# The constructive Hilbert program and the limits of Martin-Löf type theory*

Michael Rathjen
Department of Pure Mathematics, University of Leeds
Leeds LS2 9JT, United Kingdom
E-Mail: rathjen@amsta.leeds.ac.uk

## 1 Introduction

Hilbert's program is one of the truly magnificent projects in the philosophy of mathematics. To carry out this program he founded a new discipline of mathematics, called *"Beweistheorie"*, which was to perform the task of laying to rest all worries about the foundations of mathematics once and for all[1] by securing mathematics via an absolute proof of consistency. The failure of Hilbert's finitist reduction program on account of Gödel's incompleteness results is often gleefully trumpeted. Modern logic, though, has shown that modifications of Hilbert's program are remarkably resilient. These modifications can concern different parts of Hilbert's two step program[2] to validate infinitistic mathematics.

The first kind maintains the goal of a finitistic consistency proof. Here, of course, Gödel's second incompleteness theorem is of utmost relevance in that only a fragment of infinitistic mathematics can be shown to be consistent. Fortunately, results in mathematical logic have led to the conclusion that this fragment encompasses a substantial chunk of scientifically applicable mathematics (cf. [18, 72]). This work bears on the question of the indispensability of set-theoretic foundations for mathematics.

The second kind of modification gives more leeway to the methods allowed in the consistency proof. Such a step is already presaged in the work of the Hilbert school. Notably Bernays has called for a broadened or extended form of finitism (cf. [4]). Rather than a finistic consistency proof the objective here is to give a constructive and predicative consistency proof for a classical theory $T$ in which large parts of infinitistic mathematics can be developed. In order to undertake such a study fruitfully one needs to point to a particular formalization of constructive predicative reasoning $P$, and then investigate whether $P$ is sufficient to prove the consistency of $T$. The particular framework I shall be concerned with in this paper is an intuitionistic and predicative theory of types which was developed by Martin-Löf. He developed his type theory *"with the philosophical motive of clarifying the syntax and semantics of intuitionistic mathematics"* ([41]). It is intended to be a full scale system for formalizing intuitionistic mathematics. Owing to research in mathematical logic over the last 30 years - the program of reverse mathematics and Feferman's work have been especially instrumental here - one can take a certain fragment of second order arithmetic to be the system $T$. It turns out that Martin-Löf's type theory $P$

*This paper is a slightly revised and expanded version of [65].

[1] *"... die Grundlagenfragen einfürallemal aus der Welt zu schaffen."*

[2] The first step being to formalize the whole of mathematics in a formal system $T$. The second and main step consists in proving the consistency of $T$.

is strong enough to prove the consistency of $T$, thereby validating infinitistic mathematics and proving a constructive Hilbert program to be feasible. Indeed, the system $T$ alluded to above is so capacious that we do not even know of any result in ordinary mathematics which is not provable in $T$. Of course, this is a claim regarding ordinary mathematics only; highly set-theoretic topics in mathematics are not amenable to a constructive consistency proof, or, more cautiously put, we do not know how to give constructive consistency proofs for such topics.

The main goal of this paper is to find the limits of Martin-Löf type theory. A demarcation of the latter is important in determining the ultimate boundaries of a constructive Hilbert program. The aim is to single out a fragment of second order arithmetic or classical set theory which encompasses all possible formalizations of Martin-Löf type theory.

Since only a quarter of this paper will actually be concerned with exploring the limits of Martin-Löf type theory, perhaps some words of explanation are in order. The paper was originally written for my PhD students in order to acquaint them with a research area at the interface of proof theory, constructivism, and the philosophy of mathematics that is not readily available in text book form. This accounts for the at times naïve and avuncular diction. The hope, though, is that the paper might be of use to other audiences as well. Among the fields broached here are the areas of proof theory, constructivism, subsystems of arithmetic, reverse mathematics, set theory, Martin-Löf type theory, and the philosophy of mathematics. Several of the aforementioned topics are presented here ab initio with the aim of making them more accessible. The cognoscenti, though, should skim over the first couple sections and then proceed directly to sections 5 and 6. The present paper is a slightly expanded version of [65]. At first, I intended to revise the paper substantially, but time constraints did not permit me to do so. I resigned to dilating on section 6 which is concerned with the limits of Martin-Löf type theory. I also intended to excise some parts (in particular subsection 4.3) from [65] which I considered to be embarrassing given that the creators of the theories addressed therein are among the authors of this anthology. But I refrained from that also as I didn't see how to cut out pieces without mutilating the paper or revising it substantially.

The following adumbrates how the paper is organized: In section 2, fragments of second order arithmetic are introduced and their role for formalizing various parts of ordinary mathematics is discussed. Section 3 surveys different forms of constructivism. Section 4 provides an informal introduction to the ideas underlying Martin-Löf type theory and also relates them to the Dummett-Prawitz meaning-as-use theory. Section 5 is concerned with subsystems of second order arithmetic which can be shown to be consistent within Martin-Löf type theory. Section 6 is devoted to the limits of Martin-Löf type theory and thus to the limits of a constructive Hilbert program based on it. The final section briefly touches on mathematical statements whose proof depends essentially on the higher infinite in Zermelo-Fraenkel set theory and beyond.

## 2 Systems for formalizing mathematics

A natural modification of Hilbert's program consists in broadening the requirement of reduction to finitary methods by allowing reduction to constructive methods more generally.[3] The objective of our modified constructive Hilbert program is not merely the absence

[3] Such a shift from the original program is implicit in Hilbert-Bernays' [32] apparent acceptance of Gentzen's consistency proof for **PA** under the heading "Überschreitung des bisherigen methodischen Standpunktes der Beweistheorie". The need for a modified Hilbert program has clearly been recognized by Gentzen (cf. [26]) and Bernays [4]: *It thus became apparent that the "finite Standpunkt" is not the only*

of inconsistency but also the demand for a constructive conception for which there is an absolute guarantee that, whenever one proves a 'real' statement in a sufficiently strong classical theory $T$, say, a fragment of second order arithmetic or set theory, there would be an interpretation of the proof according to which the theorem is constructively true. Moreover, one would like the theory $T$ to be such as to make the process of formalization of mathematics in $T$ almost trivial, in particular $T$ should be sufficiently strong for all practical purposes. This is a very Hilbertian attitude: show once and for all that non-constructive methods do not lead to false constructive conclusions and then proceed happily on with non-constructive methods.

There are several aspects of a constructive Hilbert program that require clarification. One is to find some basic constructive principles upon which a coherent system of constructive reasoning may be built. Another is to point to a particular framework for formalizing infinitistic mathematics. The latter task will be addressed in this section. It was already observed by Hilbert-Bernays [32] that classical analysis can be formalized within second order arithmetic. Further scrutiny revealed that a small fragment is sufficient. Under the rubric of *Reverse Mathematics* a research program has been initiated by Harvey Friedman some thirty years ago. The idea is to ask whether, given a theorem, one can prove its equivalence to some axiomatic system, with the aim of determining what proof-theoretical resources are necessary for the theorems of mathematics. More precisely, the objective of reverse mathematics is to investigate the role of set existence axioms in ordinary mathematics. The main question can be stated as follows:

> *Given a specific theorem $\tau$ of ordinary mathematics, which set existence axioms are needed in order to prove $\tau$?*

Central to the above is the reference to what is called 'ordinary mathematics'. This concept, of course, doesn't have a precise definition. Roughly speaking, by ordinary mathematics we mean main-stream, non-set-theoretic mathematics, i.e. the core areas of mathematics which make no essential use of the concepts and methods of set theory and do not essentially depend on the theory of uncountable cardinal numbers. In particular, ordinary mathematics comprises geometry, number theory, calculus, differential equations, real and complex analysis, countable algebra, classical algebra in the style of van der Waerden [77], countable combinatorics, the topology of complete separable metric spaces, and the theory of separable Banach and Frechet spaces. By contrast, the theory of non-separable topological vector spaces, uncountable combinatorics, and general set-theoretic topology are not part of ordinary mathematics. Those parts of mathematics on which set-theoretic assumptions have a strong bearing will be addressed in the last section of this paper.

It is well known that mathematics can be formalized in Zermelo-Fraenkel set theory with the axiom of choice. The framework chosen for studying set existence in reverse mathematics, though, is second order arithmetic rather than set theory. Second order arithmetic, $\mathbf{Z_2}$, is a two-sorted formal system with one sort of variables ranging over natural numbers and the other sort ranging over sets of natural numbers. One advantage of this framework over set theory is that it is more amenable to proof-theoretic investigations. However, at least in my opinion, the particular choice of framework is not pivotal for the program.

For many mathematical theorems $\tau$, there is a weakest natural subsystem $S(\tau)$ of $\mathbf{Z_2}$ such that $S(\tau)$ proves $\tau$. Very often, if a theorem of ordinary mathematics is proved from

*alternative to classical ways of reasoning and is not necessarily implied by the idea of proof theory. An enlarging of the methods of proof theory was therefore suggested: instead of reduction to finitist methods of reasoning it was required only that the arguments be of a constructive character, allowing us to deal with more general forms of inferences.*

the weakest possible set existence axioms, the statement of that theorem will turn out to be provably equivalent to those axioms over a still weaker base theory. This theme is referred to as *Reverse Mathematics*. Moreover, it has turned out that $S(\tau)$ often belongs to a small list of specific subsystems of $\mathbf{Z_2}$ dubbed $\mathbf{RCA}_0$, $\mathbf{WKL_0}$, $\mathbf{WKL_0^+}$, $\mathbf{ACA}_0$, $\mathbf{ATR}_0$ and $(\mathbf{\Pi}^1_1{-}\mathbf{CA})_0$, respectively. The systems are enumerated in increasing strength. The main set existence axioms of $\mathbf{RCA}_0$ $\mathbf{ACA}_0$, $\mathbf{ATR}_0$, and $(\mathbf{\Pi}^1_1{-}\mathbf{CA})_0$ are recursive comprehension, arithmetic comprehension, arithmetical transfinite recursion and $\mathbf{\Pi}^1_1$-comprehension, respectively. Definitions of these systems will be given in the next section. For their role in reverse mathematics see [73].

The theories $\mathbf{WKL_0}$ and $\mathbf{WKL_0^+}$ (defined in section 2.1) are particularly interesting in pursuing a partial realization of the original Hilbert program. Both theories are of the same proof-theoretic strength as primitive recursive arithmetic, **PRA**, a system which is often considered to be co-extensive with finitism (cf. [75]). The principal set existence axiom of $\mathbf{WKL_0}$ is a non-constructive principle known as *weak König's lemma* which asserts that any infinite tree of finite sequences of zeros and ones has an infinite path. Friedman [23] proved via model-theoretic methods that $\mathbf{WKL_0}$ is conservative over **PRA** with respect to $\Pi^0_2$ sentences.

The question as to which parts of mathematics have applications in science, has also been studied intensively by Feferman (cf. [18], [19]). Over the years he has developed several systems for formalizing mathematics. The system $\mathbf{W}_F$ (cf. [19]) (in honor of H. Weyl) is perhaps the most streamlined. $\mathbf{W}_F$ has flexible finite types (over the natural numbers) and allows for very natural reconstructions of the real and complex numbers (as sets) and much of classical and functional analysis. In $\mathbf{W}_F$ one accepts the completed infinite set of natural numbers as well as classical logic. Though, impredicative set comprehension is taboo. $\mathbf{W}_F$ is conservative over Peano arithmetic.

## 2.1 Subsystems of second order arithmetic

The purpose of this section is to introduce the formal system of second order arithmetic and several of its subsystems so as to be able to delineate precisely its constructively justifiable parts. Another purpose is to give definitions of the subsystems figuring in reverse mathematics mentioned above.

The most basic system we shall be concerned with is primitive recursive arithmetic, **PRA**, which is a theory about the natural numbers which has function symbols for all primitive recursive functions but in contrast to Peano arithmetic, **PA**, allows for induction only quantifier free formulae.

The language $\mathcal{L}_2$ of second-order arithmetic contains (free and bound) natural number variables $a, b, c, \ldots, x, y, z, \ldots$, (free and bound) set variables $A, B, C, \ldots, X, Y, Z, \ldots$, the constant 0, function symbols $Suc, +, \cdot$, and relation symbols $=, <, \in$. $Suc$ stands for the successor function.

Terms are built up as usual. For $n{\in}\mathbb{N}$, let $\bar{n}$ be the canonical term denoting $n$. Formulae are built from the prime formulae $s = t$, $s < t$, and $s \in A$ using $\wedge, \vee, \neg, \forall x, \exists x, \forall X$ and $\exists X$ where $s, t$ are terms.

Note that equality in $\mathcal{L}_2$ is a relation only on numbers. However, equality of sets will be considered a defined notion, namely

$$A = B \quad \text{iff} \quad \forall x[x{\in}A \leftrightarrow x{\in}B].$$

The basic axioms in all theories of second-order arithmetic are the defining axioms of $0, Suc, +, \cdot, <$ and the *induction axiom*

$$\forall X(0 \in X \wedge \forall x(x \in X \to x+1 \in X) \to \forall x(x \in X)),$$

where $x+1$ stands for $Suc(x)$. With regard to a collection of $\mathcal{L}_2$ formulae $\mathcal{F}$, the *schema of $\mathcal{F}$-induction* consists of the formulae

$$\mathbf{IND}_{\mathcal{F}} \qquad \varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x),$$

where $\varphi$ belongs to $\mathcal{F}$. If $\mathcal{F}$ is the collection of all $\mathcal{L}_2$ formulae we denote the schema by **IND** rather than $\mathbf{IND}_{\mathcal{F}}$. Note that $\mathbf{IND}_{\mathbf{\Delta}^0_0}$ implies the induction axiom.

