r_2[s_1]=r_2[s_2]$ is built up corresponding to the formation of r_2 from the given derivation of $s_1=s_2$.

4.
$$\frac{\dfrac{\begin{array}{c}\vert\\ \underset{\sim}{r}\in A\end{array}\qquad \begin{array}{c}\vert\\ \underset{\sim}{s}\in B\end{array}}{\underset{\sim}{r},\underset{\sim}{s}\in A\wedge B}}{\underset{\sim}{r}\in A} \quad\mapsto\quad \begin{array}{c}\vert\\ \underset{\sim}{r}\in A\end{array}$$

$$\dfrac{\dfrac{\begin{array}{c}|\\ \underset{\sim}{r}\in A\end{array}\qquad \begin{array}{c}|\\ \underset{\sim}{s}\in B\end{array}}{\underset{\sim}{r},\underset{\sim}{s}\in A\wedge B}}{\underset{\sim}{s}\in B} \qquad\mapsto\qquad \begin{array}{c}|\\ \underset{\sim}{s}\in B\end{array}$$

5.
$$\dfrac{\dfrac{\begin{array}{c}\underset{\sim}{u}\in A^{(1)}\\ |\\ \underset{\sim}{r}\in B\end{array}}{\lambda \underset{\sim}{ur}\in A\to B}{\scriptstyle (1)}\qquad \begin{array}{c}|\\ \underset{\sim}{s}\in A\end{array}}{(\lambda \underset{\sim}{ur})\underset{\sim}{s}\in B} \qquad\mapsto\qquad \begin{array}{c}|\\ \underset{\sim}{s}\in A\\ |\\ \underset{\sim}{r}[\underset{\sim}{s}]\in B\end{array}$$

6.
$$\dfrac{\dfrac{\begin{array}{c}|\\ \underset{\sim}{r}\in A\end{array}}{\lambda x\underset{\sim}{r}\in \forall xA}}{(\lambda x\underset{\sim}{r})s\in A[s]} \qquad\mapsto\qquad \begin{array}{c}|\\ \underset{\sim}{r}[s]\in A[s]\end{array}$$

7.
$$\dfrac{\dfrac{\begin{array}{c}|\\ \underset{\sim}{r}\in A[s]\end{array}}{s,\underset{\sim}{r}\in \exists^{*}xA}}{\underset{\sim}{r}\in A[s]} \qquad\mapsto\qquad \begin{array}{c}|\\ \underset{\sim}{r}\in A[s]\end{array}$$

A derivation is called *permutatively convertible* if it ends with an elimination rule whose main part is obtained by a critical rule, i.e. O^- for terms, T , O^- for derivations or $O^{=-}$. The relation " d is permutatively convertible into e " is defined by

8.
$$\dfrac{\begin{array}{c}Et\langle s_j\rangle = \underline{i}^{\,(1)}\\ |\\ \ldots\ \underset{\sim}{r}\in A\ \ldots\ \text{for all}\ \ i<\omega\end{array}}{\underset{\sim}{r}\in A}{\scriptstyle (1)}$$

$$\to \quad \begin{array}{c} \cfrac{t=\underline{j}^{\,(2)} \quad \cdots \quad s_{j-1}=s_{j-1} \quad s_j=\underline{i}^{\,(1)} \quad s_{j+1}=s_{j+1} \quad \cdots}{Et\langle s_j\rangle = \underline{i}} \\ \vdots \\ \cfrac{\cfrac{\cdots \; \underset{\sim}{r}\in A \; \cdots \quad \text{for all} \quad i<\omega}{\cdots \; \underset{\sim}{r}\in A \qquad \cdots \quad \text{for all} \quad j<\omega}\,(1)}{\underset{\sim}{r}\in A}\,(2) \end{array}$$

9.

$$\cfrac{\cfrac{\begin{array}{c}\vdots\\ r=s\end{array} \qquad \begin{array}{c}\vdots\\ s=t\end{array}}{r=t} \qquad \cdots \; \begin{array}{c}\vdots\\ p_i=q_i\end{array} \; \cdots \quad \text{for all} \quad i<\omega}{Er\langle p_i\rangle = Et\langle q_i\rangle}$$

$$\mapsto \quad \cfrac{\cfrac{\begin{array}{c}\vdots\\ r=s\end{array} \quad \cdots \; p_i=p_i \; \cdots}{Er\langle p_i\rangle = Es\langle p_i\rangle} \qquad \cfrac{\begin{array}{c}\vdots\\ s=t\end{array} \quad \cdots \; \begin{array}{c}\vdots\\ p_i=q_i\end{array} \; \cdots}{Es\langle p_i\rangle = Et\langle q_i\rangle}}{Er\langle p_i\rangle = Et\langle q_i\rangle}$$

10.

$$\cfrac{\cfrac{\begin{array}{c}\vdots\\ r=s\end{array} \quad \begin{array}{c}\vdots\\ s=t\end{array}}{r=t} \qquad \begin{array}{c}\vdots\\ p=q\end{array}}{rp=tq} \qquad \mapsto \qquad \cfrac{\cfrac{\begin{array}{c}\vdots\\ r=s\end{array} \quad p=p}{rp=sp} \qquad \cfrac{\begin{array}{c}\vdots\\ s=t\end{array} \quad \begin{array}{c}\vdots\\ P=q\end{array}}{sp=tq}}{rp=tq}$$

11.

$$\begin{array}{c} t=\underline{i}^{\,(1)} \\ \vdots \\ \cfrac{\cfrac{\cdots \; \underset{\sim}{r}\in A \; \cdots}{\underset{\sim}{r}\in A}\,(1) \qquad \cdots \; b_j \; \cdots}{\underset{\sim}{s}\in B} \end{array}$$

$$\mapsto \qquad \dfrac{\dots\dots \quad \dfrac{\begin{matrix} t = \underline{i}^{\,(1)} \\ | \\ \underset{\sim}{r} \in A \end{matrix} \quad \dots\ b_j\ \dots}{\underset{\sim}{s} \in B} \quad \dots\dots}{\underset{\sim}{s} \in B}\ (1)$$

