The Consistency Theorems

Lecture Notes #1 for MAT 503
Princeton University
Fall 2004
Edward Nelson

Hilbert's program to secure the foundations of mathematics by finitary means is often called a failure. But it achieved several things of fundamental value and lasting importance.[1]

Before the Formalist Enlightenment, even mathematicians who like Dedekind, Peano, and Zermelo were striving to be formal did not make an absolute distinction between syntax and semantics. This distinction is essential; an axiom system in which one is required to understand the meaning of some of the axioms is not an axiomatization at all. Thanks to this enlightenment, it became possible for the first time to study theories with the same precision and clarity with which one studies groups or fields. Mathematics consists of reasoning and computation, and of the two computation is the more fundamental. Aristotle, Leibniz, Boole, and Hilbert each took a big step towards realizing the vision of reducing reasoning to computation.

Hilbert did not do it alone. Major contributions were made by Ackermann, Bernays, Gentzen, Herbrand, and von Neumann, and the intuitionistic critique led by Brouwer and seconded by Weyl, a former student of Hilbert, forced Hilbert to refine and clarify his program.

The major technical achievement of the formalist program is the Hilbert-Ackermann consistency theorem, later extended by Herbrand. The Gentzen cut elimination theorem is similar.

These three consistency theorems are syntactical results closely related to the Gödel completeness theorem, one version of which states: if a formula of a theory is valid in every model then it is a theorem—that is,

[1]The opening paragraphs are taken from my review of *Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s* by Dennis E. Hesseling, Bull. Amer. Math. Soc., vol. 41, 545–549, 2004.

there exists a proof. But the Gödel completeness theorem gives no clue as to how to find such a proof, other than blindly and systematically to search through all possible proofs until one is found—it is impossible to establish an a priori bound on the length of the proof. By contrast, the consistency theorems are algorithmic; they give an effective procedure for finding proofs. But we shall see that the algorithms are infeasible, being superexponentially long.

We shall pursue cut elimination as the basic result. A *cut* is an inference of $B \vee C$ from $A \vee B$ and $\neg A \vee C$; the formula A is the *cut formula*. Gentzen in his inaugural dissertation at Göttingen[2] showed that in his formalism cuts can be eliminated. He calls the result a normal form proof and characterizes such a proof as one that makes no detours; Kleene [Kl §77][3] calls cut elimination the unraveling of unnecessary complications. In my opinion these descriptions are somewhat misleading, as they give no hint that the algorithm for eliminating cuts from proofs is an infeasible superexponentially long procedure.

Gentzen uses the rather cumbersome formalism of sequents. Here we shall follow, with modifications, a simpler version outlined in Problems 6–8 of [Sh Ch. 4].[4] This has the virtue of remaining in the formalism of Hilbert-type systems. In Shoenfield's version of cut elimination, all cuts are eliminated except those in which the cut formula is open; this does not diminish the usefulness of the result.

1. Shoenfield's formalism

We adopt Shoenfield's formalism [Sh Ch. 2] for the predicate calculus, in which the only undefined logical operators are $\neg$, $\vee$, and $\exists$. Thus every formula is either atomic, a negation, a disjunction, or an instantiation. Basically we follow his logical terminology and notation. The defined symbols $\&$, $\rightarrow$, $\leftrightarrow$, and $\forall$ are introduced in the usual way [Sh §2.4]. As variables we take the italic letters with zero or more primes and zero or more digits as subscript.

We use the syntactical variables

[2]Gerhard Gentzen, "Untersuchungen über das logische Schliessen", Mathematische Zeitschrift, vol. 39, 176–210, 405–431, 1934-5. English translation as "Investigations into Logical Deductions", *The Collected Papers of Gerhard Gentzen*, ed. M. E. Szabo, North-Holland, Amsterdam, 1969.

[3]Stephen Cole Kleene, *Introduction to Metamathematics*, North-Holland, Amsterdam, 1952.

[4]Joseph R. Shoenfield, *Mathematical Logic*, Association for Symbolic Logic, A K Peters, Ltd., Natick, Massachusetts, 1967. `http://www.akpeters.com/`

u v	for	expressions
x y z w	for	variables
f g	for	function symbols
e	for	constants (0-ary function symbols)
a b c d	for	terms
p q	for	predicate symbols
A B C D	for	formulas.