The strength of systems of second order arithmetic is largely owed to *comprehension schemes,* which assert, roughly speaking, that if we specify a collection $X$ of numbers by a formula $\varphi$ of a particular type then $X$ is a set. We consider the axiom schema of *$\mathcal{C}$-comprehension* for formula classes $\mathcal{C}$ which is given by

$$\mathcal{C} - \mathbf{CA} \qquad \exists X \forall u(u \in X \leftrightarrow \varphi(u))$$

for all formulae $\varphi \in \mathcal{C}$ in which $X$ does not occur.

The fundamental idea of reverse mathematics is to gauge the proof-theoretical strength of a mathematical theorem by classifying how much comprehension is needed to establish the existence of the sets needed to prove the theorem. That is, we "reverse" the theorem to derive some sort of comprehension scheme. The gauging measure is that of the allowable "logical complexity" of the $\varphi$'s. Typically, this complexity might be the allowable quantifier depth of $\varphi$.

Numerical quantifiers are called bounded if they occur in the context $\forall x(x < s \rightarrow \ldots)$ or $\exists x(x < s \wedge \ldots)$ for a term $s$ which does not contain $x$. The $\mathbf{\Delta}^0_0$-formulae are those formulae in which all quantifiers are bounded numerical quantifiers. For instance, the formula asserting that "$x$ is a prime number" is a $\mathbf{\Delta}^0_0$-formula, i.e., $Prime(x) \equiv \forall u < x+1\, \forall v < x+1(u \cdot v = x \rightarrow u = 1 \vee v = 1)$, where 1 is $Suc(0)$.

For $k > 0$, $\mathbf{\Sigma}^0_k$-formulae are formulae of the form $\exists x_1 \forall x_2 \ldots Q x_k \varphi$, while $\mathbf{\Pi}^0_k$-formulae are those of the form $\forall x_1 \exists x_2 \ldots Q x_k \varphi$, where $\varphi$ is $\mathbf{\Delta}^0_0$ and the numerical quantifiers alternate in each of the prefixes. The union of all $\mathbf{\Pi}^0_k$- and $\mathbf{\Sigma}^0_k$-formulae for all $k \in \mathbb{N}$ is the class of *arithmetical* or *$\mathbf{\Pi}^0_\infty$-formulae.* The superscript "0" refers to the fact that there are no set quantifiers. We obtain a similar hierarchy if we allow set quantification by putting a superscript "1" and counting the number of alterations of set quantifiers over an arithmetical matrix. The $\mathbf{\Sigma}^1_k$-formulae ($\mathbf{\Pi}^1_k$-formulae) are the formulae $\exists X_1 \forall X_2 \ldots Q X_k \varphi$ (resp. $\forall X_1 \exists X_2 \ldots Q x_k \varphi$) for arithmetical $\varphi$, where the set quantifiers alternate in each of the prefixes.

For each axiom schema **Ax** we denote by (**Ax**) the theory consisting of the basic arithmetical axioms, the schema of induction **IND**, and the schema **Ax**. If we replace the schema of induction by the induction axiom, we denote the resulting theory by $(\mathbf{Ax})\restriction$.

An example for these notations is the theory $(\mathbf{\Pi}^1_1 - \mathbf{CA})$ which contains the induction schema, whereas $(\mathbf{\Pi}^1_1 - \mathbf{CA})\restriction$ contains only the induction axiom in addition to the comprehension schema for $\mathbf{\Pi}^1_1$-formulae.

In the framework of these theories one can introduce defined symbols for all primitive recursive functions. In particular, let $\langle , \rangle : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ be a primitive recursive and bijective pairing function.

The $x^{th}$ section of $U$ is defined by $U_x := \{y : \langle x, y \rangle \in U\}$. Observe that a set $U$ is uniquely determined by its sections on account of $\langle , \rangle$'s bijectivity.

Any set $R$ gives rise to a binary relation $\prec_R$ defined by $y \prec_R x := \langle y, x \rangle \in R$.

Using the latter coding, we can formulate the axiom of choice for formulae $\varphi$ in $\mathcal{C}$ by

$$\mathcal{C} - \mathbf{AC} \qquad \forall x \exists Y \varphi(x, Y) \rightarrow \exists Y \forall x \varphi(x, Y_x).$$

A special form of comprehension is $\mathbf{\Delta}^1_n$-comprehension, that is

$$\mathbf{\Delta}^1_n - \mathbf{CA} \qquad \forall u[\varphi(u) \leftrightarrow \vartheta(u)] \rightarrow \exists X \forall u(u \in X \leftrightarrow \varphi(u))$$

for all $\mathbf{\Pi}^1_n$-formula $\varphi$ and $\mathbf{\Sigma}^1_n$-formula $\vartheta$. $\mathbf{\Delta}^0_n$-comprehension is defined by requiring that $\varphi$ and $\vartheta$ are $\mathbf{\Pi}^0_n$-formulae and $\mathbf{\Sigma}^0_n$-formulae, respectively.

In set theory one has the principle of set induction which says that whenever a property propagates from the elements of any set to the set itself, then all sets have the property. In the context of second order arithmetic the equivalent of set induction is the schema of *transfinite induction*

$$\mathbf{TI} \qquad \forall X[\mathrm{WF}(\prec_X) \ \wedge \ \forall u(\forall v \prec_X u\varphi(v) \to \varphi(u)) \ \to \ \forall u\varphi(u)]$$

for all formulae $\varphi$, where $\mathrm{WF}(\prec_X)$ expresses that $\prec_X$ is well-founded, i.e., that there are no infinite descending sequences with respect to $\prec_X$. Classically $\mathrm{WF}(\prec_X)$ is equivalent to

$$\forall Y \, [\forall u(\forall v \prec_X u \, v{\in}Y \to u{\in}Y) \ \to \ \forall u \, u{\in}Y.]$$

We have now introduced the schemes for defining the preferred systems of reverse mathematics. $\mathbf{RCA}_0$ is the theory $(\mathbf{\Delta}^0_1 - \mathbf{CA})\restriction \ +\mathbf{IND}_{\mathbf{\Sigma}^0_1}$. $\mathbf{ACA}_0$ denotes the system $(\mathbf{\Pi}^0_\infty - \mathbf{CA})\restriction$ and $\mathbf{ATR}_0$ is $\mathbf{ACA}_0$ augmented by a schema which asserts that $\mathbf{\Pi}^0_1$-comprehension (or the Turing jump) may be iterated along any well-ordering. $(\mathbf{\Pi}^1_1{-}\mathbf{CA})_0$ denotes $(\mathbf{\Pi}^1_1{-}\mathbf{CA})\restriction$. The principal set existence axiom of $\mathbf{WKL_0}$ is an extension of $\mathbf{RCA}_0$ by *weak König's lemma* which asserts that any infinite tree of finite sequences of zeros and ones has an infinite path. The mathematically stronger system $\mathbf{WKL_0^+}$ was defined by Brown and Simpson (cf. [7, 8]). Let $2^{<\mathbb{N}}$ denote the set of finite sequences of zeros and ones. The axioms of $\mathbf{WKL_0^+}$ are those of $\mathbf{WKL_0}$ plus a scheme which amounts to a strong formal version of the Baire category theorem for the Cantor space $2^{\mathbb{N}}$. More formally, the additional scheme expresses that, given a sequence of dense subcollections of $2^{<\mathbb{N}}$ which is arithmetically definable from a given set, there exists an infinite sequence of zeros and ones which meets each of the given dense subcollections. The advantage of $\mathbf{WKL_0^+}$ over $\mathbf{WKL_0}$ is that $\mathbf{WKL_0^+}$ proves several important theorems of functional analysis which are apparently not provable in $\mathbf{WKL_0}$. Brown and Simpson used forcing to prove that $\mathbf{WKL_0^+}$ is still $\Pi^0_2$ conservative over $\mathbf{PRA}$.

## 2.2 How much of second order arithmetic is needed?

More precisely the foregoing question asks which part of second order arithmetic is needed for carrying out ordinary mathematics. The program of reverse mathematics as well as Feferman's investigations have amassed a large body of detailed results which allows one to draw the conclusion that a little bit goes a long way. Simpson [72] estimates "that at least 85% of existing mathematics can be formalized in $\mathbf{WKL_0}$ or $\mathbf{WKL_0^+}$ or stronger systems which are conservative over $\mathbf{PRA}$ with respect to $\Pi^0_2$ sentences." Similarly, Feferman conjectures that the overwhelming part of scientifically applicable mathematics can be formalized in systems (like $\mathbf{W}_F$) which are conservative over Peano arithmetic.

The focus of this section, however, is rather on the other end of the spectrum, where one is interested in mathematical theorems which encapsulate consistency strength beyond $\mathbf{PRA}$ and Peano arithmetic. Since 1931, the year Gödel's Incompleteness Theorems were published, logicians have been looking for a strictly mathematical example of an incompleteness in first-order Peano arithmetic, one which is mathematically simple and interesting and does not require the numerical coding of notions from logic. The first such examples were found early in 1977. The most elegant of these is a strengthening of the Finite Ramsey Theorem due to Paris and Harrington (cf. [52]). The original proofs of the independence of combinatorial statements from $\mathbf{PA}$ all used techniques from non-standard models of arithmetic. Only later on alternative proofs using proof-theoretic techniques were found, though results from ordinal-theoretic proof theory turned out to be pivotal in providing independence results for theories stronger than **PA**, and even led to a new combinatorial statement. The stronger theories referred to are Friedman's system $\mathbf{ATR}_0$ of *arithmetical transfinite recursion* and the system $(\mathbf{\Pi}^1_1 - \mathbf{CA}) \restriction$ based on $\mathbf{\Pi}^1_1$-comprehension. The independent combinatorial statements have their origin in certain embeddability questions in the theory of finite graphs. The first is a famous theorem of Kruskal asserting that every infinite set of finite trees has only finitely many minimal elements. It out that Kruskal's theorem is not provable in $\mathbf{ATR}_0$ (see [71]). An even more striking example of this independence phenomenon is provided by the *graph minor theorem*, *GMT*, of Robertson and Seymour which is one of the most important theorems of graph theory (see [11]). It was shown by Friedman, Robertson, and Seymour [24] that the *GMT* is not provable in $(\mathbf{\Pi}^1_1 - \mathbf{CA}) \restriction$. Actually, the original proof of the *GMT* for graphs of bounded tree width used Friedman's extended version of Kruskal's theorem with the "gap condition", which was gleaned from the proof-theoretic ordinal analysis of $(\mathbf{\Pi}^1_1 - \mathbf{CA}) \restriction$ and specifically designed to construct stronger incompletenesses in Peano arithmetic as part of the reverse mathematical program, so that the metamathematical considerations had a considerable mathematical spin-off.

Looking for an upper bound, it has turned out that the system $(\mathbf{\Pi}^1_1 - \mathbf{CA}) + \mathbf{TI}$ suffices for proving the graph minor theorem. Indeed, up till now no statements of everyday mathematics have been found that require more than means available in $(\mathbf{\Pi}^1_1 - \mathbf{CA}) + \mathbf{TI}$.[4]

# 3 Forms of constructivism

In the foregoing section we presented a framework sufficient for the needs of ordinary infinitistic mathematics. Finding a suitable system for formalizing constructive mathematics presents a more difficult task. Historically there have been differing "schools" of constructivism some of which are mutually incompatible. However, at least in my opinion, Martin-Löf's theory is the most thoroughly worked out system and, I think, philosophically the most convincing candidate. After recalling some history of constructivism in this section, the next section will provide an informal introduction to the ideas underlying Martin-Löf type theory.

One of the first facts that has to be taken into account is that 'constructive mathematics' is not a single, clearly defined body of mathematics. Indeed, several different brands of constructivism can be discerned:

1. Predicativism (Poincaré, Russell, Weyl, Lorenzen)
2. Intuitionism (Brouwer, Heyting, et. al.)
3. Russian constructivism (Markov, Shanin)
4. Bishop's constructive mathematics

*Predicativism* took shape in the writings of Poincaré and Russell in response to the paradoxes. Russell discerned the common underlying root for the paradoxes as follows:

> *Whatever we suppose to be the totality of propositions, statements about this totality generate new propositions which, on pain of contradiction, must lie outside the totality. It is useless to enlarge the totality, for that equally enlarges the scope of statements about the totality.* ([66], p. 224)

[4]However, there are certain mathematical statements whose proof requires the consistency strength of large cardinals. This topic will be briefly touched upon in the last section.

Thus Russel chimes in with Poincaré's anathemizing of so-called *impredicative definitions*. An impredicative definition of an object refers to a presumed totality of which the object being defined is itself to be a member. For example, to define a set of natural numbers $X$ as $X = \{n \in \mathbb{N} : \forall Y \subseteq \mathbb{N}\ F(n, Y)\}$ is impredicative since it involves the quantified variable '$Y$' ranging over arbitrary subsets of the natural numbers $\mathbb{N}$, of which the set $X$ being defined is one member. Determining whether $\forall Y \subseteq \mathbb{N}\ F(n, Y)$ holds involves an apparent circle since we shall have to know in particular whether $F(n, X)$ holds - but that cannot be settled until $X$ itself is determined. Impredicative set definitions permeate the fabric of Zermelo-Fraenkel set theory in the guise of the separation and replacement axioms as well as the powerset axiom. The avoidance of impredicative definitions has also been called the *Vicious Circle Principle*. This principle was taken very seriously by Hermann Weyl:

> *The deepest root of the trouble lies elsewhere: a field of possibilities open into infinity has been mistaken for a closed realm of things existing in themselves. As Brouwer pointed out, this is a fallacy, the Fall and Original Sin of set theory, even if no paradoxes result from it* ([79], p. 243).