12.
$$\dfrac{\dfrac{\begin{matrix} | \\ t_1 = t_2 \end{matrix} \quad \dots\ \begin{matrix} | \\ s_{1i} = s_{2i} \end{matrix}\ \dots\dots}{Et_1\langle s_{1i}\rangle = Et_2\langle s_{2i}\rangle} \qquad \dots\dots\ \begin{matrix} | \\ p_{1j} = p_{2j} \end{matrix}\ \dots\dots}{E(Et_1\langle s_{1i}\rangle)\langle p_{1j}\rangle = E(Et_2\langle s_{2i}\rangle)\langle p_{2j}\rangle}$$

$$\mapsto \qquad \dfrac{\begin{matrix} | \\ t_1 = t_2 \end{matrix} \qquad \dfrac{\begin{matrix} | \\ s_{1i} = s_{2i} \end{matrix} \quad \dots\dots\ \begin{matrix} | \\ p_{1j} = p_{2j} \end{matrix}\ \dots\dots}{\dots\dots\ Es_{1i}\langle p_{1j}\rangle = Es_{2i}\langle p_{2j}\rangle\ \dots\dots}}{Et_1\langle Es_{1i}\langle p_{1j}\rangle\rangle = Et_2\langle Es_{2i}\langle p_{2j}\rangle\rangle}$$

13.
$$\dfrac{\dfrac{\begin{matrix} | \\ t_1 = t_2 \end{matrix} \quad \dots\dots\ \begin{matrix} | \\ s_{1i} = s_{2i} \end{matrix}\ \dots\dots}{Et_1\langle s_{1i}\rangle = Et_2\langle s_{2i}\rangle} \qquad \begin{matrix} | \\ p_1 = p_2 \end{matrix}}{(Et_1\langle s_{1i}\rangle)p_1 = (Et_2\langle s_{2i}\rangle)p_2}$$

$$\mapsto \qquad \dfrac{\begin{matrix} | \\ t_1 = t_2 \end{matrix} \qquad \dfrac{\begin{matrix} | \\ s_{1i} = s_{2i} \end{matrix} \quad \begin{matrix} | \\ p_1 = p_2 \end{matrix}}{\dots\dots\ s_{1i}p_1 = s_{2i}p_2\ \dots\dots}}{Et_1\langle s_{1i}p_1\rangle = Et_2\langle s_{2i}p_2\rangle}$$

The relation of <u>reducibility</u> between derivations is the smallest reflexive and transitive relation containing the relation of convertibility and closed against the rules for forming derivations. A derivation is said to be in <u>normal form</u> if it contains no properly or permutatively convertible subderivation and also no subderivation convertible by evaluation.

We want to show that any derivation can be reduced to a derivation in normal form. To achieve this it suffices to produce a normal form w.r.t. proper and permutative conversions; this is carried out in Theorem 5.4 below. For after this is done one can easily eliminate all subderivations convertible by evaluation: first replace each subderivation

$$\dfrac{\ldots\; \underset{\sim}{r} \in A \;\ldots \quad \text{for all} \quad i < \omega}{\underset{\sim}{r} \in A}\,(1) \qquad \text{(each premiss derived from } \underline{j} = \underline{i}^{(1)}) \qquad \text{by} \qquad \begin{array}{c} \underline{j} = \underline{j} \\ | \\ \underset{\sim}{r} \in A \end{array},$$

and then replace each subderivation

$$\dfrac{\underline{j} = \underline{j} \qquad \ldots\; s_{1i} = s_{2i} \;\ldots}{E\underline{j}\langle s_{1i}\rangle = E\underline{j}\langle s_{2i}\rangle} \qquad \text{by} \qquad \begin{array}{c} | \\ s_{1j} = s_{2j} \end{array}.$$

We also want to note the effect of the construction of the normal form to the height of the derivations involved. The notion of the height of a derivation is not completely unproblematic: on has to take into account that certain rules for forming derivations implicitely have terms as further premisses (these are A , O^- for derivations, $\forall^-$ and $\exists^{*+}$). Also in order to obtain a proper notion of height we must look in more detail to terms in derivations. So from now on we do not identify terms any more, but rather allow that in the rules for forming derivations described in § 1 the judgements written there are replaced by "variants". A <u>variant</u> $[\underset{\sim}{r} \in A]$ of a judgement $\underset{\sim}{r} \in A$ is a judgement $\underset{\sim}{r}' \in A'$, where each r_i' is equivalent to r_i , and A' is equivalent to A , i.e. obtained from A by replacing equations $r = s$ in A by $r' = s'$, with r', s' equivalent to r, s .

We define the <u>height</u> $|d|$ <u>of a derivation</u> d for a judgement $\underset{\sim}{r} \in A$ by induction on d , as follows:
$|d| = \max(|\underset{\sim}{r} \in A|, |d|_o)$, where $|\underset{\sim}{r} \in A| = \max(|r_1|, \ldots, |r_n|, |A|)$ with $|A|$ defined by $|r = s| = \max(|r|, |s|)$, $|A \wedge B| = |A \to B| =$

$= \max(|A|, |B|)$, $|\forall xA| = |\exists^* xA| = |A|$, and where $|d|_o$ is defined by

$$|\underset{\sim}{u} \in A|_o = |x = x|_o = |\underline{j} = \underline{j}|_o = 1$$

$$\left| \frac{r \overset{|}{=} s \qquad s' \overset{|}{=} t}{[r = t]} \right|_o = \max(|r \overset{|}{=} s|, |s' \overset{|}{=} t|)$$

$$\left| \frac{\begin{array}{c} [t = \underline{i}]^{i\,(1)} \\ | \\ \ldots\ldots \; [\underset{\sim}{r} \in A]^i \; \ldots\ldots \end{array}}{\underset{\sim}{r} \in A}\,(1) \right|_o = \max\left(|t| , \sup_i \left| \begin{array}{c} [t = \underline{i}]^i \\ | \\ [\underset{\sim}{r} \in A]^i \end{array} \right| \right) + 1$$