All of these may have subscripts or superscripts.

The word "language" is used differently in computational complexity and in mathematical logic; we use it as in [Sh §2.4] to mean the logical symbols, consisting of $=\ \neg\ \vee\ \exists$ and the variables, together with certain nonlogical symbols, consisting of function symbols and predicate symbols each of which has an *index* (or "arity"). It is necessary to distinguish $=$ used as a logical symbol from its customary use in mathematics.

Terms and *formulas* of a language L are defined as in [Sh §2.4]. When infix notation for binary symbols is used to increase readability, we generally use parentheses to group terms and brackets to group formulas.

A *theory* is specified by giving its language and choosing certain formulas as the *nonlogical axioms*; a theory is *open* in case each of the nonlogical axioms is open.

If S is a formal system with finite rules [Sh §1.2], $\pi \vdash_{\mathrm{S}} \mathrm{A}$ means that π is a proof in S of A. We regard a proof as an expression, the concatenation of the formulas in the proof. (An expression in a first-order language can be a concatenation of formulas in only one way.) For any expression u, we denote by $|\mathrm{u}|$ its *expression length*, i.e. the number of occurrences of symbols in it. As we establish cut elimination we also produce bounds on the expression lengths of proofs with cuts eliminated.

2. A formal system

Let T be an open theory, and let T^{ω} be the following formal system. The language of T^{ω} is the language of T. The axioms of T^{ω} are the instances of the open axioms of T; that is [Sh §2.6] of the nonlogical axioms, the identity axioms, the equality axioms, and the propositional axioms $\neg\mathrm{A} \vee \mathrm{A}$ where A is open. The rules of T^{ω} are:

$\vee$-rule: infer B from A if every disjunct of A is a disjunct of B
$\neg\neg$-rule: infer $\neg\neg\mathrm{A} \vee \mathrm{B}$ from $\mathrm{A} \vee \mathrm{B}$
$\neg\vee$-rule: infer $\neg[\mathrm{A} \vee \mathrm{B}] \vee \mathrm{C}$ from $\neg\mathrm{A} \vee \mathrm{C}$ and $\neg\mathrm{B} \vee \mathrm{C}$
$\exists$-rule: infer $\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$ from $\mathrm{A}_{\mathrm{x}}(\mathrm{a}) \vee \mathrm{B}$
$\neg\exists$-rule: infer $\neg\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$ from $\neg\mathrm{A}_{\mathrm{x}}(\mathrm{x}') \vee \mathrm{B}$ if x' is not free in $\neg\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$
cut rule: infer $\mathrm{B} \vee \mathrm{C}$ from $\mathrm{A} \vee \mathrm{B}$ and $\neg\mathrm{A} \vee \mathrm{C}$.

Observe that the axioms of T^ω are theorems of T, by the substitution rule [Sh §3.2], and that the rules of T^ω are derived rules of T, by the tautology theorem [Sh §3.1] and the variant theorem [Sh §3.4]. Hence every theorem of T^ω is a theorem of T. The first result is the converse.

THEOREM A. *Suppose that* $\pi \vdash_T D$. *Then there exists* π' *with* $|\pi'| \leq 4|\pi|^2$ *such that* $\pi' \vdash_{T^\omega} D$.

Proof. Observe [Sh §2.6] that the only axioms of T that are not axioms of T^ω are propositional axioms $\neg A \vee A$ where A is not open and substitution axioms $\neg A_x(a) \vee \exists x A$. Recall that the rules [Sh §2.6] of T are:

expansion:	infer $B \vee A$ from A
contraction:	infer A from $A \vee A$
association:	infer $[A \vee B] \vee C$ from $A \vee [B \vee C]$
cut:	infer $B \vee C$ from $A \vee B$ and $\neg A \vee C$
$\exists$-introduction:	infer $\neg\exists x A \vee B$ from $\neg A \vee B$ if x is not free in B.