In his book *Das Kontinuum*, Weyl initiated a predicative approach to the the real numbers and gave a viable account of a substantial chunk of analysis. What are the ideas and principles that his "predicative view" is grounded in? A central tenet is that there is a fundamental difference between our understanding of the concept of natural numbers and our understanding of the set concept. As the French predicativists, Weyl accepts the completed infinite system of natural numbers as a point of departure. He also accepts classical logic but just works with sets that are of level one in Russell's ramified hierarchy, in other words only with the principle of arithmetical definitions. Logicians such as Wang, Lorenzen, Schütte, and Feferman then proposed a foundation of mathematics using layered formalisms based on the idea of predicativity which ventured into higher levels of the ramified hierarchy. The idea of an autonomous progression in an ascending ladder of systems $RA_0, RA_1, \ldots, RA_\alpha, \ldots$ was first presented in Kreisel [36] and than taken up by Schütte and Feferman to determine the limits of predicativity. The notion of autonomy therein is based on introspection and should perhaps be viewed as a 'boot-strap' condition. One takes the structure of natural numbers as one's point of departure and then explores through a process of active reflection what is implicit in accepting this structure, thereby developing a growing body of ever higher layers of the ramified hierarchy.

Brouwer regarded *intuitionistic* mathematics as the only legitimate form of mathematics, viz. the only kind of mathematics which subjectively could be described as "indubitable". Mathematics, in his view, consists of mental constructions, performed by the individual in free action. In the main, mathematics is a languageless and solitary activity, where words merely accompany mathematical "constructions originating by the self-unfolding of the primordial intuition ..." ([6]). In Brouwerian intuitionism, much attention is given to the structure of the continuum. Infinite proceeding sequences and other concepts, such as choice sequences are introduced. Consideration of the nature of these sequences as perceived by the ideal mathematician leads one to the conclusion that every operation from $\mathbb{N}^{\mathbb{N}}$ to $\mathbb{N}$ is continuous. The latter is known as *Brouwer's principle* (for natural numbers), $BP_0$. The pivotal consequence of $BP_0$ is that all functions from the reals to the reals are continuous. As a result, one arrives in Brouwer's form of constructivism at theorems which contradict classical mathematics.

The concept of algorithm or recursive function is fundamental to the *Russian schools* of Markov and Shanin. Contrary to Brouwer, this school takes the viewpoint that mathematical objects must be concrete, or at least have a constructive description, as a word in an alphabet, or equivalently, as an integer, for only on such objects do recursive functions operate. Furthermore, Markov adopts what he calls *Church's thesis*, $CT$, which asserts that whenever we see a quantifier combination $\forall n \in \mathbb{N}\, \exists m \in \mathbb{N}\, A(n,m)$, we can find a recursive function $f$ which produces $m$ from $n$, i.e. $\forall n \in \mathbb{N}\, A(n, f(n))$. On the other hand, as far as pure logic is concerned he augments Brouwer's intuitionistic logic by what is known as *Markov's principle*, $MP$, which may be expressed as

$$\forall n \in \mathbb{N}\, [A(n) \vee \neg A(n)] \;\wedge\; \neg \forall n \in \mathbb{N}\, \neg A(n) \;\rightarrow\; \exists n \in \mathbb{N}\, A(n),$$

with $A$ containing natural number parameters only. The rationale for accepting $MP$ may be phrased as follows. Suppose $A$ is a predicate of natural numbers which can be decided for each number; and we also know by indirect arguments that there should be an $n$ such that $A(n)$. Then a computer with unbounded memory could be programmed to search through $\mathbb{N}$ for a number $n$ such that $A(n)$ and we should be convinced that it will eventually find one. As an example for an application of $MP$ to the reals one obtains $\forall x \in \mathbb{R}\, (\neg x \leq 0 \;\rightarrow\; x > 0)$.

In 1967 Bishop published his *Foundations of constructive analysis* [5] in which he carried out an informal development of constructive analysis which went substantially further mathematically than anything done before by constructivists. Bishop was a confirmed constructivist, as was Brouwer. However, what was novel about Bishop's work was that it could be read as a piece of classical mathematics as well. Bishop works with informal notions of constructive function and set. Among the constructivists schools, his standpoint uses the fewest assumptions in the mathematics. In fact, letting $BC$ stand for Bishop-style constructivism, one may characterize the other philosophies roughly as follows:

- Russian constructivism = $BC + MP + CT$
- Brouwerian intuitionism = $BC + BP_0 + BI$,

where $BI$ is bar induction, which is a form of transfinite induction on well-founded trees. With classical logic, $BI$ is equivalent to the principle **TI** defined in section 2.1.[5]

Several frameworks for constructivism that relate to Bishop's constructive mathematics as theories like **ZFC** relate to Cantorian set theory have been proposed by Myhill, Martin-Löf, and Feferman. Myhill (cf. [46]) developed his constructive set theory with the aim of isolating the principles underlying Bishop's conception of what sets and functions are. Moreover, he wanted "these principles to be such as to make the process of formalization completely trivial, as it is in the classical case" ([46], p. 347). Indeed, while he uses other primitives in his set theory $CST$ besides the notion of set, it can be viewed as a subsystem of **ZF**. The advantage of this is that the ideas, conventions and practise of the set theoretical presentation of ordinary mathematics can be used in the set theoretical development of constructive mathematics, too. Feferman's *Explicit Mathematics* is a theory of operations and classes ([14, 15]), $T_0$, which is suitable for representing Bishop-style constructive mathematics as well as generalized recursion, including direct expression of structural concepts which admit self-application.

The intuitionistic theory of types **MLTT** as developed by Martin-Löf (cf. [40],[42]) is also intended to be a system for formalizing intuitionistic mathematics. However, Martin-Löf probes far deeper in several respects. Not only has he developed a formal system **MLTT** but also a philosophical underpinning of constructivism which is encapsulated in his informal semantics for **MLTT**, called "meaning explanation". The latter is a systematic method of assigning meaning to the assertions of **MLTT** which enables him to justify the rules of **MLTT** by showing their validity with respect to that semantics. It

[5] This is the reason why the principle **TI** is often referred to as *bar induction*.

is perhaps not an exaggeration to say that Martin-Löf's theory gives rise to a full scale philosophy of constructivism.

Type theory is a logic free theory of constructions within which the logical notions can be defined whereas systems of explicit mathematics and constructive set theory leave the logical notions unanalyzed. For this reason I consider type theory philosophically more fundamental.[6]

Integral to an understanding of Martin-Löf's theory of types, notably his justification of the logical laws, is an awareness of the distinction between the notions of *judgement*[7] and *proposition*. The point of view he adopts is that logical operations (constants) operate on propositions whereas the logical laws (rules of inference) operate on judgements. Performing the logical operation $\vee$ leads from propositions $A$ and $B$ to a new proposition $A \vee B$. By contrast, the premises and conclusion of a logical inference are always judgements. In the order of conceptual priority the notion of judgement comes before the notion of proposition since the assertion *$A$ is a proposition* is a judgement.

The notion of proposition is a semantic notion. In a first approach, a proposition could be construed as a meaningful statement describing a state of affairs. Traditionally, a proposition is something that is either true or false. In the case of mathematical statements involving quantifiers ranging over infinite domains, however, by adopting such a view one is compelled to postulate a transcendent realm of mathematical objects which determines their meaning and truth value. Like Brouwer, Martin-Löf repudiates such an account as a myth. He explains that the meaning of a mathematical statement is not independent of our cognitive activity, that it is subject related, relative to the knowing subject in Kantian terminology. Kolmogorov observed that the laws of the intuitionisic propositional calculus become evident upon conceiving propositional variables as ranging over problems or tasks.[8] In a similar vein, Martin-Löf explains the notion of proposition as follows:

> *Returning now to the form of judgement* '$A$ is a proposition', *the semantical explanation which goes together with it is this, and here I am using the knowledge theoretical formulation, that to know a proposition, which may be replaced, if you want, by problem, expectation or intention, you must know what counts as a verification, solution, fulfillment, or realization of it. ([45], p. 34)*

In keeping with the above meaning explanation of the judgement *'$A$ is a proposition'*, to make the judgement *'$A$ is true'* you must have knowledge how to verify $A$, and for a mathematical proposition $A$ a method to verify $A$ is nothing else but a proof of $A$.

The most forceful criticism of the idea of a knowledge transcendent notion of mathematical truth has been put forward by Dummett. His arguments will be related in subsection 4.3.

# 4 Martin-Löf's theory of types

When we talk about Martin-Löf type theory we refer to more than just a formal system as the meaning explanations for the rules form an essential ingredient of it. The language

[6]Incidentally, this view was shared by Myhill: *Since completing this paper I have become familiar with some recent unpublished work of Martin-Löf, .... While the complexity of Martin-Löf's system makes extremely unlikely its general adoption as a definitive formalization of constructive mathematics, it pushes its analysis far deeper than we do, .... In any case no further work should be done in this area without familiarity with Martin-Löf's work, whose speedy publication we anticipate with pleasure. ([46], p. 355)*

[7]The word "judgement" is closely related to the German word "Urteil". Urteil was the central notion of logic in Kant's philosophy. In his transcendental logic, which forms part of the *Kritik der reinen Vernunft*, Kant arrives at his categories by discerning the various forms of judgements.

[8]He used the German word *Aufgabe*.

with which we will be concerned here is a full scale system which accounts for essentially everything one does in mathematics. The origins of this system can be traced to the natural deduction systems of Gentzen, taken in conjunction with Prawitz's reduction procedures, and to Gödel's Dialectica system. Moreover, **MLTT** is an open ended framework in that one may always add new types and rules providing they are amenable to a meaning explanation which validates them. In point of fact, a particular powerful way of going beyond an existing formalization $T$ of **MLTT** is by reflection about $T$. This is formally mirrored by the introduction of universes into **MLTT**.

Distinctive features of **MLTT** are the following:

- *meaning explanations* for rules
- the full use of the *propositions-as-types* paradigm to represent logic
- strength and expressiveness are obtained through the use of *inductive data types* and *reflection*, i.e. type universes.

These features will be addressed one by one in what follows.

## 4.1 A first glimpse of type theory

Martin-Löf's formal language has a system of rules for deriving judgements. This is in contrast to the standard formal systems which involve rules for deriving propositions. The distinction between propositions and judgements is essential for Martin-Löf. What we combine by means of the logical operations ($\rightarrow, \wedge, \vee, \neg, \forall, \exists$) and hold true are propositions. When we hold a proposition to be true, we make a judgement. The fundamental notions of type theory are introduced in the four forms of judgement:

1. $A$ is a type (abbr. $A$ *type*),
2. $A$ and $B$ are equal types (abbr. $A = B$),
3. $a$ is an element of type $A$ (abbr. $a : A$),
4. $a, b$ are equal elements of type $A$ (abbr. $a = b : A$).

There is a qualitative distinction between arbitrary elements of a type and canonical elements of a type. A type $A$ is defined by stating what we have to do in order to construct a *canonical* element of the type and what conditions need to be satisfied for such canonical elements to be equal. By contrast, an arbitrary element of type $A$ is a method or program which when executed yields a canonical element of type $A$. Two arbitrary elements $a, b$ of $A$ are equal if, when executed, $a$ and $b$ evaluate to equal canonical elements of $A$. Equality between canonical elements of the same type should be reflexive, symmetric, and transitive. For each type there are formation rules for that type and introduction rules to construct canonical elements of that type. These are best illustrated by means of a concrete example. For ease of presentation we will neglect all rules dealing with equality. The type of natural number is introduced by the following rules:

$$(\mathbb{N}\text{-formation}) \qquad \mathbb{N}\ \mathit{type}$$

$$(\mathbb{N}\text{-introduction}) \qquad 0 :\ \mathbb{N} \qquad \frac{a :\ \mathbb{N}}{Sa :\ \mathbb{N}}$$

The introduction rules state how canonical elements of $\mathbb{N}$ are generated. On the other hand, $2+2$ should be regarded as an element of type $\mathbb{N}$ as well, since we can evaluate it to a canonical element of type $\mathbb{N}$. But obviously $2+2$ is not obtainable from the given rules, viz. the judgement $2+2 : \mathbb{N}$ is not derivable via $\mathbb{N}$-introduction. Martin-Löf regards the expression, "$2+2$" as a program, which gives instructions for its own evaluation. In its evaluated form (in his theories) it will be the canonical element $SSSS0$. Therefore we shall need further rules, called *elimination rules*, which allow one to derive judgements of the form $t : A$ for non-canonical $t$. However, before stating any elimination rules let us study some further forms of types and their canonical elements, called canonical elements henceforth.

$$A \times B \qquad A \to B \qquad (\Pi x : A)B(x) \qquad A + B \qquad (\Sigma x : A)B(x).$$

A canonical element of the Cartesian product type $A \times B$ is a pair, and can be written as $(a, b)$, where $a : A$ and $b : A$. The pertaining rules are:

$$(\times\text{-introduction}) \qquad \frac{a : A \quad b : B}{(a, b) : A \times B} \qquad \frac{a = a' : A \quad b = b' : B}{(a, b) = (a', b') : A \times B}$$

A canonical element of $A \to B$ is an expression $(\lambda x)t(x)$ which (informally) denotes a function given by a term $t(x)$ such that $t(a)$, the result of substituting $a$ for $x$, is an element of $B$ for each $a : A$. In the natural deduction calculus, ($\to$-introduction) takes the forms

$$\frac{\begin{array}{c}[x : A] \\ t(x) : B\end{array}}{(\lambda x)t(x) : A \to B} \qquad \frac{\begin{array}{c}[x : A] \\ t(x) = s(x) : B\end{array}}{(\lambda x)t(x) = (\lambda x)s(x) : A \to B},$$

where $[x : A]$ indicates an assumption which gets discharged by the inference.