$$\left| \frac{t_1 \overset{|}{=} t_2 \quad \ldots\ldots \quad s_{1i} \overset{|}{=} s_{2i} \quad \ldots\ldots}{[Et_1 \langle s_{1i} \rangle = Et_2 \langle s_{2i} \rangle]} \right|_o = \max(|t_1 \overset{|}{=} t_2| , \sup_i |s_{1i} \overset{|}{=} s_{2i}|) + 1$$

$$\left| \frac{r \overset{|}{=} s}{[\lambda xr = \lambda xs]} \right|_o = |r \overset{|}{=} s|$$

$$\left| \frac{t_1 \overset{|}{=} t_2 \qquad s_1 \overset{|}{=} s_2}{[t_1 s_1 = t_2 s_2]} \right|_o = \max(|t_1 \overset{|}{=} t_2|, |s_1 \overset{|}{=} s_2|) + 1$$

$$\left| \frac{\underset{\sim}{r} \overset{|}{\in} A \qquad \underset{\sim}{s} \overset{|}{\in} B}{[\underset{\sim}{r}, \underset{\sim}{s} \in A \wedge B]} \right|_o = \max(|\underset{\sim}{r} \overset{|}{\in} A|, |\underset{\sim}{s} \overset{|}{\in} B|)$$

$$\left|\frac{\overset{|}{\underset{\sim}{r},\underset{\sim}{s} \in A \wedge B}}{[\underset{\sim}{r} \in A]}\right|_{0} = \left|\frac{\overset{|}{\underset{\sim}{r},\underset{\sim}{s} \in A \wedge B}}{[\underset{\sim}{s} \in B]}\right|_{0} = |\overset{|}{\underset{\sim}{r},\underset{\sim}{s} \in A \wedge B}| + 1$$

$$\left|\frac{\begin{array}{c}\underset{\sim}{u} \in A^{(1)} \\ | \\ \underset{\sim}{r} \in B\end{array}}{[\lambda \underset{\sim}{u}\underset{\sim}{r} \in A \rightarrow B]}\,(1)\right|_{0} = \left|\begin{array}{c}\underset{\sim}{u} \in A \\ | \\ \underset{\sim}{r} \in B\end{array}\right|$$

$$\left|\frac{\overset{|}{\underset{\sim}{t} \in A \rightarrow B} \qquad \overset{|}{\underset{\sim}{s} \in A'}}{[\underset{\sim}{t}\underset{\sim}{s} \in B]}\right|_{0} = \max(|\overset{|}{\underset{\sim}{t} \in A \rightarrow B}|\ ,\ |\overset{|}{\underset{\sim}{s} \in A'}|) + 1$$

$$\left|\frac{\overset{|}{\underset{\sim}{r} \in A}}{[\lambda x \underset{\sim}{r} \in \forall x A]}\right|_{0} = |\overset{|}{\underset{\sim}{r} \in A}|$$

$$\left|\frac{\overset{|}{\underset{\sim}{t} \in \forall x A}}{[\underset{\sim}{t} s \in A[s]]}\right|_{0} = \max(|\overset{|}{\underset{\sim}{t} \in \forall x A}|\ ,\ |s|) + 1$$

$$\left|\frac{\overset{|}{\underset{\sim}{r} \in A[s]}}{[s,\underset{\sim}{r} \in \exists^{*} x A]}\right|_{0} = \max(|\overset{|}{\underset{\sim}{r} \in A[s]}|\ ,\ |s|)$$

$$\left|\frac{\overset{|}{s,\underset{\sim}{r} \in \exists^{*} x A}}{[\underset{\sim}{r} \in A[s]]}\right|_{0} = |\overset{|}{s,\underset{\sim}{r} \in \exists^{*} x A}| + 1 \ .$$

One can easily see that with this definition we have

<u>Lemma 5.1</u> (Substitution Lemma)

a. Let d be a derivation

$$d \begin{cases} \underset{\sim}{u} \in A \\ \quad | \\ \underset{\sim}{r} \in B \end{cases},$$

x be a variable different from all assumption variables (i.e. variables u_i in assumptions $\underset{\sim}{u} \in A$) and s be a term of the same type as x . Then we can find a derivation

$$d[s] \begin{cases} \underset{\sim}{u} \in A[s] \\ \quad | \\ \underset{\sim}{r}[s] \in B[s] \end{cases}$$

of height $|d[s]| \leq |s| + |d|$.

b. Let d,e be derivations

$$d \begin{cases} \underset{\sim}{u} \in A \\ \quad | \\ \underset{\sim}{r} \in B \end{cases} \qquad e \begin{cases} \\ \quad | \\ \underset{\sim}{s} \in A' \end{cases}$$

with A' equivalent to A . Then we can find a derivation

$$d[e] \begin{cases} \quad | \\ \underset{\sim}{s} \in A \\ \quad | \\ \underset{\sim}{r}[\underset{\sim}{s}] \in B[\underset{\sim}{s}] \end{cases}$$

of height $|d[e]| \leq |e| + |d|$.

Furthermore it is clear that conversions by evaluation can only lower the the height of a term.

The level LA of a formula A is defined by $L(r=s)=Lr$, $L(A \wedge B) = \max(LA,LB)$, $L(A \to B) = \max(LA+1, LB)$, $L(\forall xA) = \max(Lx+1, LA)$ and $L(\exists^* xA) = \max(Lx,LA)$. If $\rho_1,\ldots,\rho_n$ are the types of A (cf. § 1), then we clearly have $L\rho_i \leq LA$. The <u>level</u> Ld of a derivation d is the level of the formula in the judgement derived by d . By the <u>rank</u> Rd of a derivation d we mean the least number greater than all levels of main parts of properly or permutatively convertible subderivations of d .

<u>Lemma 5.2</u> (Reduction Lemma) Let d be a derivation ending with an elimination with main part c and minor parts b_i. If $Lc \leq k$ and $Rc, Rb_i \leq k$, then d can be reduced to a derivation e with $Re \leq k$ and $|e| \leq \max(|d|, (\sup_i |b_i|) + |c|)$.

Note that in the case of the rule 0^- for derivations the main part is a term t, and in the case of the rule $\forall^-$ the minor part is a term s.