Now every rule of T is a rule of T^ω: the expansion, contraction, and association rules are special cases of the $\vee$-rule; the cut rule is a rule of T^ω; and $\exists$-introduction is the special case of the $\neg\exists$-rule in which x' is x. Therefore we need only consider the case that D is a non-open propositional axiom or a substitution axiom of T, and show that there is a proof π' of D in T^ω with $|\pi'| \leq 4|D|^2$ (since the shortest proof of an axiom D is D itself).

I claim that for all A there exists π' with $|\pi'| \leq 4|A|^2$ such that $\pi' \vdash_{T^\omega} \neg A \vee A$. The proof is by induction on the expression length of A.

If A is atomic this is clear since $\neg A \vee A$ is an axiom of T^ω.

Let A be $\neg A_0$ and $\alpha = |A_0|$. By the induction hypothesis there exists π_0 with $|\pi_0| \leq 4\alpha^2$ such that $\pi_0 \vdash_{T^\omega} \neg A_0 \vee A_0$. Let π' be π_0 followed by

$$\begin{aligned} A_0 \vee \neg A_0 \quad & (\vee\text{-rule}) \\ \neg\neg A_0 \vee \neg A_0 \quad & (\neg\neg\text{-rule}). \end{aligned}$$

Then $\pi' \vdash_{T^\omega} \neg A \vee A$ and $|\pi'| \leq 4\alpha^2 + 4\alpha + 6 \leq 4(\alpha+1)^2 = 4|A|^2$.

Let A be $A_0 \vee B_0$, $\alpha = |A_0|$ and $\beta = |B_0|$. By the induction hypothesis there exist π_1 and π_2 with $|\pi_1| \leq 4\alpha^2$ and $|\pi_2| \leq 4\beta^2$ such that $\pi_1 \vdash_{T^\omega} \neg A_0 \vee A_0$ and $\pi_2 \vdash_{T^\omega} \neg B_0 \vee B_0$. Let π' be $\pi_1\pi_2$ followed by

$$\begin{aligned} \neg A_0 \vee A_0 \vee B_0 \quad & (\vee\text{-rule}) \\ \neg B_0 \vee A_0 \vee B_0 \quad & (\vee\text{-rule}) \\ \neg[A_0 \vee B_0] \vee A_0 \vee B_0 \quad & (\neg\vee\text{-rule}). \end{aligned}$$

Then $\pi' \vdash_{\mathrm{T}^\omega} \neg\mathrm{A}\vee\mathrm{A}$ and $|\pi'| \leq 4\alpha^2+4\beta^2+5\alpha+5\beta+10 \leq 4(\alpha+\beta+1)^2 = 4|\mathrm{A}|^2$.

Let A be $\exists\mathrm{x}\mathrm{A}_0$ and $\alpha = |\mathrm{A}_0|$. By the induction hypothesis there exists π_0 with $|\pi_0| \leq 4\alpha^2$ such that $\pi_0 \vdash_{\mathrm{T}^\omega} \neg\mathrm{A}_0 \vee \mathrm{A}_0$. Let π' be π_0 followed by

$$\begin{aligned} \mathrm{A}_0 \vee \neg\mathrm{A}_0 \quad & (\vee\text{-rule}) \\ \exists\mathrm{x}\mathrm{A}_0 \vee \neg\mathrm{A}_0 \quad & (\exists\text{-rule}) \\ \neg\mathrm{A}_0 \vee \exists\mathrm{x}\mathrm{A}_0 \quad & (\vee\text{-rule}) \\ \neg\exists\mathrm{x}\mathrm{A}_0 \vee \exists\mathrm{x}\mathrm{A}_0 \quad & (\neg\exists\text{-rule with x}' \text{ the same as x}). \end{aligned}$$

Then $\pi' \vdash_{\mathrm{T}^\omega} \neg\mathrm{A}\vee\mathrm{A}$ and $|\pi'| \leq 4\alpha^2+8\alpha+16 \leq 4(\alpha+2)^2 = 4|\mathrm{A}|^2$.

This establishes the claim.