Suppose $A$ is a type and $B(a)$ is a type for each $a : A$. Then we can form a new type $(\Pi x : A)B$ whose canonical elements are of the form $(\lambda x)t(x)$ where $t(a)$ is an element of $B$ for each $a : A$.

The canonical elements of $A + B$ are of the forms $i(a)$ and $j(b)$, where $a : A$ and $b : B$, respectively.

If $B(a)$ is a type for each $a : A$, then $(\Sigma x : A)B(x)$ is a type whose canonical elements are pairs $(a, b)$ with $a : A$ and $b : B(a)$.

There is a certain pattern for forming canonical elements in all the above cases. Each element in canonical form is built from its components using special *constructors*. $\mathbb{N}$ has the constructors $0$ and $S$; $A \to B$ has the constructor $\lambda$; $A + B$ has the constructors $i$ and $j$; and so forth.

The dual of the introduction rules for obtaining canonical elements of a type are its *elimination rules*. The elimination rules for a type $A$ are, as it were, natural consequences of its introduction rules. They amount to saying that all canonical elements of $A$ are generated by exactly the means laid down in the introduction rules, viz. there are no other ways to form canonical elements. Martin-Löf is expanding here on Gentzen's ideas of dividing logical rules into introduction and elimination rules. Referring to the logical constants, Gentzen [25] explains the harmony between these two kinds of rules as follows: *"an introduction rule gives, as it were, a definition of the constant in question"* while *"an elimination rule is only a consequence of the corresponding introduction rule, which may be expressed somewhat as follows: when making an inference by an elimination rule, we are allowed to 'use' only what the principal sign of the major premiss 'means' according to the introduction rule for this sign."*

In the case of the type of natural numbers $\mathbb{N}$ the elimination rules amount to familiar rules for structural induction and recursion over $\mathbb{N}$. The formal rendering of elimination rules for an arbitrary type requires a new constant. Associated with each type will be a *selector* (dual to the notion of a constructor, above), given as an implicitly defined constant, whose defining equations express the elimination rules for the type. As the elimination rules for $\mathbb{N}$ are rather involved (but see section 4.4 below) let us study the much simpler case of the type $A \to B$. Here the rules are:[9]

$$(\to\text{-elimination}) \qquad \frac{c : A \to B \quad a : A}{\mathbf{App}(c, a) : B}$$

$$\frac{c = d : A \to B \quad a = b : A}{\mathbf{App}(c, a) = \mathbf{App}(d, b) : B}$$

$$\frac{\begin{array}{c}[x : A] \\ t(x) : B\end{array} \quad a : A}{\mathbf{App}((\lambda x)t(x), a) = t(a) : B}$$

$$\frac{c : A \to B}{c = (\lambda x)\mathbf{App}(c, x) : A \to B}$$

To explain the meaning of the selector **App** suppose $c : A \to B$ and $a : A$. Then $\mathbf{App}(c, a)$ is a method for obtaining a canonical element of $B$ which is executed as follows. $c$ is a method which yields a canonical element $(\lambda x)t(x)$ of $A \to B$. Substituting $a$ for $x$ leads to $t(a) : B$. Thus, through evaluating $t(a)$ we finally arrive at a canonical element of $B$.

## 4.2 The proposition-as-types interpretation

So far we have avoided the question how standard logical operations are to be treated in type theory. It turns out that the first glimpse of type theory given above suffices for that task. The point of view to be adhered to is that propositions are types and that logical operations on propositions correspond to the appropriate type forming operations in line with the *Brouwer-Heyting-Kolmogorov interpretation* (BHK-interpretation) of the constructive meaning of the logical constants, whereby each proposition is identified with the type of its proofs.[10] For example the constructive meaning of an implication $A \supset B$ consists in saying that a proof of $A \supset B$ is a constructive procedure that transforms each proof of $A$ into a proof of $B$. It seems that the notion of constructive procedure used here must be taken as a primitive notion. The table below gives a dictionary for translating logical operations into type theory.

| **Logical notion** | **Type-theoretic notion** |
| --- | --- |
| proposition | type |
| proof of $A$ | element of $A$ |
| $A$ is true | $A$ has an element |
| $A \wedge B$ | $A \times B$ |

[9]Martin-Löf actually breaks up the rules which we collectively call elimination rules, into two groups dubbed elimination and equality rules, respectively. His elimination rules explain how non-canonical elements are formed via the selector, and the equation rules explain how the selector operates on the canonical elements.

[10]On the formal level, the analogy between propositions and types was discovered by Curry and Feys [10] and further developed by Howard [34].

| | |
|---|---|
| $A \supset B$ | $A \to B$ |
| $A \vee B$ | $A + B$ |
| $(\forall x{\in}A)B(x)$ | $(\Pi x:\ A)B(x)$ |
| $(\exists x{\in}A)B(x)$ | $(\Sigma x:\ A)B(x)$ |

Note that the treatment of quantifiers in **MLTT** strictly adheres to Brouwer's dictum that quantified variables should range only over already-defined sets/types $A$.

In rendering propositions as types, the elements of a proposition are to be understood as proofs and thus its canonical elements could be termed *canonical proofs*. The inductive nature of the canonical proofs can be made more explicit by the following table:

| **a canonical proof of** | **has the form** |
|---|---|
| $A \wedge B$ | $(a, b)$, where $a$ is a proof of $A$ and $b$ is a proof of $B$ |
| $A \supset B$ | $(\lambda x)t(x)$, where $t(a)$ is a proof of $B$ whenever $a$ is a proof of $A$ |
| $A \vee B$ | $i(a)$, where $a$ is a proof of $A$, or $j(b)$, where $b$ is a proof of $B$ |
| $(\forall x{\in}A)B(x)$ | $(\lambda x)t(x)$, where $t(a)$ is a proof of $B(a)$ whenever $a$ is an element of $A$ |
| $(\exists x{\in}A)B(x)$ | $(a, b)$, where $a$ is an element of $A$ and $b$ is a proof of $B(a)$ |

### 4.3 The Dummett-Prawitz meaning-as-use theory and MLTT

Inspired by ideas of Wittgenstein, Dummett (cf. [12, 13]) has brought forward philosophical arguments against a platonistic theory of meaning, amounting to a rejection of classical logic in favour of intuitionistic logic. In order for such a rejection to be conclusive, Prawitz [55] has expanded on Dummett's ideas who suggested that a theory of meaning formulated in terms of proofs or rules also ought to take note of Gentzen's fundamental insight that a correspondence or duality obtains between the rules for asserting a sentence and the rules for drawing consequences from it. In many aspects the Dummett-Prawitz theory of meaning-as-use is closely related to Martin-Löf's understanding of the meaning of the logical operators and thus I consider it useful to intersperse a brief account of their theory here. On the one hand this might be instructive and illuminating for an understanding of certain aspects of Martin-Löf's meaning explanations. On the other hand, it seems that **MLTT** overcomes a deficiency of the Dummett-Prawitz semantics which is due to the latter's rather narrow focus on logical reasoning. In mathematics, logic retreats into the background and mathematical objects and constructions occupy center stage. The interrelationship between logical inferences and mathematical constructions connects together logic and mathematics. Logic gets intertwined with mathematical objects and operations, and it appears that its role therein cannot be separated from mathematical constructions. Indeed, not only does Martin-Löf's work demonstrate that a coherent theory of meaning can be developed along the lines of Dummett-Prawitz for a rich and elaborate part of mathematics but also that logical operators can be construed as special cases of more general mathematical operations.

Below I shall not undertake any detailed analysis of all the facets of of Dummett's and Prawitz's arguments and confine myself to a rough sketch of the main points.

### 4.3.1 The meaning-as-use thesis

The classical/platonist mathematician holds that the meaning of a mathematical statement is determined by its truth conditions in the (abstract) realm of mathematical objects existing independently of us. This view, however, appears to be very detached from mathematical practice. If we look back at the way that mathematics is actually taught, it seems that what we learn is not to establish truth-conditions of sentences in a transcendent world but rather what is to be counted as establishing the truth of sentences, that is to say, we are trained in the art of proving mathematical sentences. And it therefore seems that the invocation of this transcendence serves merely as an *Überbau* which fails to add anything to the elucidation of what it is to know a mathematical statement.

Dummett argues that meaning cannot be separated from use and must be recognizable by us. His tenet is that the meaning of a sentence must be fully manifest in its use, where use is to be taken in a very broad sense covering all its aspects. Use exhaustively determines meaning in the sense that two expressions which are always used in the same way ought to have the same meaning. The main lines of support for Dummett's thesis are the following:

1. *"... that meaning has to be communicable and that communication has to be observable: to assume that there is some ingredient in the meaning of a sentence which cannot become manifest in the use made of it is to assume that part of the meaning cannot be communicated. This part of the meaning would then be irrelevant when the statement was used in communication. Such a meaning would therefore be of no importance for mathematics understood as a social enterprise in which many people can cooperate and exchange results with each other."* ([55], p. 4)

2. When we learn a language in a primordial way, i.e. not by translation into another language, all we learn is how to use expressions correctly. The grasp of the meaning of a sentence in such a language must therefore consists in our ability to use it correctly.

3. Knowledge of the meaning of an expression can sometimes be demonstrated in the manner of Socrates by explicitly defining it in terms of other expressions of which the meaning is already known. However, if there is not to be an infinite regress, the meaning of a statement cannot consist solely of explicit verbalizable knowledge and thus knowledge of meaning must ultimately be traced back to implicit knowledge. Implicit knowledge of the meaning of a sentence $A$ can only manifest itself in the ability to use it or to respond to its use by others in a certain observable way. Thus the only way of acquiring such knowledge is by observing and learning its use. Therefore the meaning of a sentence cannot transcend its total use. In short: We know the meaning of $A$ if we know under what conditions $A$ may be correctly asserted.

Specialized to the case of mathematics, the above views on the meaning of a sentence amount to agreeing with the intuitionists that meaning in mathematics has somehow to be understood in terms of proofs.

### 4.3.2 Molecular versus holistic views on meaning

The principle that use determines meaning does not rule out the possibility that the meaning of a single statement can only be understood with regard to the framework of language as a whole. An extreme holistic view claims "that nothing less than the total use of the language determines the meaning of an individual sentence." ([55], p. 7) On such a view, it is not possible to develop a meaning theory which explains the meaning of single sentences in terms of their constituent parts or even to adhere to a milder form of holism where one singles out a privileged class of sentences which are amenable to such a molecular meaning explanation and which then gives meaning to other sentences by their deductive relationships with this class. Dummett rejects the drastic form of holism on the grounds that meaning must be recognizable by us through use in particular situations and 'total use of language' blatantly escapes recognizability.

A prominent example of a restricted form of holism in mathematics is, of course, underlying Hilbert's program, where the privileged sentences are called 'real sentences'. Partial holism is also a common standpoint in the philosophy of science. Here one distinguishes between theoretical and observational sentences. The latter are, by and large, endowed with a meaning in isolation whereas the meaning of the former is understood as determined by their role in deducing and refuting privileged observational sentences.

Dummett, however, repudiates all kinds of holistic meaning theories and maintains a molecular view. Following Frege, he sees the meaning of a sentence as being determined by the way it is built up from its constituent parts. Thus a sentence carries an individual meaning which is generated from its atomic components (having immediate meaning) via logical operations.

### 4.3.3 Meaning-as-use and the BHK-interpretation

The meaning-as-use thesis asserts that the meaning of a statement is determined by the conditions under which it may be correctly asserted. Turning to a mathematical statement $A$, its meaning could be explained by what it means to give a proof of $A$ or what counts as a proof of $A$. The BHK-interpretation mentioned above provides a precise rendering of what counts as a proof of a mathematical proposition. It also adheres to the molecular semantical view in that the proofs of a complex proposition are explained in terms of proof conditions for its components. To distinguish the notion of proof supported by the BHK-interpretation from others, let us call them *canonical proofs* or *direct proofs.* A first attempt at a meaning theory for mathematical statements $A$ could be framed as follows:

($MT_1$) To know the meaning of $A$ is to know the conditions for asserting $A$.

The condition for asserting $A$ is to know a canonical proof of $A$.

Under the proposition-as-types view, the canonical proofs of a proposition are precisely the ones generated by the introduction rules for the pertaining type. To give an example, a direct proof $(p_A, p_B)$ of a proposition $A \wedge B$ is obtained from a direct proof $p_A$ of $A$ and a direct proof $p_B$ of $B$. In the case of a conditional $A \supset B$ a direct proof consists of a construction $f$ which when fed a direct proof of $A$ returns a direct proof of $B$. However, the notion of a direct proof is too restrictive. Even from an intuitionistic point of view there are perfectly legitimate proofs which cannot be obtained in this direct way. For instance, in arithmetic we frequently infer a statement $F(n, m)$ for large numbers $n, m$ from a proof of the universal statement $(\forall x{\in}\mathbb{N})(\forall y{\in}\mathbb{N})F(x, y)$ by instantiation. Or, taking an example from [55], p. 21, "we may assert even intuitionistically that $A(n) \vee B(n)$ for some numeral $n$ without knowing a proof of $A(n)$ or $B(n)$; it would be sufficient, e.g., if we know a proof of $A(0) \vee B(0)$ and a proof of $(\forall x{\in}\mathbb{N})([A(x) \vee B(x)] \rightarrow [A(x+1) \vee B(x+1)])$." Thus from an intuitionistic point of view it is sufficient that we know a procedure to obtain a direct proof but it is not necessary that such a direct proof be actually constructed. As a result, the problems met with ($MT_1$) can be remedied by saying:

($MT_2$) To know the meaning of $A$ is to know the conditions for asserting $A$.

The condition for asserting $A$ is that we either know a canonical proof of $A$ *or* know a procedure for obtaining such a proof.