The proof is by induction on c. We may assume that c ends with a non-numerical introduction or with a critical rule, for otherwise we already would have $Rd \leq k$. Corresponding the ten last rule applied in c we distinguish the following cases.

<u>Case $\to^{=+}$:</u> Then the given derivation d has the form

$$\dfrac{c\left\{\dfrac{\begin{matrix} | \\ r_1 = r_2 \end{matrix}}{[\lambda x r_1 = \lambda x r_2]}\right. \qquad \left.\begin{matrix} | \\ s_1 = s_2 \end{matrix}\right\} b}{[(\lambda x r_1)s_1 = (\lambda x r_2)s_2]} \quad .$$

For the derivation e to be constructed we take

$$\left.\dfrac{\begin{matrix} | \\ r_1[s_1] = r_2[s_1] \end{matrix} \qquad \begin{matrix} | \\ r_2[s_1] = r_2[s_2] \end{matrix}}{r_1[s_1] = r_2[s_2]}\right\} e$$

Since $L(\lambda x r_i) \leq k$ we have $Ls_i < k$. Hence the derivation of $r_1[s_1] = r_2[s_1]$ has a rank $\leq k$, and also the derivation of $r_2[s_1] = r_2[s_2]$ has a rank $\leq k$. So $Re \leq k$. For the height we obtain by Lemma 5.1

$$\begin{aligned} |e| &= \max(|\overset{|}{r_1[s_1] = r_2[s_1]}|, |\overset{|}{r_2[s_1] = r_2[s_2]}|) \\ &\leq \max(|s_1| + |\overset{|}{r_1 = r_2}|, |\overset{|}{s_1 = s_2}| + |r_2|) \\ &\leq |\overset{|}{s_1 = s_2}| + |\overset{|}{r_1 = r_2}| \\ &\leq |b| + |c| \end{aligned}$$

The cases $\wedge^+$ and $\exists^{*+}$ are trivial: for e we can take a subderivation of d .

Case $\to^+$:

$$c\left\{\begin{array}{c} \underset{\sim}{u} \in A^{(1)} \\ | \\ \underset{\sim}{r} \in B \\ \hline [\lambda \underset{\sim}{u}\underset{\sim}{r} \in A \to B] \end{array}\right. \quad \left.\begin{array}{c} | \\ \underset{\sim}{s} \in A' \end{array}\right\}b \atop \overline{\qquad [(\lambda \underset{\sim}{u}\underset{\sim}{r})\underset{\sim}{s} \in B] \qquad} \qquad \mapsto \qquad \left.\begin{array}{c} | \\ \underset{\sim}{s} \in A \\ | \\ \underset{\sim}{r}[\underset{\sim}{s}] \in B \end{array}\right\}e$$

Since $L(A \to B) \le k$ we have $LA < k$ and hence also $Ls_i < k$ for each s_i from $\underset{\sim}{s}$. So $Re \le k$. From Lemma 5.1 we obtain $|e| \le |b| + |c|$.

Case $\forall^+$:

$$c\left\{\begin{array}{c} | \\ \underset{\sim}{r} \in A \\ \hline [\lambda x \underset{\sim}{r} \in \forall x A] \end{array}\right. \atop \overline{[(\lambda x \underset{\sim}{r}) s \in A[s]]} \qquad \mapsto \qquad \left.\begin{array}{c} | \\ \underset{\sim}{r}[s] \in A[s] \end{array}\right\} e$$

Since $L(\forall x A) \le k$ we have $Ls < k$. Hence $Re \le k$. From Lemma 5.1 we obtain $|e| \le |s| + |c|$.

Case O^- for terms: Then the given derivation d has the form

$$\frac{\ldots\; b_i\left\{\begin{array}{c} [Et\langle s_j\rangle = \underline{i}]^{i\ (1)} \\ | \\ [\underset{\sim}{r} \in A]^i \end{array}\right. \;\ldots \quad \text{for all } i < \omega}{\underset{\sim}{r} \in A} \;.$$

For the derivation e to be constructed we take

$$\frac{\ldots \quad \mathrm{IH}\left(\dfrac{\ldots \quad \begin{array}{c} \dfrac{t^i = \underline{j}^{(2)} \;\ldots\; s^i_{j-1} = s^i_{j-1},\ s^i_j = \underline{i}^{(1)},\ s^i_{j+1} = s^i_{j+1} \;\ldots}{Et^i\langle s^i_j\rangle = \underline{i}} \\ | \\ [\underset{\sim}{r} \in A]^i \end{array} \;\ldots \text{ for all } i < \omega}{\underset{\sim}{r} \in A}(1)\right) \;\ldots \text{ for all } j < \omega}{\underset{\sim}{r} \in A}(2) ,$$

where t^i, s^i_j are the terms equivalent to t, s_j used in $[Et\langle s_j\rangle = \underline{i}]^i$. IH(....) is the result of applying the induction hypothesis to the derivation shown; this application is possible since s_j is a proper subterm of $Et\langle s_i\rangle$. We clearly have $Re = 0$ (note that $k = 0$ in this case). For the height we obtain

$$|e| \le \max(|t|\ ,\ \sup_j[\max(|\underset{\sim}{r} \in A|\ ,(\sup_i \left|\begin{array}{c} t^i = \underline{j}\ \ldots\ s^i_j = \underline{i}\ \ldots \\ \hline Et^i\langle s^i_j\rangle = \underline{i} \\ | \\ [\underset{\sim}{r} \in A]^i \end{array}\right|) + |s_j|)]) + 1$$

$$\le \max(|t|\ ,\ (\sup_i|b_i|) + \sup_j|s_j|) + 1$$

$$\le (\sup_i|b_i|) + |Et\langle s_j\rangle|\ .$$

<u>Case T :</u> We have two subcases according to whether the last rule in d is $0^{=-}$ or $\to^{=-}$. Since both subcases are treated similarly, we restrict ourselves to the first one. Then the given derivation d has the form

$$\begin{array}{c} c\left\{\begin{array}{ll} \dfrac{r = s \qquad s' = t}{[r = t]} & \ldots\ p_i = q_i \end{array}\right\} b_i \quad \ldots \\ \hline [Er\langle p_i\rangle = Et\langle q_i\rangle] \end{array}$$

