The Independence of Peano's Fourth Axiom from Martin-Löf's Type Theory without Universes

Jan M. Smith
Department of Computer Science
University of Göteborg/Chalmers
S-412 96 Göteborg
Sweden

March 1987

1 Introduction

In Hilbert-Ackermann [2] there is given a simple proof of the consistency of first order predicate logic by reducing it to propositional logic. Intuitively, the proof is based on interpreting predicate logic in a domain with only one element. Tarski [7] and Gentzen [1] have extended this method to simple type theory by starting with an individual domain consisting of a single element and then interpreting a higher type by the set of truth valued functions on the previous type.

I will use the method of Hilbert and Ackermann on Martin-Löf's type theory without universes to show that $\neg\,\mathsf{Eq}(A, a, b)$ cannot be derived without universes for any type A and any objects a and b of type A. In particular, this proves the conjecture in Martin-Löf [5] that Peano's fourth axiom $(\forall\, x \in \mathsf{N})\neg\,\mathsf{Eq}(\mathsf{N}, 0, \mathsf{succ}(x))$ cannot be proved in type theory without universes. If we by consistency mean that there is no closed term of the empty type, then the construction will also give a consistency proof by finitary methods of Martin-Löf's type theory without universes. So, without universes, the logic obtained by interpreting propositions as types is surprisingly weak. This is in sharp contrast with type theory as a computational system, since, for instance, the proof that every object of a type can be computed to normal form cannot be formalized in first order arithmetic.

The nonderivability of $\neg\,\mathsf{Eq}(\mathsf{N}, 0, 1)$ for the version of type theory given in Martin-Löf [4] was already shown in Smith [6] as a corollary to a somewhat less straightforward construction made with a different purpose. The proofs in this paper will work for any of the different formulations of Martin-Löf's type theory.

2 The construction of the interpretation

We define a truth valued function φ on the types of Martin-Löf's type theory without universes. Intuitively, $\varphi(A) = \top$ means that the interpretation of the type A is a set with one element and $\varphi(A) = \bot$ means that A is interpreted as the empty set. φ is defined for each type expression $A(x_1, \ldots, x_n)$ by recursion on the length of the derivation of $A(x_1, \ldots, x_n)$ *type* $[x_1 \in A_1, \ \ldots, \ x_n \in A_n(x_1, \ldots, x_{n-1})]$, using the clauses

$$
\begin{aligned}
\varphi(\mathsf{N}_0) &= \bot \\
\varphi(\mathsf{N}_k) &= \top \qquad (k = 1, 2, \ldots\ ) \\
\varphi(\mathsf{N}) &= \top \\
\varphi(\mathsf{Eq}(A, a, b)) &= \varphi(A) \\
\varphi(A + B) &= \varphi(A) \vee \varphi(B) \\
\varphi((\Pi x \in A) B(x)) &= \varphi(A) \to \varphi(B(x)) \\
\varphi((\Sigma x \in A) B(x)) &= \varphi(A) \wedge \varphi(B(x)) \\
\varphi((\mathsf{W} x \in A) B(x)) &= \varphi(A) \wedge (\neg \varphi(B(x))) \\
\varphi(\{x \in A \mid B(x)\}) &= \varphi(A) \wedge \varphi(B(x))
\end{aligned}
$$

$\wedge$, $\vee$, $\to$, and $\neg$ denote the usual boolean operations.

That φ really interprets type theory in the way we have intended is the content of the following theorem.

Theorem. *Let* $a(x_1, \ldots, x_n) \in A(x_1, \ldots, x_n)$ $[x_1 \in A_1, \ \ldots, \ x_n \in A_n(x_1, \ldots, x_{n-1})]$ *be derivable in type theory without universes. Then* $\varphi(A(x_1, \ldots, x_n)) = \top$ *provided that* $\varphi(A_1) = \ \cdots \ = \varphi(A_n(x_1, \ldots, x_{n-1})) = \top$.

In the proof of this theorem we will use two lemmas. The first says that the truth value assigned to a type expression is preserved under substitution. The second lemma says that equality between types is preserved by φ.

Lemma 1. *If* $A(x_1, \ldots, x_n)$ *type* $[x_1 \in A_1, \ \ldots, \ x_n \in A_n(x_1, \ldots, x_{n-1})]$ *and* $a_1 \in A_1, \ \ldots, \ a_n \in A_n(a_1, \ldots, a_{n-1})$ *are derivable in type theory without universes, then* $\varphi(A(x_1, \ldots, x_n)) = \varphi(A(a_1, \ldots, a_n))$.