### 4.3.4 On the harmony between the rules for making assertions and drawing consequences

Dummett points out that the view expressed in ($MT_2$) may be somewhat immature in that it neglects an important aspect of the use of an assertive sentence which manifests itself in the commitments made by asserting it. In the case of mathematics, the full grasp of the meaning of a statement not only comprises the ability to recognize a proof of it when one is presented to us but also our capacity to draw consequences from it. He therefore suggests that in a theory of meaning a harmony should obtain between the rules for asserting a sentence and the rules for drawing consequences from it. Thus, in addition to the molecular view, there is an important strand of ideas, originating with Gentzen, running through Dummett's theory of meaning. Gentzen's elimination rules are rules for inferring consequences from a sentence and run parallel to the introduction rules. Hence, taking elimination rules into account we arrive at an enriched concept of meaning of the following form:

($MT_3$) To know the meaning of $A$ is to know the conditions for asserting $A$.

The condition for asserting $A$ is that we either know a canonical proof of $A$ *or* know a procedure for obtaining such a proof, and, moreover, that we know the rules for drawing consequences from $A$.

($M_3$) doesn't specify the exact nature of the rules for drawing consequences from $A$. In a sense, there are many ways of inferring further sentences from a given one. Nonetheless, the elimination rules pertaining to $A$ seem to occupy a privileged position in that they flow naturally from the corresponding introduction rules. Such a view seems to prevail in **MLTT** where this idea is developed at full scale. Prawitz embraces the notion of meaning as developed in ($M_3$) as the most desirable candidate but also enters the caveat that "in order for such a meaning theory to appear reasonable, it seems that one has to argue for the unique position of the elimination rules." ([55], p. 33) He conjectures that what distinguishes the elimination rules from other possible candidates might be that they constitute in some sense the strongest rules for drawing consequences from an assertion.

To my mind the question of the position of the elimination rules is best discussed in a wider context where one views propositions as types and in addition to the type forming operations pertaining to the BHK-interpretation takes also mathematically important types into account. These so-called *inductive data types* will be discussed in the next subsection. In the case of inductive data types, the elimination rules amount to the central tool for performing mathematical operations on them. They enable one to prove assertions about all the elements of the type by structural induction and at the same time allow one to define functions on the elements of the type by structural recursion. Owing to in-depths proof-theoretic research it is known that the omission of elimination rules in formal systems featuring inductive types gives rise to theories of weak proof-theoretic strength whereas the inclusion of elimination rules results in theories of considerable proof-theoretic strength. The latter clearly corroborates the view that elimination rules are pivotal in determining the meaning of statements involving mathematical types as e.g. the natural numbers.

### 4.4 Inductive data types

Thus far I have only addressed a small fragment of **MLTT** which is sufficient for developing logic. The concept of an *inductive data type* is central to Martin-Löf's constructivism. It was very succinctly stated by Gödel in a handwritten text for an invited lecture which he delivered in 1933. Therein he described constructive mathematics by the following two characteristics:

> *1. The application of the notion "all" or "any" is to be restricted to those infinite totalities for which we can give a* **finite procedure for generating all their elements** *(as we can, e.g., for the totality of integers by the process of forming the next greater integer and as we cannot, e.g., for the totality of all properties of integers).*
>
> *2. Negation must not be applied to propositions stating that something holds for all elements, because this would give existence propositions. Or to be more exact: Negatives of general propositions (i.e., existence propositions) are to have meaning in our system only in the sense that we have found an example but, for the sake of brevity, do not state it explicitly. I.e., they serve merely as an abbreviation and could be entirely dispensed with if we wished.*
>
> *From the fact that we have discarded the notion of existence and the logical rules concerning it, it follows that we are left with essentially only one method for proving general propositions, namely,* **complete induction applied to the generating process of our elements.** [...] *and so we may say that our system is based exclusively on the* **method of complete induction in its definitions as well as its proofs.** *([27], p. 51)*

The paradigm of an inductive data type is that of the type of natural numbers. The elimination rules for ℕ correspond to the principles of induction and recursion over ℕ. Likewise, for an arbitrary inductive type $A$ the elimination rules encapsulate the principles of structural induction and recursion over $A$. The induction principle for $A$ tells one how to prove properties for all elements of a type by induction on the build-up of its canonical elements and, correspondingly, the recursion principle tells one how to define a function on all canonical elements of the type by recursion on their build-up. More formally, elimination rules explain how non-canonical elements are formed via the selector, and the equation rules explain how the selector operates on the canonical elements. The selector pertaining to the type ℕ is $\mathbf{R}_{\mathbb{N}}$. The elimination rules is the following:

$$(\mathbb{N}\text{-elimination}) \qquad \frac{c : \mathbb{N} \qquad d : C(0) \qquad \begin{matrix}[x : \mathbb{N},\ y : C(x)] \\ e(x,y) : C(Sx)\end{matrix}}{\mathbf{R}_{\mathbb{N}}(c, d, (x,y)e(x,y)) : C(c)}$$

$\mathbf{R}_{\mathbb{N}}(c, d, (x,y)e(x,y))$ is explained as follows: first execute $c$, getting a canonical element of ℕ, which is either 0 or $Sa$ for some $a : \mathbb{N}$. In the first case, continue by executing $d$, which yields a canonical element $f : C(0)$; but, since $c = 0 : \mathbb{N}$ in this case, $f$ is also a canonical element of $C(c) = C(0)$. In the second case, substitute $a$ for $x$ and $\mathbf{R}_{\mathbb{N}}(a, d, (x,y)e(x,y))$ (namely, the preceding value) for $y$ in $e(x,y)$ so as to get $e(a, \mathbf{R}_{\mathbb{N}}(a, d, (x,y)e(x,y))$. Executing it, we get a canonical $f$ which, by the right premiss, is in $C(Sa)$ by the first case. Otherwise, continue as in the second case, until we eventually reach the value 0. This explanation of the above elimination rule also explains the equality rules.

$$(\mathbb{N}\text{-equality I}) \qquad \frac{d : C(0) \qquad \begin{matrix}[x : A,\ y : C(x)] \\ e(x,y) : C(Sx)\end{matrix}}{\mathbf{R}_{\mathbb{N}}(0, d, (x,y)e(x,y)) = d : C(c)}$$

$$(\mathbb{N}\text{-equality II}) \qquad \frac{a : \mathbb{N} \qquad d : C(0) \qquad \begin{array}{c}[x : \mathbb{N},\ y : C(x)] \\ e(x,y) : C(Sx)\end{array}}{\mathbf{R}_{\mathbb{N}}(Sa, d, (x,y)e(x,y)) = e(a, \mathbf{R}_{\mathbb{N}}(a, d, (x,y)e(x,y))) : C(Sa)}$$

An infinitary example of an inductive definition in Martin-Löf type theory is the $W$-type built over a family of types. A special case of the latter is the type of Brouwer ordinals or well-founded trees over **N** which may be taken as a constructive version of the second number class of ordinals. The inductive generation proceeds by the following introduction rules:

$$\bar{0} : \mathcal{O} \qquad \frac{a : \mathcal{O}}{a' : \mathcal{O}} \qquad \frac{f : \mathbb{N} \to \mathcal{O}}{\sup(f) : \mathcal{O}}$$

An inductive definition often involves the characterization of a collection of objects as the smallest collection satisfying certain closure conditions. In classical set theory the inductively defined set is usually obtained as the intersection of all collections that satisfy the closure conditions. Such an explicit definition is thoroughly impredicative in that the collection is defined using quantification over all collections. It would seem therefore that if one wants to make sense of infinitary inductive definitions it is necessary at some point to make use of some impredicative definitions. This is certainly the way that inductive definitions are accounted for in classical set theory. But there is something unsatisfying about this conclusion. Many inductive definitions can be intuitively understood directly in their own terms and impredicative definitions are only required in order to represent them within a particular framework such as classical set theory.

## 4.5 Reflection via universes

The openendedness of Martin-Löf type theory is particularly manifest in the introduction of so-called *universes*. Type universes encapsulate the informal notion of reflection whose role may be explained as follows. During the course of developing a particular formalization of type theory, the type theorist may look back over the rules for types, say $\mathcal{C}$, which have been introduced hitherto and perform the step of recognizing that they are valid according to Martin-Löf's informal semantics of meaning explanation. This act of 'introspection' is an attempt to become aware of the conceptions which have governed our constructions in the past. It gives rise to a *"***reflection principle** *which roughly speaking says whatever we are used to doing with types can be done inside a universe"* ([40], p. 83). On the formal level, this leads to an extension of the existing formalization of type theory in that the type forming capacities of $\mathcal{C}$ become enshrined in a type universe $\mathbf{U}_{\mathcal{C}}$ mirroring $\mathcal{C}$. If, e.g., $\mathcal{C}$ consists of the type forming operations $\mathbb{N}, +, \Pi$ and their rules, then this gives rise to the rules:

$$(\mathbf{U}_{\mathcal{C}}\text{-formation}) \qquad \mathbf{U}_{\mathcal{C}} : \mathit{type} \qquad \frac{a : \mathbf{U}_{\mathcal{C}}}{\mathbf{T}_{\mathcal{C}}(a) : \mathit{type}}$$

$$(\mathbf{U}_{\mathcal{C}}\text{-introduction}) \qquad \hat{\mathbb{N}} : \mathbf{U}_{\mathcal{C}} \qquad \mathbf{T}_{\mathcal{C}}(\hat{\mathbb{N}}) = \mathbb{N}$$

$$\frac{a : \mathbf{U}_{\mathcal{C}} \quad b : \mathbf{U}_{\mathcal{C}}}{a \hat{+} b : \mathbf{U}_{\mathcal{C}}} \qquad \frac{a : \mathbf{U}_{\mathcal{C}} \quad b : \mathbf{U}_{\mathcal{C}}}{\mathbf{T}_{\mathcal{C}}(a \hat{+} b) = \mathbf{T}_{\mathcal{C}}(a) + \mathbf{T}_{\mathcal{C}}(b)}$$

$$\frac{a : \mathbf{U}_{\mathcal{C}} \qquad \begin{array}{c}[x : \mathbf{T}_{\mathcal{C}}(a)] \\ t(x) : \mathbf{U}_{\mathcal{C}}\end{array}}{\hat{\Pi}(a, (\lambda x)t(x)) : \mathbf{U}_{\mathcal{C}}} \qquad \frac{a : \mathbf{U}_{\mathcal{C}} \qquad \begin{array}{c}[x : \mathbf{T}_{\mathcal{C}}(a)] \\ t(x) : \mathbf{U}_{\mathcal{C}}\end{array}}{\mathbf{T}_{\mathcal{C}}(\hat{\Pi}(a, (\lambda x)t(x))) = (\Pi x : \mathbf{T}_{\mathcal{C}}(a))\mathbf{T}_{\mathcal{C}}(t(x))}$$

$x : \mathbf{U}_\mathcal{C}$ and $\mathbf{T}_\mathcal{C}(x)$ are defined by a simultaneous induction. The elements of type $\mathbf{U}_\mathcal{C}$ are codes or Gödel numbers of types generated from $\mathbb{N}$ via the type forming operation $+$ and $\Pi$. Each time an element of $\mathbf{U}_\mathcal{C}$ is generated there is a declaration, by means of the decoding function $\mathbf{T}_\mathcal{C}$, of the type for which it stands.

It might seem that type universes are not that different from inductive data types. However, as opposed to inductive data types their elements are not simply defined by a positive inductive definition. They incorporate the extra feature of being equipped with a type-valued function defined on them, and, moreover, this simultaneously defined decoding function is allowed to occur negatively in its introduction rules as, e.g., in those for $\hat{\Pi}$.

As in the case of inductive data types, the introduction rules for type universes give rise to their dual, the elimination rules, which entail that the elements of $\mathbf{U}_\mathcal{C}$ are exactly generated by the forms of type that $\mathbf{U}_\mathcal{C}$ is reflecting.[11]

### 4.6 Reflection in higher order universes

It is the nature of reflection to aim at higher degrees of introspection. The first level is encapsulated in universes containing codes for types which reflect previously accepted type forming operations. In [40, 42] Martin-Löf considered an infinite, externally indexed tower of universes $\mathbf{U}_0 \in \mathbf{U}_1 \in \ldots \in \mathbf{U}_n \in \ldots$ all of which are closed under the same standard ensemble of set forming operations. The next natural step was to implement a *universe operator* into type theory which takes a family of sets and constructs a universe above it. Such a universe operator was formalized by Palmgren while working on a domain-theoretic interpretation of the logical framework with an infinite sequence of universes (cf. [50]). Aiming at extensions of type theory with more powerful axioms, Martin-Löf then suggested finding axioms for a universe $\mathbb{V}$ which itself is closed under the universe operator. The type-theoretic formalization of the pertinent rules is due to Palmgren [51], where the universe was referred to as a *superuniverse* for intuitionistic type theory.

After the superuniverse one can go to a new level of abstraction which consists in gaining insight into how type forming operations are obtained. One can, e.g., internalize the introduction of new universe operators in a formal system of type theory wherein each operator $F : \mathbf{Fam} \to \mathbf{Fam}$ from families of types to families of types gives rise to a new universe operator $\mathbb{S}(F) : \mathbf{Fam} \to \mathbf{Fam}$. When applied to a family of sets, $\mathbb{S}(F)$ produces a universe above it which, in addition to the standard set constructors, is closed under $F$. Such a system, denoted **MLF**, was presented in [62].

A further level of abstraction gives rise to types having codes for such operations as elements which may be called universes of type two. Such a step has, e.g., been taken in [60] which introduces a formal type theory **MLQ**. Central to **MLQ** are a universe **M** and a universe of codes for type constructors **Q** which are defined simultaneously. In more generality, this idea has been developed by Palmgren (cf. [51]) who introduced type universes of arbitrary finite levels.