Finally, consider a substitution axiom $\neg\mathrm{A}_\mathrm{x}(\mathrm{a}) \vee \exists\mathrm{x}\mathrm{A}$ of T. Let $\alpha = |\mathrm{A}_\mathrm{x}(\mathrm{a})|$ and $\beta = |\exists\mathrm{x}\mathrm{A}|$. By what was just proved, there exists π_0 with $|\pi_0| \leq 4\alpha^2$ such that $\pi_0 \vdash_{\mathrm{T}^\omega} \neg\mathrm{A}_\mathrm{x}(\mathrm{a}) \vee \mathrm{A}_\mathrm{x}(\mathrm{a})$. Let π' be π_0 followed by

$$\begin{aligned} \mathrm{A}_\mathrm{x}(\mathrm{a}) \vee \neg\mathrm{A}_\mathrm{x}(\mathrm{a}) \quad & (\vee\text{-rule}) \\ \exists\mathrm{x}\mathrm{A} \vee \neg\mathrm{A}_\mathrm{x}(\mathrm{a}) \quad & (\exists\text{-rule}) \\ \neg\mathrm{A}_\mathrm{x}(\mathrm{a}) \vee \exists\mathrm{x}\mathrm{A} \quad & (\vee\text{-rule}). \end{aligned}$$

Then $\pi' \vdash_{\mathrm{T}^\omega} \neg\mathrm{A}_\mathrm{x}(\mathrm{a})\vee\exists\mathrm{x}\mathrm{A}$ and $|\pi'| \leq 4\alpha^2+4\alpha+2\beta+6 \leq 4(\alpha+\beta+2)^2 = 4|\neg\mathrm{A}_\mathrm{x}(\mathrm{a}) \vee \exists\mathrm{x}\mathrm{A}|^2$. □

3. Level

The *level* of A, denoted by $\Lambda(\mathrm{A})$, is defined recursively as follows: if A is open, $\Lambda(\mathrm{A}) = 0$; if $\neg\mathrm{A}$ is not open, $\Lambda(\neg\mathrm{A}) = \Lambda(\mathrm{A}) + 1$; if $\mathrm{A}\vee\mathrm{B}$ is not open, $\Lambda(\mathrm{A}\vee\mathrm{B}) = \max\big(\Lambda(\mathrm{A}), \Lambda(\mathrm{B})\big) + 1$; and $\Lambda(\exists\mathrm{x}\mathrm{A}) = \Lambda(\mathrm{A}) + 1$.

Let T^λ be the formal system that is T^ω with the cut rule replaced by the λ-*cut rule*: infer $\mathrm{B}\vee\mathrm{C}$ from $\mathrm{A}\vee\mathrm{B}$ and $\neg\mathrm{A}\vee\mathrm{C}$ provided that $\Lambda(\mathrm{A}) \leq \lambda$. Notice that any proof in T^λ is a proof in T^ω, and that for any proof in T^ω there is a λ such that it is a proof in T^λ. We shall show, in Theorem L below, that every theorem of T^ω (and hence every theorem of T) is a theorem of T^0.

4. Tree proofs

Let S be a formal system with finite rules. By a *tree proof* in S we mean a proof in S such that each occurrence of a formula, except the last, is used exactly once as a premise to a rule to infer a later formula. At the cost of repetition and an exponential growth in the length of a proof, we can always replace a proof in S by a tree proof, as follows. Start with the last formula and picture it as the root of a tree, at the

bottom. If it is not an axiom, list immediately above it the formulas from which it is inferred by a rule, and repeat. In this way we obtain a tree whose root is the theorem and whose leaves are occurrences of axioms. The portion of the proof lying strictly above an occurrence of a formula will be called the *ancestry* of that occurrence. By $\tau\colon \vdash_{\mathrm{S}} \mathrm{D}$ we mean that τ is a tree proof in S of D.

Theorems concerning tree proofs are frequently proved by *downward tree induction*. Let $\tau\colon \vdash_{\mathrm{S}} \mathrm{D}$; we want to prove that D has the property P. We do so by showing that each axiom has the property P and for each rule, if the premises have the property P so does the conclusion. There is also *upward tree induction*. Again let $\tau\colon \vdash_{\mathrm{S}} \mathrm{D}$; we want to prove that every formula of τ has the property P. We do so by showing that D has the property P and for each rule, if the conclusion has the property P so do the premises.