For the derivation e to be constructed we take

$$\dfrac{IH\left(\dfrac{r = s \qquad p_i = p_i\ \ldots}{Er\langle p_i\rangle = Es\langle p_i\rangle}\right) \qquad IH\left(\dfrac{s' = t\ \ \ldots\ p_i = q_i\ \ldots}{Es'\langle p_i\rangle = Et\langle q_i\rangle}\right)}{Er\langle p_i\rangle = Et\langle q_i\rangle}\ .$$

It is clear that $Re = 0$ (note that again $k = 0$ in this case), and for the height we obtain

$$|e| \le \max(\max(|\mathrm{Er}\langle p_i\rangle \overset{|}{=} \mathrm{Es}\langle p_i\rangle|\ ,\ (\sup_i |p_i|) + |r \overset{|}{=} s|)\ ,$$

$$\max(|\mathrm{Es}'\langle p_i\rangle \overset{|}{=} \mathrm{Et}\langle q_i\rangle|\ ,\ (\sup_i |p_i \overset{|}{=} q_i|) + |s' \overset{|}{=} t|))$$

$$\le \max(|\overset{|}{r} = s| + 1\ ,\ (\sup_i \overset{|^i}{|} b_i|) + |r=s|\ ,\ |s'=t| + 1\ ,\ (\sup_i |b_i|) + |s'=\overset{|}{t}|)$$

$$\le \max(|d|\ ,\ (\sup_i |b_i|) + |c|)\ .$$

<u>Case O^- for derivations:</u> Then the given derivation d has the form

$$\dfrac{c\left\{\dfrac{\begin{array}{c}[t=\underline{i}]^{i}\ (1)\\ |\\ \dots[\underset{\sim}{r}\in A]^{i}\dots\end{array}}{\underset{\sim}{r}\in A}(1)\right. \qquad \dots\ b_j\ \dots}{\underset{\sim}{s}\in B}\ \text{Elimination.}$$

For the derivation e to be constructed we take

$$\dfrac{\dots\quad \mathrm{IH}\left(\dfrac{\begin{array}{c}[t=i]^{i}\ (1)\\ |\\ [\underset{\sim}{r}\in A]^{i}\end{array} \qquad \dots\ b_j\ \dots}{\underset{\sim}{s}\in B}\right)\quad \dots}{\underset{\sim}{s}\in B}(1)\ .$$

It is clear that $\mathrm{Re} \le k$, and for the height we obtain

$$|e| \le \max(|t|, \sup_i(\max(\left|\begin{array}{c}[t=i]^{i}\\ |\\ \underset{\sim}{s}\in B\end{array}\right|\ ,\ (\sup_j |b_j|) + \left|\begin{array}{c}[t=i]^{i}\\ |\\ [\underset{\sim}{r}\in A]^{i}\end{array}\right|))) + 1$$

$$= \max(|t|\ ,\ \sup_i(\max(\left|\begin{array}{c}[t=i]^{i}\\ |\\ [\underset{\sim}{r}\in A]^{i}\end{array}\right| + 1\ ,\ (\sup_j |b_j|) + \left|\begin{array}{c}[t=i]^{i}\\ |\\ [\underset{\sim}{r}\in A]^{i}\end{array}\right|))) + 1$$

$$\le \max(|c|\ ,\ (\sup_j |b_j|) + \sup_i \left|\begin{array}{c}[t=i]^{i}\\ |\\ [\underset{\sim}{r}\in A]^{i}\end{array}\right|) + 1$$

$$\le \max(|d|\ ,\ (\sup_j |b_j|) + |c|)\ .$$

Case $O^{=-}$: We again have two subcases according to whether the last rule in d is $O^{=-}$ or $\rightarrow^{=-}$. Since both subcases are treated similarly, we restrict ourselves to the second one. Then the given derivation d has the form

$$\dfrac{c\left\{\dfrac{\overset{|}{t_1 = t_2} \quad \ldots \quad \overset{|}{s_{1i} = s_{2i}} \quad \ldots}{[Et_1\langle s_{1i}\rangle = Et_2\langle s_{2i}\rangle]}\right. \qquad \left.\overset{|}{p_1 = p_2}\right\}b}{[(Et_1\langle s_{1i}\rangle)p_1 = (Et_2\langle s_{2i}\rangle)p_2]} \quad .$$

For the derivation e to be constructed we take

$$\dfrac{\overset{|}{t_1 = t_2} \quad \ldots . \quad IH\left(\dfrac{\overset{|}{s_{1i} = s_{2i}} \quad \overset{|}{p_1 = p_2}}{s_{1i}p_1 = s_{2i}p_2}\right) \ldots .}{Et_1\langle s_{1i}p_1\rangle = Et_2\langle s_{2i}p_2\rangle} \quad .$$

It is clear again that $Re \leq k$, and for the height we obtain

$$\begin{aligned}
|e| &\leq \max(|\overset{|}{t_1 = t_2}|, \sup_i \max(|\overset{|}{s_{1i}p_1 = s_{2i}p_2}|, |\overset{|}{p_1 = p_2}| + |\overset{|}{s_{1i} = s_{2i}}|)) + 1 \\
&= \max(|\overset{|}{t_1 = t_2}|, \sup_i \max(|\overset{|}{s_{1i} = s_{2i}}| + 1, |\overset{|}{p_1 = p_2}| + |\overset{|}{s_{1i} = s_{2i}}|)) + 1 \\
&\leq \max(|c|, |b| + \sup_i |\overset{|}{s_{1i} = s_{2i}}|) + 1 \\
&\leq \max(|d|, |b| + |c|) \ .
\end{aligned}$$

Lemma 5.3 (Structural Lemma) Consider a derivation of rank $\leq k+1$ ending with an elimination with main part c of a level $Lc \geq k+1$. Then c is obtained by in general multiple uncritical eliminiations from an assumption, from an axiom $x = x$ or from a variable x.