Essentially stronger universe constructions require new ideas and their potential seems to be limited by creativity only. However, in [62] I have stated a conjecture about the proof-theoretic strength of universe constructions based on higher types which predicts that they can always be mimicked in a classical set theory called **KPM**. **KPM** was designed to formalize segments $L_\mu$ of the constructible hierarchy $L$, where $\mu$ is a recursively Mahlo ordinal (cf. [56]).

[11] Martin-Löf has expressed the view that it is natural to keep the type $\mathbf{U}_\mathcal{C}$ open to reflect any additional forms of type that can arise in the future and therefore not to incorporate elimination rules for universes. In my view such a decision is a mere convenience lest one be forced to specify different forms of universes.

Another important step in advancing constructive type theory has been taken by Setzer who gave a formalization of a Mahlo universe in type theory (cf. [69]), dubbed **TTM**. Constructive Zermelo-Fraenkel set theory conjoined with an axiom asserting the existence of a Mahlo set can be interpreted in **TTM** as has been shown in [64]. Setzer's theory is stronger than those based on higher type universes. It provides an important step for expanding the realm of Martin-Löf type theory. The difference between **TTM** and the systems above is that **TTM** introduces a new construction principle which is not foreshadowed in Martin-Löf's original papers. This is witnessed by the fact that models for Setzer's Mahlo universe are generated by a non-monotonic inductive definition (see section 6) and, furthermore, by an observation due to Palmgren which shows it to be incompatible with elimination rules for the universe. In a sense, **TTM** means a paradigm shift to a *new* Martin-Löf type theory in that the rules for forming the elements of a type are no longer required to be monotonic. Palmgren's proof that the usual elimination rules for universes yield an inconsistency when applied to **TTM** prompted Martin-Löf to respond that universes need not be equipped with elimination rules and that such rules may be rather alien to the idea of a universe. From the point of view of the classical theory of inductive definitions (cf. [33]) though, Palmgren's inconsistency proof presents no surprise. The usual elimination rules for universes are tailored for monotone inductive definitions and thus spring from the crucial property that the inductively defined type (or set) is the least fixed point of the corresponding operator. To be more precise, in set theory a *monotone inductive definition* over a given set $A$ is derived from a mapping

$$\Phi : \mathbf{pow}(A) \to \mathbf{pow}(A)$$

that is monotone, i.e., $\Phi(X) \subseteq \Phi(Y)$ whenever $X \subseteq Y \subseteq A$. Here $\mathbf{pow}(A)$ denotes the class of all subsets of $A$. The set inductively defined by $\Phi$, $\Phi^\infty$, is the smallest set $Z$ such that $\Phi(Z) \subseteq Z$. Due to the monotonicity of $\Phi$ such a set exists. In the case of Setzer's type **TTM**, however, the pertinent operator is non-monotone and does not possess a least fixed point, hence the inconsistency arises. The classical view on non-monotone inductive definitions is that the inductively defined set is obtained in stages by iteratively applying the corresponding operator to what has been generated at previous stages along the ordinals until no new objects are generated in this way. More precisely, if $\Phi : \mathbf{pow}(A) \to \mathbf{pow}(A)$ is an arbitrary mapping then the set-theoretic definition of the set inductively defined by $\Phi$ is given by

$$\begin{aligned} \Phi^\infty &:= \bigcup_\alpha \Phi^\alpha, \\ \Phi^\alpha &:= \Phi(\bigcup_{\beta<\alpha} \Phi^\beta) \cup \bigcup_{\beta<\alpha} \Phi^\beta, \end{aligned}$$

where $\alpha$ ranges over the ordinals. I agree with Setzer (cf. [69], p. 157) that the correct type-theoretic elimination rules for non-monotone type universes have yet to be found. Furthermore, I'd like to conjecture that the "correct" elimination rules can be found by viewing non-monotone inductive types as being equipped with a $W$-type (or well-ordering) which provides the stages of its inductive generation.

# 5 Which fragments of second order arithmetic are secured by MLTT?

The central notions of *proof-theoretic reducibility* and *proof-theoretic strength* will be used below. For the readers convenience I shall insert a brief account of them. All theories

$T$ considered in the following are assumed to contain a modicum of arithmetic. For definiteness let this mean that the system **PRA** of Primitive Recursive Arithmetic is contained in $T$, either directly or by translation.

**Definition 5.1** Let $T_1$, $T_2$ be a pair of theories with languages $\mathcal{L}_1$ and $\mathcal{L}_2$, respectively, and let $\Phi$ be a (primitive recursive) collection of formulae common to both languages. Furthermore, $\Phi$ should contain the closed equations of the language of **PRA**.

We then say that $T_1$ is *proof-theoretically $\Phi$-reducible to* $T_2$, written $T_1 \leq_\Phi T_2$, if there exists a primitive recursive function $f$ such that

$$\mathbf{PRA} \vdash \forall \phi \in \Phi \, \forall x \, [\mathsf{Proof}_{T_1}(x, \phi) \; \rightarrow \; \mathsf{Proof}_{T_2}(f(x), \phi)]. \tag{1}$$

$T_1$ and $T_2$ are said to be *proof-theoretically $\Phi$-equivalent*, written $T_1 \equiv_\Phi T_2$, if $T_1 \leq_\Phi T_2$ and $T_2 \leq_\Phi T_1$.

The appropriate class $\Phi$ is revealed in the process of reduction itself, so that in the statement of theorems we simply say that $T_1$ is *proof-theoretically reducible* to $T_2$ (written $T_1 \leq T_2$) and $T_1$ and $T_2$ are *proof-theoretically equivalent* (written $T_1 \equiv T_2$), respectively. Alternatively, we shall say that $T_1$ and $T_2$ have the *same proof-theoretic strength* when $T_1 \equiv T_2$.

Proof-theoretic investigations undertaken by Buchholz, Pohlers and Sieg in the late 1970s (see [9]) showed that the intuitionistic theory of below $\epsilon_0$-iterated inductive tree classes, $\mathbf{ID}^i_{<\epsilon_0}(\mathcal{O})$, is of the same proof-theoretic strength as $(\mathbf{\Sigma}^1_2\textbf{-AC})$. With the help of one type universe reflecting the $W$-type one can interpret $\mathbf{ID}^i_{<\epsilon_0}(\mathcal{O})$ in type theory thereby reducing $(\mathbf{\Sigma}^1_2\textbf{-AC})$ to type theory. However, a considerably stronger theory than $(\mathbf{\Sigma}^1_2\textbf{-AC})$ can be proven consistent in **MLTT**. Jäger and Pohlers [35]) gave an ordinal analysis $(\Sigma^1_2\textbf{-AC})$ plus bar induction in terms of an ordinal representation system. Utilizing the pertaining ordinal representation system the following result was obtained independently by Rathjen [57] and Setzer [67]:

**Theorem 5.2** *The consistency of* $(\Sigma^1_2\textbf{-AC})$+**TI** *is provable in Martin-Löf's 1984 type theory.*

On the part of the intuitionists/constructivists, the following objection could be raised against the significance of consistency proofs: even if it had been constructively demonstrated that the classical theory $T$ cannot lead to mutually contradictory results, the theorems of $T$ would nevertheless be propositions without sense and their investigation therefore an idle pastime.[12] Well, it turns out that the constructive well-ordering proof of the representation system used in the analysis of $(\Sigma^1_2\textbf{-AC})$+**TI** yields more than the mere consistency of the latter system. For the important class of $\Pi^0_2$ statements one obtains a conservativity result.

**Theorem 5.3** (Rathjen [57]; Setzer [68])

- *The soundness of the negative arithmetic fragment of* $(\Sigma^1_2\textbf{-AC})$+**TI** *is provable in Martin-Löf's 1984 type theory.*
- *Every* $\Pi^0_2$ *statement provable in* $(\Sigma^1_2\textbf{-AC})$+**TI** *has a proof in Martin-Löf's 1984 type theory.*

[12] This is reminiscent of Russell's bon mot that not everything in an inconsistent theory can be true but that every axiom of a consistent theory may be false.

Theorem 5.2 is by no means the strongest result of its kind. The paper [59] introduces a theory of second order arithmetic which is based on $(\Sigma^1_2\text{-}\mathbf{AC})\restriction$ but in addition has axioms stating that there exist many $\beta$-models of $(\Sigma^1_2\text{-}\mathbf{AC})$. A $\beta$-model is a model with respect to which the notion of well-foundedness is absolute. More precisely, let $\mathbf{T}$ be the theory $(\Sigma^1_2\text{-}\mathbf{AC})\restriction$ plus the scheme asserting that every true $\Pi^1_3$ statement is reflected in a $\beta$-model of $(\Sigma^1_2\text{-}\mathbf{AC})$. It follows from [59], Theorem 5.15 that the type theory **MLF** and $\mathbf{T}$ have the same proof-theoretic strength. It also follows from [59], Theorem 5.15 that the type theory **MLQ** proves the consistency and $\Pi^0_2$ soundness of $T$.

# 6 The limits of Martin-Löf type theory

Above we have seen that ordinary mathematics is demonstrably consistent relative to **MLTT**. Thus from the point of view of justifying mathematical practice in a Hilbertian way the existing formalizations of **MLTT** are already powerful enough. Notwithstanding it is still of great interest to ponder where the limits of **MLTT** lie. This is a somewhat ambiguous question, though. It is, of course, fundamental to Martin-Löf's outlook that it is possible to apply reflection to any particular formalization of **MLTT**, and thus obtain a stronger system. Consequently, Martin-Löf type theory as a foundational undertaking cannot be captured by a formal system that the type theorist can recognize as being sound. Moreover, it doesn't seem to be possible to single out a particular fragment $T_M$ of **ZFC** and argue that **MLTT** exhausts and furthermore is equiconsistent with $T_M$ since drawing such an exact limit to **MLTT** appears to entail that this limit could be approached from within Martin-Löf type theory and therefore be overcome.

Nonetheless, it appears to be possible to transcend the type theorists realm by adopting, as it were, a hypothetical 'eagle's' point of view as in classical set theory and reflect from such an advanced position on all possible moves the type theorist can ever perform. This thought experiment should then allow one to delineate bounds for the realm which the type theorist will never be leaving, or, more precisely, it should enable the Cantorian set theorist to draw a line in his set-theoretic world beyond which the type theorist will never be able to reach. An intellectual difficulty in pursuing this project is to become aware of the conceptions which govern all possible thinking of a Martin-Löf type theorist. A first and rather coarse reflection suggests three basic principles of Martin-Löf type theory.

(A0) (Predicativism) The realm of types is built in stages (by the idealized type theorist). It is not a completed totality. In declaring what are the elements of a particular type it is disallowed to make reference to all types.

(A1) A type $A$ is defined by describing how a canonical element of $A$ is formed as well as the conditions under which two canonical elements of $A$ are equal.

(A2) The canonical elements of a type must be namable, that is to say, they must allow for a symbolic representation, as a word in a language whose alphabet, in addition to countably many basic symbols, consists of the elements of previously introduced types. Here "previously" refers back to the stages of (A0).

Principles (A0), (A1), and (A2) are liable to different readings and are in need of further rumination. The predicativism of **MLTT** is much more liberal than the one based on autonomous progressions of theories delineated in Feferman's and Schütte's work. In addition to the inductively defined set of natural numbers it allows for other inductive definitions of a rather general kind. This type of predicativism is often referred to as "generalized predicative".

The types of **MLTT** are comprised roughly of three different kinds: explicitly defined types (e.g. the empty type and the type of Booleans $N_1$) as well types defined explicitly from given types or families of types (e.g. $A+B$, $(\Sigma x : A)B(x)$), inductively defined types, and functions types (e.g. $A \to B$, $(\Pi x : A)B(x)$). The understanding of function types in **MLTT** requires an elucidation of the notion of function. In [45] the notion of function is explained in terms of the notion of hypothetical proof which is declared to be a primitive notion (see [45], p.41). Similarly, ([42], p.23) Martin-Löf writes that: *"The reason that $B^A$ can be constructed as a set is that we take the notion of function as primitive, instead of defining a function as a set of ordered pairs or a binary relation satisfying the usual existence and uniqueness conditions, which would make it a category (like $\mathcal{P}(A)$) instead of a set."* These explanations, however, are not very informative and the latter seems to be rather misguided as the distinction between a set of functions $\{0,1\}^{\mathbb{N}}$ and a proper class of subsets of $\mathbb{N}$ can be made in constructive set theories notwithstanding that functions are defined as sets of ordered pairs (see [1]).