Proof. The proof is by induction on the length of the derivation of $A(x_1, \ldots, x_n)$ *type* $[x_1 \in A_1, \ \ldots, \ x_n \in A_n(x_1, \ldots, x_{n-1})]$. The only type forming rule where free variables may be introduced is Eq-formation. Since $\varphi(\mathsf{Eq}(A, a, b)) = \varphi(A)$ the induction hypothesis directly gives the result.

Lemma 2. *If* $A(x_1, \ldots, x_n) = B(x_1, \ldots, x_n)$ $[x_1 \in A_1, \ \ldots, \ x_n \in A_n(x_1, \ldots, x_{n-1})]$ *is derivable in type theory without universes, then* $\varphi(A(x_1, \ldots, x_n)) = \varphi(B(x_1, \ldots, x_n))$.

Proof. This lemma is straightforwardly proved by induction on the length of the derivation of $A(x_1, \ldots, x_n) = B(x_1, \ldots, x_n)$ $[x_1 \in A_1, \; \ldots, \; x_n \in A_n(x_1, \ldots, x_{n-1})]$. Note that lemma 1 is needed for the rule

$$\frac{a = b \in A \qquad C(x)\ \mathit{type}\ [x \in A]}{C(a) = C(b)}$$

Proof of the theorem. The proof is by induction on the length of the derivation of $a(x_1, \ldots, x_n) \in A(x_1, \ldots, x_n)$ $[x_1 \in A_1, \; \ldots, \; x_n \in A_n(x_1, \ldots, x_{n-1})]$. I will only discuss a few of the rules; the remaining can be handled in the same way.

Equality of types

$$\frac{a \in A \qquad A = B}{a \in B}$$

By the induction hypothesis we have that $\varphi(A) = \top$ and, by lemma 2, that $\varphi(A) = \varphi(B)$. Hence, $\varphi(B) = \top$.

There are different formulations of the rules for the Eq-type in Martin-Löf [3] and Martin-Löf [4, 5]. I will here use the earlier formulation which is the one now used by Martin-Löf since it does not destroy the decidability of the judgemental equality
$a = b \in A$.

Eq-introduction

$$\frac{a \in A}{\mathsf{eq}(a) \in \mathsf{Eq}(A, a, a)}$$

Since, by the definition of φ, $\varphi(\mathsf{Eq}(A, a, a)) = \varphi(A)$, the induction hypothesis directly gives $\varphi(\mathsf{Eq}(A, a, a)) = \top$.

Eq-elimination

$$\frac{c \in \mathsf{Eq}(A, a, b) \qquad d(x) \in C(x, x, \mathsf{eq}(x))\ \ [x \in A]}{\mathsf{J}(c, d) \in C(a, b, c)}$$

By the induction hypothesis we have that $\varphi(\mathsf{Eq}(A, a, b)) = \top$ and that $\varphi(C(x, x, \mathsf{eq}(x))) = \top$ if $\varphi(A) = \top$. Hence, since $\varphi(\mathsf{Eq}(A, a, b)) = \varphi(A)$, $\varphi(C(x, x, \mathsf{eq}(x))) = \top$ which, by lemma 1, gives $\varphi(C(a, b, c)) = \top$.

If we instead had considered the Eq-rules in Martin-Löf [4] we could simplify the definition of φ by putting $\varphi(\mathsf{Eq}(A, a, b)) = \top$.

Π-introduction

$$\frac{b(x) \in B(x)\ \ [x \in A]}{\lambda(b) \in (\Pi x \in A)B(x)}$$

By the induction hypothesis we know that $\varphi(B(x)) = \top$ if $\varphi(A) = \top$. Since $\varphi((\Pi x \in A)B(x)) = \varphi(A) \to \varphi(B(x))$ this gives that $\varphi((\Pi x \in A)B(x)) = \top$.

Π-elimination

$$\frac{a \in A \qquad c \in (\Pi x \in A)B(x)}{\mathsf{apply}(c,a) \in B(a)}$$

According to the induction hypothesis, we have $\varphi(A) = \top$ and $\varphi((\Pi x \in A)B(x)) = \top$, which, since $\varphi((\Pi x \in A)B(x)) = \varphi(A) \to \varphi(B(x))$, gives that $\varphi(B(x)) = \top$. Hence, by lemma 1, $\varphi(B(a)) = \top$.