Proof: If $c \equiv: c_o$ is obtained by an uncritical elimination, take its main part c_1. If c_1 again is obtained by an uncritical elimination, take again its main c_2, and so on. Note that all c_i have levels $\geq k+1$. The procedure terminates with a main part c_n. By construction c_n is not obtained by an uncritical elimination. Since by assumption the whole derivation has a rank $\leq k+1$, we also know that c_n can neither be obtained by an introduction nor by a critical rule. Hence the claim follows.

Theorem 5.4 (Existence of the Normal Form)

a. Any derivation d of rank $\leq k+1$ can be reduced to a derivation d' of rank $\leq k$, such that $|d'| \leq 2^{|d|}$.

b. Any derivation d (of arbitrary rank $\leq \omega$) can be reduced to a derivation d^* of rank 0, such that $|d^*| \leq \varepsilon_{|d|}$.

The proof is by induction on d. If d doesn't end with an elimination claims a and b follow immediately from the induction hypothesis. Otherwise d has the form

$$\frac{c\left\{\begin{array}{l} \vert \\ \underset{\sim}{r} \in A \end{array}\right. \qquad \ldots\ b_i\ \ldots}{\underset{\sim}{s} \in B}\ \text{Elimination.}$$

Note that in the case of O^- for derivations the term t occurring there is to be considered as main part c, and in the case of $\forall^-$ the term s occurring there is to be considered as the single minor part b_j.

Part a: By induction hypothesis we have c', b_i' with Rc', $Rb_i' \leq k$ and $|c'| \leq 2^{|c|}$, $|b_i'| \leq 2^{|b_i|}$.

Case 1: The main part c has a level $Lc \geq k+1$. Then by the Structural Lemma 5.3 we know that c is obtained by in general multiple uncritical eliminations from an assumption, from an axiom $x=x$ or from a variable x. As in the proof of Theorem 3.4 we know the form of c' by the induction hypothesis of a claim to be proved simultaneously: c' is obtained from the same assumption, the same axiom $x=x$ or the same variable, respectively, by the same uncritical elimination rules, where in each of those rules the minor parts have to be reduced corresponding to the induction hypothesis. We let d' be

$$\dfrac{c'\left\{\begin{array}{l} \vert \\ \underset{\sim}{r} \in A \end{array}\right. \qquad \ldots\ b_i'\ \ldots}{\underset{\sim}{s} \in B} \text{ Elimination.}$$

Then we know from the form of d' that $Rd' \leq k$. For the height we have

$$\begin{aligned} |d'| &= \max(|c'|, \sup_i |b_i'|) + 1 \\ &\leq \max(2^{|c|}, \sup_i 2^{|b_i|}) + 1 \\ &< 2^{\max(|c|, \sup_i |b_i|) + 1} \\ &= 2^{|d|}. \end{aligned}$$

Case 2 : The main part c has a level $Lc \leq k$. Then we can apply the Reduction Lemma 5.2. We obtain the required derivation d' with rank $Rd' \leq k$. For the height we have

$$\begin{aligned} |d'| &\leq \max(\max(|c'|, \sup_i |b_i'|) + 1, (\sup_i |b_i'|) + |c'|) \\ &= \max(|c'| + 1, (\sup_i |b_i'|) + |c'|) \\ &\leq \max(2^{|c|} + 1, (\sup_i 2^{|b_i'|}) + 2^{|c|}) \\ &\leq \max(2^{|c|+1}, 2^{\max(|c|, \sup_i |b_i|) + 1}) \\ &= 2^{|d|}. \end{aligned}$$

<u>Part b :</u> By induction hypothesis we have c^*, b_i^* with Rc^*, $Rb_i^* = 0$ and $|c^*| \leq \varepsilon_{|c|}$, $|b_i^*| \leq \varepsilon_{|b_i|}$. Let $k < \omega$ be the level of the main part c. Then we construct the required derivation d^* by a k-fold application of Part a to

$$\dfrac{c^*\left\{\begin{array}{l} \vert \\ \underset{\sim}{r} \in A \end{array}\right. \qquad \ldots\ b_i^*\ \ldots}{\underset{\sim}{s} \in B} \text{ Elimination.}$$

For then we have $Rd^* = 0$ and

$$|d^*| \leq \exp^k(\max(|c^*|\ ,\ \sup_i|b_i^*|) + 1) \quad \text{with } \exp\alpha := 2^\alpha$$

$$\leq \exp^k(\max(\varepsilon_{|c|}\ ,\ \sup_i \varepsilon_{|b_i|}) + 1)$$

$$< \varepsilon_{\max(|c|\ ,\ \sup_i|b_i|) + 1}$$

$$= \varepsilon_{|d|}\ .$$

§ 6 Restrictions on infinite derivations

We now want to impose some restrictions on the infinite rules of T . Such restrictions are necessary, for otherwise our system would be unreasonably strong. For instance, any true Π_1^0 - formula, i.e. formula of the form $\forall x(fx = 0)$ with f primitive recursive, could be derived trivially.

Now what requirements should be satisfied by a properly restricted class of derivations ? First of all, we want the resulting system to be equivalent to the finitary version of T , where we have induction axioms instead of the ω-rule 0^- , and constants for the recursion operators of finite types instead of the infinite terms. Second, we obviously want that the formation of normal forms described in § 5 does not lead us out of our class.

To fulfill those requirements we introduce code numbers for derivations. This can be done in a similar way as it was done for terms in § 4 ; hence we omit the details. However, the mere existence of a code number is not yet a strong enough restriction: the trivial derivation mentioned above of an arbitrary true Π_1^0 - formula can certainly be coded. What seems to be necessary here is a reference to a formal system, e.g. the finitary version of T , and to require that a derivation not only has a code, but furthermore that this fact is derivable in the finitary version of T .