The axiom that the class of functions $A \to B$ of a set $A$ into a set $B$ is a set is known as the *Exponentiation* axiom. The difference between the axioms of Power set and Exponentiation was explained by Myhill ([46], p.351) as follows: *"Power set seems especially nonconstructive and impredicative compared with the other axioms: it does not involve, as the others do, putting together or taking apart sets that one has already constructed but rather selecting, out of the totality of all sets those that stand in the relation of inclusion to a given set. One could make the same, admittedly vague, objection to the existence of the set $A \to B$ of mappings of $A$ to $B$ but I do not think the situation is parallel–a mapping or function is a rule, a finite object which can actually be given;.."*

To be able draw any limits of **MLTT** within classical set theory we must be able to understand how functions are conceived in **MLTT**. If the function type $A \to B$ is to be taken in the classical set-theoretic way it will be a fully impredicative set and the difference between **MLTT** and classical set theory wanes in importance. The charge of impredicativity against the intuitionistic notion of proof is well known. The problem is vividly related in ([13], pp.274-275): *"This may make it appear that, in order to recognize an operation as a proof of a statement of the form $(B \to C) \to D$, we must survey all possible proofs of $B \to C$ to which the operation might be applied; and in order to do this, we should need to know the whole of (existing) mathematics, since we cannot tell on what mathematical results a recognition of the efficacy of an operation proving $B \to C$ might not draw. If this were so, the intuitionistic acceptance of compositionality and consequent rejection of mathematical holism would be spurious: an understanding of $(B \to C) \to D$ would not rest on a* prior *understanding of $B \to C$, since an understanding of $B \to C$ would already involve a knowledge of all mathematics. Futhermore, meaning and even proof itself would be unstable. As mathematics advances, we become able to conceive of new operations and to recognize them and others as effectively transforming proofs of $B$ into proofs of $C$: and so the meaning of $B \to C$ would change, if a grasp of it required us to circumscribe such operations in thought. Moreover, an operation which would transform proofs of $B \to C$ available to us now into a proof of $D$ might not so transform proofs of $B \to C$ which became available to us with the advance of mathematics: and so what would now count as a valid proof of $(B \to C) \to D$ would no longer count as one."* Dummett ([13],p.274) assures us that these fears are groundless: *"In order to recognize an operation as a proof of $(B \to C) \to D$, we must think of it as acting on anything we may ever recognize as a proof of $B \to C$. Of such a proof, we know in advance only what is specified by the intuitive explanation of $\to$: namely, that we recognize it as an effective operation, and as one that will transform any proof of $B$ into a proof of $C$."* It's not obvious to me what to make of Dummett's rebuttal of the charge of impredicativity but one aspect about it seems to be clear enough, namely that in intuitionistic mathematics operations must be effective. This is also confirmed on the next page [13],p.275: *"An operation $f$ defined over a domain $D$ carries each element $x$ of $D$ into an element $f(x)$ of its range $R$. In intuitionistic mathematics the operation must be given as an effective means of determining the result $f(x)$ from the way in which $x$ is given ... ."*

[49] (p.48) offers the following explanation of the concept of function in **MLTT**: *"The basic notion of function is an expression formed by abstraction."* Martin-Löf's remark about function types (or sets) in [42] (p.27) is somewhat cryptic and seems to support an impredicative reading: *"Since, in general, there are no exhaustive rules for generating all functions from one set to another, it follows that we cannot generate inductively all the elements of a set of the form* $(\Pi x \in A)B(x)$ *(or, in particular, of the form $B^A$, like $N^N$)."* However, this quote allows for a much more benign interpretation. In **MLTT** the canonical objects of $A \to B$ are formed by abstraction $(\lambda x)t(x)$ from functional expressions (terms) $t(x)$. As Martin-Löf type theory is an open-ended system, the addition of new types with their pertaining constructors and selectors is always possible and will give rise to new terms involving these constants. As a result, terms will be created which, by putting a lambda in front of them, may give rise to new canonical elements of function types $A \to B$ created at earlier stages. This seeming circularity in defining elements of a type after the creation of this type not only affects function types but also inductively defined types like the $W$-types. To my mind, the distinction between canonical and non-canonical elements given by Martin-Löf is not a happy one in the case of function types. The meaning of an implication $A \supset B$ is explained by saying what counts as a canonical proof of it. If the meaning depended on all that counts as a proof of it, then the meaning would change each time we found a new proof of it. However, in **MLTT** any non-canonical proof $b$ of the proposition $A \supset B$ gives rise to the canonical proof $(\lambda x)Ap(b, x)$ of $A \supset B$, rendering the distinction between canonical and non-canonical proofs just a formal one. Be that as it may, there is agreement among intuitionists and constructivists that a function $f$ is always given by rule which effectively determines the result $f(x)$ for every argument $x$ in the domain of $f$. In **MLTT** functions are presented by functional expressions that encapsulate a programme for calculating $f(x)$ for $x$ in the domain of $f$. In light of this, the main argument for conceiving of function types as predicative will be based on the tenet that functions be concrete objects given by rules. At first blush, this appears to entail that all functions from $\mathbb{N}$ to $\mathbb{N}$ must be recursive, and thus point to Church's thesis. There are, however, cherished principles of Brouwerian mathematics such as the Fan theorem that are incompatible with $CT$. Also the extensional version of Martin-Löf type theory refutes $CT$ for the simple reason that the propositions-as-types interpretation forces $CT$ to be a $\Pi\Sigma$-type whose inhabitedness conjoined with the principle of extensionality for functions with domain $\mathbb{N} \to \mathbb{N}$ implies the solvability of the halting problem. This result, though, does not mean that extensional **MLTT** proves the existence of non-recursive functions as one can easily construct models of **MLTT** with extensionality wherein all functions are recursive. Furthermore, intensional **MLTT** is compatible with $CT$.

The problem thus remains to delineate a class of functions that comprises all functions acceptable in Martin-Löf type theory. I will argue that all functions that deserve to be called effective must at least be definable in a way that is persistent with expansions of the universe of types.

To put flesh to this idea, I consider it fruitful to investigate a rigorous model of the principles underlying **MLTT** within set theory. In the following, let us adopt a classical Cantorian point of view and analyze the principles (A0),(A1),(A2) on this basis. Firstly, types are to be interpreted as sets. By Gödel numbering, (A2) hereditarily has the consequence that nothing will be lost by considering all types to be surjective images of subsets of $\mathbb{N}$. In combination with (A1), such an encoding yields that every inductive type $A$ can be emulated by an inductive definition $\Phi$ over the natural numbers together with a decoding function $D$, where

$$\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$$

is a (class) function from the class of all subsets of $\mathbb{N}$, $\mathbf{pow}(\mathbb{N})$, to $\mathbf{pow}(\mathbb{N})$. The set inductively defined by $\Phi$, $\Phi^\infty$, has the set-theoretic definition

$$\begin{aligned} \Phi^\infty &:= \bigcup_\alpha \Phi^\alpha, \\ \Phi^\alpha &:= \Phi(\bigcup_{\beta<\alpha} \Phi^\beta) \cup \bigcup_{\beta<\alpha} \Phi^\beta, \end{aligned}$$

where $\alpha$ ranges over the ordinals. The elements of type $A$ are then considered to be represented by the elements of $\Phi^\infty$. Thus the type $A$ will be identified with the set

$$\{D(x) : x \in \Phi^\infty\}.$$

The vast majority of inductively defined types of **MLTT** is given by monotone operators[13] $\Phi$ whose iterations satisfy $\Phi^\alpha = \Phi(\bigcup_{\beta<\alpha} \Phi^\beta)$. But since Setzer's type theory **TTM** (cf. [69]) features a type which is not generated by a monotone operator, I shall not impose the restriction of monotonicity on operators.

A further step in delineating **MLTT** consists in describing the allowable operators $\Phi$ and decoding functions $D_A$. A common way of classifying inductive definitions proceeds by their syntactic complexity. To find such a syntactic bound it is in order to recall that the type theorists develop their universe of types in stages. Introducing a new type $A$ consists in describing a method for generating its elements. Taking into account that the type-theoretic universe is always in a state of expansion it becomes clear that each time a new element of $A$ is formed by the method of generation for $A$, this method can only refer to types that have been built up hitherto. Furthermore, the method of generation of elements should also obey a persistency condition of the following form: If at a certain stage an object $t$ is recognized as an element of $A$ then an expansion of the type-theoretic universe should not nullify this fact, i.e. the method should remain to be applicable and yield $t : A$ in the expanded universe as well. And in the same vein, if $A$ is a type of codes of types which comes endowed with a type-valued decoding function $D$ (like in the case of type universes), then the validity of equations between types of the form $D(x) = B$ with $x : A$ should remain true under expansions of the universe of types.

Framing the foregoing in set-theoretic terms amounts to saying that the truth of formulas describing $t \in \Phi(X)$ and $D(t) = b$, respectively, ought to be persistent under adding more sets to a set-theoretic universe. In more technical language this means that whenever $\mathbb{M}$ and $\mathbb{P}$ are transitive sets of sets such that $t, X \in \mathbb{M}$, $\mathbb{M} \subseteq \mathbb{P}$ and $(\mathbb{M}, \in_{\mathbb{M}}) \models t \in \Phi(X)$, then $(\mathbb{P}, \in_{\mathbb{P}}) \models t \in \Phi(X)$[14] should obtain as well. The same persistency property should hold for formulas of the form '$D(x) = b$'.

The formulas which can be characterized by the latter property are known in set theory as the $\Sigma$-formulas. They are exactly the collection of set-theoretic formulas generated from the atomic and negated atomic formulas by closing off under $\wedge, \vee$, bounded quantifiers $(\forall x{\in}a), (\exists x{\in}a)$ and unbounded existential quantification $\exists x$ (cf. [2],I.8).

In view of the preceding, one thus is led to impose restrictions on the complexity of inductive definitions for generating types in **MLTT** as follows.

[13] $\Phi$ is said to be *monotone* if $X \subseteq Y$ implies $\Phi(X) \subseteq \Phi(Y)$.

[14] $\in_{\mathbb{M}}$ stands for the elementhood relation restricted to sets in $\mathbb{M}$.

(A3) Every inductive definition $\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$ for generating the elements of an inductive type $A$ in **MLTT** and its pertinent decoding function are definable by set-theoretic $\Sigma$-formulas. These formulas may contain further sets as parameters, corresponding to previously defined types.

To avoid misunderstandings, I'd like to emphasize that (A3) is not meant to say that every such $\Sigma$ inductive definition gives rise to a type acceptable in **MLTT**. (A3) is intended only as a delineation of an upper bound.

Having determined that the case for the predicativity of function types rests on the requirement that functions be given by rules that enable one to compute their values effectively, it is plain that any such function must be definable in an absolute way. In view of the foregoing arguments for restrictions imposed on inductive types in conjunction with (A2) one is led to require the following:

(A4) All the functions figuring in **MLTT** belong to the set

$$\mathbf{Func} := \{f \subseteq \mathbb{N} \times \mathbb{N} : f \text{ is a } \Sigma\text{-definable function}\}.$$

Note that the functions in **Func** are required to have a lightface $\Sigma$ definition, that is to say definitions must not involve parameters (oracles).

The functions in **Func** are known from generalized recursion theory on ordinals. **Func** consists all $\infty$-partial recursive functions from $\mathbb{N}$ to $\mathbb{N}$ (see [33]). In terms of the analytical hierarchy, **Func** can be characterized as the class of all (lightface) $\Sigma^1_2$-definable partial functions from $\mathbb{N}$ to $\mathbb{N}$.

(A4) and previous considerations induce us to delineate the interpretation of product types as follows:

(A5) Every product type $(\Pi x : A)B(x)$ in **MLTT** is a set of functions from $A$ to $\bigcup_{x:A} B(x)$ $\Sigma$-definable (with parameters) from previously defined types and the set **Func**. Moreover, $(\Pi x : A)B(x)$ is a subset of **Func**.

The principles (A0)-(A5) will allow us to draw a limit to **MLTT** in the guise of a small fragment of **ZF**. This fragment, notated **T**, will be based on the ubiquitous Kripke-Platek set theory, **KP**. Kripke-Platek set theory is a truly remarkable subsystem of **ZF**. Though considerably weaker than **ZF**, a great deal of set theory requires only the axioms of this subsystem. **KP** arises from **ZF** by omitting the power set axiom and restricting separation and collection to bounded formulas, that is formulas without unbounded quantifiers. **KP** has been a major site of interaction between many branches of logic (for more information see the book by Barwise [2]). The transitive models of **KP** are called *admissible sets*.

To describe **T**, we have to alter **KP** slightly. Among the axioms of **KP** is the foundation scheme which says that every non-empty definable class has an $\in$-least element. Let $\mathbf{KP}^r$ result from **KP** by restricting the foundation scheme to sets. In addition to $\mathbf{KP}^r$, **T** has an axiom asserting that every set is contained in a transitive set which is a $\Sigma_1$ elementary substructure of the set-theoretic universe $V$ (written $M \prec_1 V$). To be more precise, let $M \prec_1 V$ stand for the scheme

$$\forall a {\in} M\, [\exists x\, \phi(x, a) \to \exists x \in M\, \phi(x, a)]$$

for all bounded formulas $\phi(x, y)$ with all free variables exhibited. Using a $\Sigma_1$ satisfaction predicate, $M \prec_1 V$ can actually be expressed via a single formula.

We take **T** to be the theory

$$\mathbf{KP}^r + \forall x\, \exists M\, (x {\in} M \ \wedge\ M \prec_1 V).$$

The following theorems are provable in **T**.

**Theorem 6.1** **Func** *is a set.*

**Proof**: Let $X$ be a set satisfying $X \prec_1 V$. One easily checks that every element of **Func** is in $X$ since it is $\Sigma$ definable without parameters. Moreover, **Func** is subset of $M$ which is definable in $M$, whence **Func** is a set by bounded separation. □

**Theorem 6.2** *If* $\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$ *is definable by a* $\Sigma$ *formula with parameters in* $M$ *and* $M \prec_1 V$*, then* $\Phi^\infty \in M$.

The above theorem supports the claim that everything a Martin-Löf type theorist can ever develop can be emulated in **T** or, to put it more pictorially, that the boundaries of the type theorist world are to be drawn inside $M$, where $M$ satisfies $M \prec_1 V$.

Before elaborating further on this question, it might be interesting to give an equivalent characterization of **T** which is couched in terms of subsystems of second order arithmetic.