THEOREM B. *Suppose that* $\pi \vdash_{\mathrm{T}^\lambda} \mathrm{D}$. *Then there exists* τ *with* $|\tau| \leq 2^{|\pi|}$ *such that* $\tau\colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$.

Proof. Construct, as above, the tree proof τ corresponding to π. Since each rule of T^λ has at most two premises, $|\tau| \leq 2^{|\pi|}$. □

Consider a tree proof τ in T^λ, and let the variables $\mathrm{x}_1, \ldots, \mathrm{x}_\nu$ be given. We say that τ is *adjusted for* $\mathrm{x}_1, \ldots, \mathrm{x}_\nu$ in case for each occurrence of a formula $\neg\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$ inferred by the $\neg\exists$-rule from $\neg \mathrm{A}_{\mathrm{x}}(\mathrm{x}') \vee \mathrm{B}$ where x' is not free in $\neg\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$, the variable x' occurs only in the ancestry of that occurrence, occurs only free, and is not one of $\mathrm{x}_1, \ldots, \mathrm{x}_\nu$; then we call x' a *restricted variable* for τ. We say that τ is *adjusted* in case it is adjusted for some sequence of variables, possibly empty.

THEOREM C. *Suppose that* $\tau\colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$ *and let* $\mathrm{x}_1, \ldots, \mathrm{x}_\nu$ *be given. Then there exists* τ' *with* $|\tau'| \leq |\tau|$ *such that* $\tau'\colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$ *and* τ' *is adjusted for* $\mathrm{x}_1, \ldots, \mathrm{x}_\nu$.

Proof. For each occurrence in τ of a formula $\neg\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$ inferred by the $\neg\exists$-rule from $\neg \mathrm{A}_{\mathrm{x}}(\mathrm{x}') \vee \mathrm{B}$ where x' is not free in $\neg\exists \mathrm{x}\mathrm{A} \vee \mathrm{B}$, replace all free occurrences of x' in the ancestry of that occurrence by x'', where x'' is distinct from $\mathrm{x}_1, \ldots, \mathrm{x}_\nu$ and occurs nowhere else in the proof. □

We say that τ is a *first* $(\lambda+1)$*-cut* in case $\tau\colon \vdash_{\mathrm{T}^{\lambda+1}} \mathrm{B} \vee \mathrm{C}$ where $\mathrm{B} \vee \mathrm{C}$ is inferred by a cut with cut formula A of level $\lambda+1$ and there is no $(\lambda+1)$-cut in the ancestry of $\mathrm{B} \vee \mathrm{C}$. That is, we have

(1) $\tau\colon \vdash_{\mathrm{T}^{\lambda+1}} \mathrm{B} \vee \mathrm{C}$

(2) $\tau_1\colon \vdash_{\mathrm{T}^\lambda} \mathrm{A} \vee \mathrm{B}$

(3) $\tau_2\colon \vdash_{\mathrm{T}^\lambda} \neg\mathrm{A} \vee \mathrm{C}$

where $\Lambda(\mathrm{A}) = \lambda + 1$, τ_1 is the restriction of τ to $\mathrm{A} \vee \mathrm{B}$ and its ancestry, and τ_2 is the restriction of τ to $\neg\mathrm{A} \vee \mathrm{C}$ and its ancestry. There are three cases:

(i) A is a negation $\neg\mathrm{A}_0$
(ii) A is a disjunction $\mathrm{A}_0 \vee \mathrm{A}_1$
(iii) A is an instantiation $\exists \mathrm{x} \mathrm{A}_0$.

We treat these cases in turn.

5. Negations

In case (i), $\neg\mathrm{A}$ is $\neg\neg\mathrm{A}_0$. First we prove a lemma, Theorem D, showing that $\tau_2 \colon \vdash_{\mathrm{T}^\lambda} \neg\neg\mathrm{A}_0 \vee \mathrm{C}$ can be replaced by $\tau_2' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{A}_0 \vee \mathrm{C}$ with $|\tau_2'| \leq |\tau_2|$ and then in Theorem E we use A_0, which is of level λ, as the cut formula.