Note that there is a problem here: Since the notion of a code number of a derivation (or of a term) is defined inductively, it is not immediately clear how this notion is to be represented in T . To solve this problem, let us first argue that the set of code numbers of terms is Π_1^0 . To see this, think of terms as given by well-founded trees, where at each node there is either no branching at all (i.e., it is a topmost node) and a variable or a numeral is affixed, or there is a 1-fold branching (corresponding to the rule $\to^+$), or a 2-fold branching (corresponding to the rule $\to^-$), or an ω-fold branching (corresponding to the rule O^-). Then any code number u of a term r can be thought of as obtained inductively by affixing to each node of the tree corresponding to r a code number of the corresponding subterm. Hence the property "u is a code number of a term" is equivalent to u having such a well-founded genealogic tree. But the latter fact can be easily written in Π_1^0 - form: One has to express that at any node (= sequence number) n the term is locally correct, i.e. that the code number u_n affixed there (u_n can be easily defined by induction on n) and all its predecessors $u_{n*\langle i\rangle}$ fulfill a relation as given in the definition of code numbers for terms. The well-foundedness is then obtained automatically, since in particular $|u_{n*\langle i\rangle}| \prec |u_n|$ is required and $\prec$ is a well-ordering.

Similarly, one can see that the set of code numbers for derivations is Π_1^0 . The only additional difficulty here is that our rules for defining derivations involve the notion of equivalence of terms. Hence we have to know that the relation "u and v are code numbers of equivalent terms" is Π_1^0 . Now this means that the code numbers u^* and v^* of the normal forms of u and v (which can be obtained recursively from u and v ; cf. § 4) code terms which are the same up to a change of bound variables. But from any code number u of a term one can define primitive recursively a code number u' of an equivalent term whose bound variables are choosen in a fixed, standard way: e.g., at a node n where λx^ρ is introduced (by the rule $\to^+$) use the variable z^ρ_{n+k} instead of x^ρ, where $k-1$ is the maximal index of variables free in the term coded by u. Hence the above relation between u^* and v^* is Π^0_1.

So we restrict the infinite rules of T as follows: We only allow a derivation if it has a code number u, and furthermore if the Π^0_1-formula expressing the fact that u codes a derivation is derivable in the finitary version of T (or equivalently, in pure arithmetic Z).

We have to show now that the restricted system T 1. is equivalent to the finitary version of T, and 2. is closed against the formation of normal forms of derivations as described in § 5. The second fact can be seen easily by formalizing the arguments in § 5. For the first fact, one direction is obviously true, since we can embed the finitary version of T into our restricted system. For the other direction we use the formalization of the normal form theorem of § 5, as follows.

First note that a normal derivation has the subformula property, i.e. it contains only subformulas of the derived formula or of an assumption. Here the notion of a subformula is as usual, with the sole exception that any equation $r^\rho_1 = r^\rho_2$ is a subformula of $s^\rho_1 = s^\rho_2$ (this is a consequence of the transitivity rule T). Now assume that a code u of height $\alpha < \varepsilon_0$ for a derivation of a judgement $\underset{\sim}{r} \in A$ from assumptions $\underset{\sim}{s}_i \in B_i$ is given. We may assume that u is in normal form. Hence only subformulas of A and the finitely many B_i's can occur in this derivation. So, using a truth predicate choosen accordingly, we can prove by α-induction that $\underset{\sim}{r} \in A$ is true provided that the $\underset{\sim}{s}_i \in B_i$ are true, and hence can derive the judgement $\underset{\sim}{r} \in A$ from the assumptions $\underset{\sim}{s}_i \in B_i$ in the finitary version of T.

§ 7 The Markov rule

As an example of possible applications of the normal form theorem of § 5 we want to show now that T (i.e., from now on, T with its infinite rules restricted as described in § 6) is closed against the Markov rule. This is to say that from a derivation of $\exists y(fxy = \underline{0})$ with the classical existential quantifier $\exists y$ and variables x,y of type 0 (and f a closed term of type $0,0 \to 0$) we can construct a term h of type $0 \to 0$ (so h denotes a $<\varepsilon_0$-recursive function) and a derivation of $hx \in \overset{*}{\exists} y(fxy = \underline{0})$, now with the constructive existential quantifier $\overset{*}{\exists} y$.

Note that the well-known method due to [Dragalin 1980] and [H. Friedman 1978] of proving closure under the Markov rule does not work for our system T involving realizing terms. For this method consists of replacing each equation $r = s$ in each formula of a given derivation of $\exists y(fxy = 0)$ by $r = s \vee^* \overset{*}{\exists} y(fxy = 0)$, and observing 1. that this transformation does not destroy the property of being a derivation, and 2. that the derived formula $\exists y(fxy = 0)$ is transformed into a formula equivalent to $\overset{*}{\exists} y(fxy = 0)$. Now this argument does not provide the realizing terms we would need in order to obtain a derivation in T.

To prove closure of T under the Markov rule, assume that a derivation d of $\exists y(fxy = 0)$ or equivalently of $\neg\neg\overset{*}{\exists} y(fxy = 0)$ is given. By § 5 we can assume that d is in normal form. We may write $\neg\neg\overset{*}{\exists} y(r[x,y] = 0)$ for the endformula of d with $r[x,y]$ being the normal form of fxy, and obviously can assume that x is the only variable free in d. Let us first analyse how our normal derivation d must look like.

If we follow a path from the root of the derivation tree upwards, after possibly some applications of the ω-rule 0^- we must come to an application of $\to^+$ of the form

$$\frac{\underline{0} = \underline{1}}{\neg\neg\overset{*}{\exists} y(r[x,y] = \underline{0})} ,$$

where some assumptions $\neg\overset{*}{\exists} y(r[x,y] = \underline{0})$ are cancelled. Following our path further upwards we may pass through type-0-equations with rules concerning equations, or stop at an axiom, or stop at an assumption $t[x] = \underline{i}$ cancelled at an application of the ω-rule 0^- we passed earlier, or else come to an application of $\to^-$ of the form

$$\frac{\neg\overset{*}{\exists} y(r[x,y] = \underline{0}) \qquad s[x] \in \overset{*}{\exists} y(r[x,y] = \underline{0})}{\underline{0} = \underline{1}} .$$

Here we may assume that the left hand premiss $\neg\overset{*}{\exists} y(r[x,y] = \underline{0})$ is cancelled by the application of $\to^+$ we passed previously. If we follow our path further upwards through the right hand premiss, after possibly some applications of the ω-rule 0^- we must come to an application of $\exists^{*+}$ of the form

$$\frac{r[x,s[x]] = \underline{0}}{s[x] \in \overset{*}{\exists} y(r[x,y] = 0)} .$$

Now we have an equation again, and following our path further upwards we may again pass through type-0-equations with rules concerning equations, or stop at an axiom, or stop at an assumption $t[x] = \underline{i}$ cancelled at an application of the ω-rule 0^- we passed earlier, or else come to an application of $\to^-$ as described above, and so on.