**Theorem 6.3** ([61], Theorem 5.3) *The theories* $(\mathbf{\Pi}^1_2\textbf{-CA})\upharpoonright$ *and* **T** *prove the same statements of second order arithmetic.*

Resuming the question of the type theorist's limit, I shall now argue on the basis of **T** that every set $M \prec_1 V$ with $\mathbf{Func} \in M$ is a model of **MLTT**, i.e. it contains all the types that may ever be constructed in **MLTT**. The argument may run in this way: Types are interpreted as sets. At a certain stage the idealized type theorist, called $ITT$, has a certain repertoire of type forming operations, say $\mathcal{C}$. The operations correspond to a collection $\mathcal{C}_{Set}$ of $\Sigma$-definable operations on sets. Further, assume that $ITT$ introduces a new type $A$ by utilizing $\mathcal{C}$. Inductively we may assume that any set $M$ with $M \prec_1 V$ and $\mathbf{Func} \in M$ is a model of $ITT$'s reasoning as developed up to this point. Thus any such $M$ is closed under $\mathcal{C}_{Set}$. According to (A3), the generation of the elements of $A$ gives rise to an operator $\Phi_M : \mathbf{pow}(\mathbb{N}) \cap M \to \mathbf{pow}(\mathbb{N}) \cap M$ and a decoding function $D_M$ which are both $\Sigma$-definable on $M$ whenever $M \prec_1 V$. Moreover, $\Phi_M$ and $D_M$ are uniformly definable on all $M \prec_1 V$, that is to say, there are $\Sigma$-formulas $\psi(x, y)$ and $\delta(u, v)$ such that $\Phi_M(X) = Y$ iff $(M, \in_M) \models \psi(X, Y)$ and $D_M(u, v)$ iff $(M, \in_M) \models \delta(u, v)$ whenever $X, Y \in \mathbf{pow}(\mathbb{N}) \cap M$, $u, v \in M$, and $M \prec_1 V$. Now define

$$\Phi : \mathbf{pow}(\mathbb{N}) \to \mathbf{pow}(\mathbb{N})$$

by letting $\Phi(X) = \Phi_M(X)$, where $X \in M$ and $M \prec_1 V$. $\Phi$ defines a function since the $\Phi_M$ are $\Sigma$ definable and for every $X \subseteq \mathbb{N}$ there exists $M \prec_1 V$ such that $X \in M$. Thus **T** proves that $\Phi$ is a $\Sigma$-definable operator, i.e.,

$$\mathbf{T} \vdash \forall X \subseteq \mathbb{N}\, \exists Y\, \Phi(X) = Y.$$

Employing Theorem 6.2, one can deduce that $\Phi^\infty$ is a set. Moreover, as $\Phi^\infty$ is $\Sigma$ definable too, one can infer that $\Phi^\infty \in M$ and thus

$$A = \{D(u) \,:\, u \in \Phi^\infty\} = \{D_M(u) \,:\, u \in \Phi^\infty_M\} \in M$$

for every $M \prec_1 V$.

# 7 The higher infinite and new axioms

At the end of this article it is, perhaps, in order to point out that there are parts of mathematics which are permeated by set theory and thus are not capable of a constructive consistency proof. In particular the structure of sets of reals is affected by set-theoretic axioms. Examples of results that require uncountably many iterations of the power set operation are D. Martin's theorem that all Borel games are determined and Friedman's Borel diagonalization theorem.

The set existence axioms considered in modern set theory go way beyond Zermelo-Fraenkel set theory. These so-called large cardinal axioms imply that certain infinite games played on sets of reals always possess a winning strategy. Perhaps, most notably, projective determinacy, **PD**, asserts that all games which are definable in the language of $\mathbf{Z_2}$ have a winning strategy. The relation to large cardinals is that **PD** is a consequence of the existence of infinitely many Woodin cardinals. **PD** leads to a very satisfying structure theory for projective sets of reals in that under **PD** every projective set $A$ of reals is Lebesgue measurable, has the property of Baire, and if $A$ is uncountable, then $A$ has a perfect subset (see [80]).

The ASL 2000 annual meeting also saw a panel discussion devoted to the question: *Does mathematics need new axioms?* (See [21]). Two of the panelists, Harvey Friedman and John Steel, maintained that mathematics needs new axioms, i.e. principles not already provable in **ZFC**. Steel's line of argument was that the descriptive set theory emanating from large cardinals is a reason why mathematicians should adopt these large cardinal axioms. Harvey Friedman held that his most recent discovery, called *Boolean relation theory* (BT), provides strong reasons for adopting new axioms as BT has consequences for the core of mathematics which are hard to dismiss. Roughly speaking, BT is concerned with the relationship between sets and their images under multivariate functions. What is most striking about Friedman's results in BT is that they encapsulate the proof-theoretic strength of certain large cardinals. Only the future can tell whether BT is ever going to play a role in the dealings of everyday mathematics.

**Acknowledgement.** I wish to thank John Derrick for reading an earlier version of this paper, bringing several inaccuracies to my attention, and suggesting improvements. Notwithstanding that we hold differing views on the foundations of mathematics, our discussions of the paper at Monk Fryston Hall (while indulging in a cream tea) were most enjoyable.

# References

[1] P. Aczel, M. Rathjen: *Notes on constructive set theory*, Technical Report 40, Institut Mittag-Leffler (The Royal Swedish Academy of Sciences, 2001). *http://www.ml.kva.se/preprints/archive2000-2001.php*

[2] J. Barwise: *Admissible Sets and Structures* (Springer, Berlin 1975).

[3] P. Bernays: *Hilbert, David,* Encyclopedia of philosophy, Vol. 3 (Macmillan and Free Press, New York, 1967) 496–504.

[4] E. Bishop: *Foundations of constructive analysis.* (McGraw-Hill, New York, 1967).

[5] L.E.J. Brouwer: *Weten, willen, spreken* (Dutch). Euclides 9 (1933) 177-193.

[6] D.K. Brown: *Functional analysis in weak subsystems of second order arithmetic.* Ph.D. Thesis (Pennsylvania State University, University Park,1987).

[7] D.K. Brown and S. Simpson: *Which set existence axioms are needed to prove the separable Hahn-Banach theorem?*AAnnals of Pure and Applied Logic 31 (1986) 123-144.

[8] W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981).

[9] H.B. Curry, R. Feys:*Combinatory logic, vol. I.* (North-holland, Amsterdam, 1958)

[10] R. Diestel: *Graph Theory* (Springer, New York-Berlin-Heidelberg, 1997).

[11] M. Dummett: *The philosophical basis of intuitionistic logic.* In: H.E. Rose et al.(eds.): *Logic Colloquium '73* (North-Holland, Amsterdam, 1973) 5-40.

[12] M. Dummett: *Elements of intuitionism.* 2nd edition (Clarendon Press,Oxford,2000).

[13] S. Feferman: *A Language and Axioms for Explicit Mathematics*, Lecture Notes in Math. 450 (Springer, Berlin, 1975), 87–139.

[14] S. Feferman: *Constructive Theories of Functions and classes* in: Boffa, M., van Dalen, D., McAloon, K. (eds.), *Logic Colloquium '78* (North-Holland, Amsterdam, 1979) 159–224.

[15] S. Feferman: *Why a little bit goes a long way* In: S. Feferman: *In the light of logic.*(Oxford University Press, Oxford, 1998).

[16] S. Feferman: *Weyl vindicated: "Das Kontinuum" 70 years later.* In: S. Feferman: *In the light of logic.*(Oxford University Press, Oxford, 1998).

[17] S. Feferman, H. Friedman, P. Maddy, J. Steel: *Does mathematics need new axioms?* Bulletin of Symbolic Logic 6 (2000) 401-446.

[18] H. Friedman, N. Robertson, P. Seymour: *The metamathematics of the graph minor theorem*, Contemporary Mathematics 65 (1987) 229–261.

[19] G. Gentzen: *Untersuchungen über das logische Schliessen.* Mathematische Zeitschrift 39 (1935) 176-210, 405-431.

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

[21] K. Gödel: *The present situation in the foundations of mathematics.* In: Collected Works, vol. III (Oxford University Press, New York, 1995).

[22] D. Hilbert: *Über das Unendliche.* Mathematische Annalen 95 (1926) 161-190. English translation in: J. van Heijenoort (ed.): *From Frege to Gödel. A Source Book in mathematical logic, 1897-1931.*(Harvard University Press, Cambridge, Mass.,1967).

[23] D. Hilbert and P. Bernays: *Grundlagen der Mathematik II* (Springer, Berlin, 1938).

[24] P. Hinman: *Recursion-theoretic hierarchies* (Springer, Berlin, 1978).

[25] W.A. Howard: *The formulae-as-types notion of construction.* (Privately circulated notes, 1969).

[26] G. Jäger and W. Pohlers: *Eine beweistheoretische Untersuchung von* $\Delta^1_2$*–CA+ BI und verwandter Systeme*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch–Naturwissenschaftliche Klasse (1982).

[27] G. Kreisel: *Ordinal logics and the characterization of informal concepts of proof.* In: *Proceedings of the 1958 International Congress of Mathematicians,*(Edinburgh,1960) 289–299.

[28] P. Maddy: *Realism in mathematics.* (Clarendon Press, Oxford,1990).

[29] P. Maddy: *Naturalism in mathematics.* (Clarendon Press,OOxford,1997).

[30] P. Martin-Löf: *An intuitionistic theory of types: predicative part*, in: H.E. Rose and J. Shepherdson (eds.): *Logic Colloquium '73* (North-Holland, Amsterdam, 1975) 73–118.

[31] P. Martin-Löf: *Constructive mathematics and computer programming*, in: L.J. Cohen, J. Los, H. Pfeiffer, K.-P. Podewski:*LMPS IV* (North-Holland, Amsterdam, 1982).

[32] P. Martin-Löf: *Intuitionistic Type Theory*, (Bibliopolis, Naples, 1984).

[33] P. Martin-Löf: *Truth of a proposition, evidence of a judgement, validity of a proof* Synthese (1987) 73: 87–99.

[34] P. Martin-Löf: *Verificationism then and now*, in: W. DePauli-Schimanovich, E. Köhler, and F. Stadler (eds.): *The Foundational Debate: Complexity and Constructivity in Mathematics and Physics* (Kluwer, Dordrecht, 1995).

[35] P. Martin-Löf: *On the meanings of the logical constants and the justifications of the logical laws*, Nordic Journal of Philosophical Logic (1996) 1: 11–60.

[36] J. Myhill: *Constructive Set Theory*, J. Symbolic Logic 40 (1975) 347–382.

[37] B. Nordström, K. Petersson and J.M. Smith: *Programming in Martin–Löf's Type Theory*, (Clarendon Press, Oxford, 1990).

[38] E. Palmgren: *An information system interpretation of Martin-Löf's partial type theory with universes*, Information and Computation1106 (1993) 26–60.

[39] E. Palmgren: *On universes in type theory.* In: G. Sambin, J. smith (eds.): *Twenty-five years of type theory* (Oxford University Press, Oxford, 1998) 191–204.

[40] J. Paris, L. Harrington: *A mathematical incompleteness in Peano arithmetic.* In: J. Barwise (ed.): *Handbook of Mathematical Logic* (North Holland, Amsterdam, 1977) 1133-1142.

[41] D. Prawitz: *Meaning and proofs: on the conflict between classical and intuitionistic logic.* Theoria 43 (1977)11–40.

[42] M. Rathjen: *Proof-Theoretic Analysis of KPM*, Arch. Math. Logic 30 (1991) 377–403.

[43] M. Rathjen: *The strength of some Martin–Löf type theories.* Preprint, Department of Mathematics, Ohio State University (1993) 39 pages.

[44] M. Rathjen: *Proof theory of reflection.* Annals of Pure and Applied Logic 68 (1994) 181–224.

[45] M. Rathjen: *The recursively Mahlo property in second order arithmetic.* Mathematical Logic Quarterly 42 (1996) 59–66.

[46] M. Rathjen, E. Palmgren: *Inaccessibility in constructive set theory and type theory.* Annals of Pure and Applied Logic 94 (1998) 181–200.

[47] M. Rathjen: *Explicit mathematics with the monotone fixed point principle. II: Models.* Journal of Symbolic Logic 64 (1999) 517–550.

[48] M. Rathjen: *The superjump in Martin-Löf type theory.* In: S. Buss, P. Hajek, P. Pudlak (eds.): *Logic Colloquium '98*, Lecture Notes in Logic 13. (Association for Symbolic Logic, 2000) 363–386.

[49] M. Rathjen: *Realizing Mahlo set theory in type theory.* Arch. Math. Logic 42 (2003) 89–101.

[50] M. Rathjen: *The constructive Hilbert programme and the limits of Martin-Löf type theory* Synthese 147 (2005) 81 - 120.

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

[52] T. Setzer: *Proof theoretical strength of Martin–Löf type theory with W–type and one universe* (Thesis, University of Munich, 1993).

[53] A. Setzer: *A well-ordering proof for the proof theoretical strength of Martin-Löf type theory*, Annals of Pure and Applied Logic 92 (1998) 113–159.

[54] A. Setzer: *Extending Martin-Löf type theory by one Mahlo-universe.* Archive for Mathematical Logic (2000) 39: 155-181.

[55] S. Simpson: *Nichtbeweisbarkeit von gewissen kombinatorischen Eigenschaften endlicher Bäume*, Archiv f. Math. Logik 25 (1985) 45–65.

[56] S. Simpson: *Partial realizations of Hilbert's program.* Journal of Symbolic Logic 53 (1988) 349-363.

[57] S. Simpson: *Subsystems of second order arithmetic* (Springer, Berlin, 1999).

[58] W. Tait: *Finitism.* Journal of Philosophy 78 (1981) 524-546.

[59] B. van der Waerden: *Moderne Algebra I/II* (Springer, Berlin,1930/31).

[60] H. Weyl: *Die Stufen des Unendlichen.* (Verlag von Gustav Fischer, Jena, 1931).

[61] H. Weyl: *Philosophy of Mathematics and Natural Sciences.* (Princeton University Press, Princeton, 1949)

[62] W.H. Woodin: *Large cardinal axioms and independence: The continuum problem revisited*, The Mathematical Intelligencer vol. 16, No. 3 (1994) 31–35.