N-elimination

$$\frac{n \in \mathsf{N} \qquad d \in C(0) \qquad e(x,y) \in C(\mathsf{succ}(x)) \;\; [x \in \mathsf{N},\; y \in C(x)]}{\mathsf{rec}(n,d,e) \in C(n)}$$

By the induction hypothesis we have that $\varphi(C(0)) = \top$ which, by lemma 1, gives $\varphi(C(n)) = \top$.

3 Some consequences of the interpretation

3.1 The unprovability of Peano's fourth axiom

By the interpretation we can now see that for no type A and terms a and b does there exist a closed term t such that

$$t \in \neg\, \mathsf{Eq}(A,a,b) \qquad (*)$$

is derivable in type theory without universes. Assume that $(*)$ holds. Then there must exist a derivation of $\mathsf{Eq}(A,a,b)$ *type* and, hence, also a derivation of $a \in A$. So, by the theorem, $\varphi(A) = \top$ which, together with the definitions of φ and $\neg$, gives

$$\varphi(\neg\, \mathsf{Eq}(A,a,b)) = \varphi(\mathsf{Eq}(A,a,b) \to \mathsf{N}_0) = \varphi(\mathsf{Eq}(A,a,b)) \to \varphi(\mathsf{N}_0) = \varphi(A) \to \bot = \bot$$

Hence, by the theorem, $\neg\, \mathsf{Eq}(A,a,b)$ cannot be derived in type theory without universes.

Assume that Peano's fourth axiom can be derived, that is, that we have a derivation of

$$s \in (\Pi x \in \mathsf{N})\, \neg\, \mathsf{Eq}(\mathsf{N}, 0, \mathsf{succ}(x))$$

for some closed term s. By Π-elimination we then get $\mathsf{apply}(s,0) \in \neg\, \mathsf{Eq}(\mathsf{N},0,\mathsf{succ}(0))$ which is of the form $(*)$ and therefore impossible to derive in type theory without universes.

That no negated equalities can be proved reflects the intuition behind φ, which is that it interprets type theory in a domain with a single element. We can make this explicit inside type theory by introducing a new constant $\star$ and for each type A such that $\varphi(A) = \top$ adding a new rule

$$\star \in A$$

The theorem can still be proved with this new rule added, so the extension is consistent. Since $\varphi((\Pi x \in A)\mathsf{Eq}(A, x, \star)) = \varphi(A) \to \varphi(A) = \top$ we have that $\star \in (\Pi x \in A)\mathsf{Eq}(A, x, \star)$, that is, all objects of a type are equal to $\star$. Note that the extension is classical because $\star \in A \vee (\neg A)$. Since $\star \in \mathsf{Eq}(\mathsf{N}, 0, 1)$, type theory with universes becomes inconsistent if the $\star$-rule is added.

3.2 Well-orderings

The definition of φ on well-orderings, $\varphi((\mathsf{W}x \in A)B(x)) = \varphi(A) \wedge (\neg\varphi(B(x)))$, is made as to validate the rules in Martin-Löf [4]. The W-introduction rule in [4] does not have a bottom clause $\mathsf{0} \in (\mathsf{W}x \in A)B(x)$ since such a clause can be derived using a universe. We can now see that this use of a universe is necessary. Since $\varphi((\mathsf{W}x \in A)B(x)) = \top$ implies $\varphi(A) = \top$ and $\varphi(B(x)) = \bot$ we get, by the theorem, that if $(\mathsf{W}x \in A)B(x)$ is not empty then $B(a)$ must be empty for all a in A. This gives that all elements of a well-ordering type are initial, that is, without predecessors. So, only very trivial well-orderings can be constructed.

If we add a bottom clause to the W-rules and change the definition of φ by $\varphi((\mathsf{W}x \in A)B(x)) = \top$, we get the full computational strength of the well-ordering types and can still prove our theorem.