The first step in our proof of closure of T under the Markov rule consists in constructing from the given normal derivation d of $\neg\neg\overset{*}{\exists} y(r[x,y] = \underline{0})$ a $<\varepsilon_0$-recursive function h such that $r[\underline{n}, \underline{h(n)}] = \underline{0}$ holds, for any n. So let n be given. Call a node k (i.e., a finite sequence of numbers) of d an <u>n-node</u> if k is in d and if at each application of the ω-rule 0^- in d

$$\frac{\ldots\ \underset{\sim}{r} \in A\ (t[x] = \underline{i})\ \ldots \qquad \text{for all } i < \omega}{\underset{\sim}{r} \in A}$$

the j-th branch with $j :=$ the value of $t[\underline{n}]$ is in k, provided k is long enough. Call an n-node k <u>critical</u> if at k and at each initial segment of k there is affixed in d either the endformula $\neg\neg\exists\overset{*}{y}(r[x,y]=\underline{0})$, or an equation which is false for $x=\underline{n}$, or a judgement $s[x]\in\exists\overset{*}{y}(r[x,y]=\underline{0})$ which is false for $x=\underline{n}$. Obviously the bottom node 0 is critical. We define a function $g(n,k)$ by α-recursion, $\alpha := |d| < \varepsilon_0$, as follows. If k is not a critical n-node of d, let $g(n,k)=0$. If k is a topmost critical n-node of d, we know that at k we must have the conclusion $\underline{0}=\underline{1}$ of an application of $\to^{-}$ of the form

$$\frac{\neg\exists\overset{*}{y}(r[x,y]=\underline{0}) \qquad s[x]\in\exists\overset{*}{y}(r[x,y]=\underline{0})}{\underline{0}=\underline{1}}$$

where $s[x]\in\exists\overset{*}{y}(r[x,y]=\underline{0})$ is true for $x=\underline{n}$. Let $g(n,k)$ be the value of $s[\underline{n}]$; then $r[\underline{n}, \underline{g(n,k)}]=\underline{0}$ holds. If k is a critical n-node of d but not a topmost one, let $g(n,k) = g(n, k*\langle i\rangle)$ with i minimal such that $k*\langle i\rangle$ is a critical n-node. Hence with $h(n) := g(n,0)$ we have $r[\underline{n}, \underline{h(n)}]=\underline{0}$.

Obviously the above argument can be formalized in the finitary version of T (or equivalently in pure arithmetic Z). Then the definition of the $<\varepsilon_0$-recursive function h yields a closed term of type $0\to 0$ also denoted by h, and we obtain a derivation in the finitary version of T for $fx(hx)=\underline{0}$. Now consider the derivation

$$\cfrac{\cdots \quad \cfrac{\cfrac{\cfrac{\overset{|}{f=f} \quad x=\underline{n}^{(1)}}{fx=f\underline{n}} \qquad \cfrac{\overset{|}{h=h} \quad x=\underline{n}^{(1)}}{hx=h\underline{n}}}{fx(hx)=\underbrace{f\underline{n}(h\underline{n})}_{\underline{0}}}}{hx\in\exists\overset{*}{y}(fxy=\underline{0})} \quad \cdots \quad \text{for all } n<\omega}{hx\in\exists\overset{*}{y}(fxy=\underline{0})}\,(1)\ .$$

By what we just said its local correctness is provable in the finitary version of T, and hence we have a derivation in T of $hx\in\exists\overset{*}{y}(fxy=\underline{0})$, as required.

References

Dragalin, A.G., New Forms of Realizability and Markov's Rule (in Russian). Dokl. Akad. Nauk SSSR 251 (1980). English Translation in: Soviet Math. Dokl. 21 (1980), 461-464.

Friedman, H., Classically and Intuitionistically Provably Recursive Functions. In: G.H. Müller and D.S. Scott, eds., Higher Set Theory (Springer, Berlin, 1978), 21-27.

Gentzen. G., Untersuchungen über das logische Schließen. Math. Zeitschr. 39 (1934), 176-210, 405-431.

Gödel, K., Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. Dialectica 12 (1958), 280-287.

Kreisel, G., Interpretation of Analysis by Means of Constructive Functionals of Finite Types. In: A. Heyting, ed., Constructivity in Mathematics (North-Holland, Amsterdam, 1959), 101-128.

Maaß, W., Eine Funktionalinterpretation der prädikativen Analysis. Dissertation, Universität München, 1974.

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

Minc, G.E., Finite Investigations of Transfinite Derivations (in Russian). Zapiski Leningrad 49 (1975), 67-121. English Translation in: J. Soviet Math. 10 (1978), 548-596.

Ruckert, M., Church-Rosser Theorem für Reduktionskalküle mit permutativen Reduktionen. Manuskript, München, 1983.

Schwichtenberg, H., Einige Anwendungen von unendlichen Termen und Wertfunktionalen. Habilitationsschrift, Münster, 1973.

Schwichtenberg, H., Elimination of Higher Type Levels in Definitions of Primitive Recursive Functionals by Means of Transfinite Recursion. In: H.E. Rose and J.C. Shepherdson, eds., Logic Colloquium '73 (North-Holland, Amsterdam, 1975), 279-303.

Schwichtenberg, H., Proof Theory: Some Applications of Cut-Elimination. In: J. Barwise, ed., Handbook of Mathematical Logic (North-Holland, Amsterdam, 1977), 867-895.

Tait, W.W., Infinitely Long Terms of Transfinite Type. In: J.N. Crossley and M.A.E. Dummet, eds., Formal Systems and Recursive Functions (North-Holland, Amsterdam, 1965), 176-185.

Troelstra, A.S., Metamathematical Investigation of Intuitionistic Arithmetic and Analysis. Springer, Berlin, 1973.