3.3 Consistency

Since absurdity is interpreted in type theory by the empty type N_0, the obvious way of defining consistency for type theory is to say that there is no closed term of type N_0. Since $\varphi(\mathsf{N}_0) = \bot$, the theorem shows that there cannot be a closed term of type N_0. Clearly, this consistency proof is finitary in the sense of Hilbert and can be carried out in primitive recursive arithmetic. This may seem surprising since the proof theoretic strength of type theory without universes measured in terms of provable well-orderings is, without well-ordering types, the same as first order arithmetic and, with well-ordering types, even far beyond ε_0. However, this is not in conflict with Gödel's second incompletness theorem, because in order to prove Gödel's theorem, primitive recursive predicates must be numeralwise expressible in the theory and, as we have seen, not even equality is numeralwise expressible in type theory without universes.

If we instead by consistency mean that there is no closed term of type $\mathsf{Eq}(\mathsf{N}, 0, 1)$, the consistency of type theory cannot be proved using the theorem since $\varphi(\mathsf{Eq}(\mathsf{N}, 0, 1)) = \top$. Actually, defining absurdity by $\mathsf{Eq}(\mathsf{N}, 0, 1)$ instead of N_0, first order arithmetic can be interpreted in type theory without universes. So, with this definition of consistency, there cannot be a finitary consistency proof.

3.4 Universes

If φ was extended to a universe, then $\varphi(\mathsf{T}(a))$ has to be defined for each object a of the universe U because of the rule

$$\frac{a \in \mathsf{U}}{\mathsf{T}(a)\ \mathit{type}}$$

which says that if a is the code of a type then $\mathsf{T}(a)$ is the type that a encodes. Let n_0 and n_1 be the codes of N_0 and N_1 respectively. Since

$$\mathsf{T}(\mathsf{n}_0) = \mathsf{N}_0 \quad \text{and} \quad \mathsf{T}(\mathsf{n}_1) = \mathsf{N}_1$$

we must have

$$\varphi(\mathsf{T}(\mathsf{n}_0)) = \bot \quad \text{and} \quad \varphi(\mathsf{T}(\mathsf{n}_1)) = \top$$

Hence, lemma 1, which is crucial for the proof of the theorem, would no longer hold.

An obvious way of extending type theory in order to obtain the strength of first order arithmetic is to add Peano's fourth axiom. This would not, however, follow the general pattern of introduction and elimination rules in type theory which is very natural, particularly when viewing a type as a set and not as a proposition: the elements of a set are defined by the introduction rules and the elimination rule makes it possible to define functions by recursion on the set.

Martin-Löf has instead suggested to extend type theory without universes by using the objects 0_2 and 1_2 of type N_2 as codes for N_0 and N_1 respectively. We then have to add the type formation rule

$$\frac{a \in \mathsf{N}_2}{\mathsf{T}(a)\ \mathit{type}}$$

and the type equalities

$$\mathsf{T}(0_2) = \mathsf{N}_0 \qquad\qquad \mathsf{T}(1_2) = \mathsf{N}_1$$

This makes the two element type N_2 function as a very small universe, containing codes only for the types N_0 and N_1, and Peano's fourth axiom can now be proved as in [5].

Acknowledgements.

I would like to thank Per Martin-Löf and Anne Troelstra for helpful comments on a draft of this paper.

References

[1] G. Gentzen. Die Widerspruchsfreiheit der Stufenlogik. *Mathematische Zeitschrift 41, No. 3, 1936, pp.357-366.*

[2] D. Hilbert and W. Ackermann. Grundzüge der Theoretischen Logik. *Springer-Verlag 1928.*

[3] P. Martin-Löf. An intuitionistic theory of types: predicative part. In *Logic Colloquium '73, North-Holland, 1975.*

[4] P. Martin-Löf. Constructive Mathematics and Computer Programming. In *Sixth International Congress for Logic, Methodology, and Philosophy of Science, pp. 153-175. North-Holland, 1982.*

[5] P. Martin-Löf. Intuitionistic Type Theory. *Studies in Proof Theory, Lecture Notes, Bibliopolis, Napoli, 1984.*

[6] J.M. Smith. On a Nonconstructive Type Theory and Program Derivation. To appear in the proceedings of *Conference on Logic and its Applications, Bulgaria 1986 (Plenum Press).*

[7] A. Tarski. Einige Betrachtungen über die Begriffe der ω-Wiederspruchsfreiheit und der ω-Vollständigkeit. *Monatsh. Math. Phys. 40, 1933. pp.279-295.*