THEOREM D. *Let $\neg\neg\mathrm{A}_0$ not be open, and let D' be obtained from* D *by replacing an occurrence of $\neg\neg\mathrm{A}_0$ as a disjunct of* D *by* A_0. *If $\tau \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$ then there exists τ' with $|\tau'| \leq |\tau|$ such that $\tau' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}'$.*

Proof. For the purposes of this proof, a *slice* of a formula D is a formula obtained by replacing zero or one occurrences of $\neg\neg\mathrm{A}_0$ as a disjunct of D by A_0. We shall prove by downward tree induction that if $\tau \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$ and D' is a slice of D, there exists τ' with $|\tau'| \leq |\tau|$ such that $\tau' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}'$.

If D is an axiom of T^λ it is open but $\neg\neg\mathrm{A}_0$ is not, so the only slice of D is D itself. If D is inferred from D_1 (or from D_1 and D_2) by any rule other than the $\neg\neg$-rule, a slice D' of D can be inferred from a slice D_1' of D_1 (or slices D_1' and D_2' of D_1 and D_2) by the same rule.

If D is $\neg\neg\mathrm{A}_1 \vee \mathrm{C}$ and is inferred by the $\neg\neg$-rule from $\mathrm{A}_1 \vee \mathrm{C}$, and D' is a slice of D, there are two cases: either D' is $\neg\neg\mathrm{A}_1 \vee \mathrm{C}'$ where C' is a slice of C, or else A_1 is A_0 and D' is $\mathrm{A}_0 \vee \mathrm{C}$. In the first case D' can be inferred by the $\neg\neg$-rule from the slice $\mathrm{A}_1 \vee \mathrm{C}'$ of $\mathrm{A}_1 \vee \mathrm{C}$, and in the second case we already have D'. □

THEOREM E. *Let $\tau \colon \vdash_{\mathrm{T}^{\lambda+1}} \mathrm{B} \vee \mathrm{C}$ be a first $(\lambda+1)$-cut, with cut formula $\neg\mathrm{A}_0$. Then there exists τ' with $|\tau'| \leq |\tau|$ such that $\tau' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{B} \vee \mathrm{C}$.*

Proof. We use the notation of (1)–(3). By Theorem D there exists τ_2' with $|\tau_2'| \leq |\tau_2|$ such that $\tau_2' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{A}_0 \vee \mathrm{C}$. Then $\Lambda(\mathrm{A}_0) = \lambda$, so $\mathrm{B} \vee \mathrm{C}$ can be inferred from $\mathrm{A}_0 \vee \mathrm{C}$ and $\neg\mathrm{A}_0 \vee \mathrm{B}$ by the λ-cut rule. □

6. Disjunctions

In case (ii), A is $A_0 \vee A_1$ and $\neg A$ is $\neg[A_0 \vee A_1]$. First we prove a lemma, Theorem F, showing that there exist τ_2' and τ_2'', with $|\tau_2'| \leq |\tau_2|$ and $|\tau_2''| \leq |\tau_2|$, such that $\tau_2' \colon \vdash_{T^\lambda} \neg A_0 \vee C$ and $\tau_2'' \colon \vdash_{T^\lambda} \neg A_1 \vee C$. Then in Theorem G we eliminate the $(\lambda+1)$-cut by λ-cuts with cut formulas A_0 and A_1.

THEOREM F. *Let* $\neg[A_0 \vee A_1]$ *not be open, and let* D' *be obtained from* D *by replacing an occurrence of* $\neg[A_0 \vee A_1]$ *as a disjunct of* D *by* $\neg A_0$*, or by* $\neg A_1$*. If* $\tau \colon \vdash_{T^\lambda} D$*, there exists* τ' *with* $|\tau'| \leq |\tau|$ *such that* $\tau' \colon \vdash_{T^\lambda} D'$*.*

Proof. For the purposes of this proof, a *slice* of a formula D is a formula obtained by replacing zero or one occurrences of $\neg[A_0 \vee A_1]$ as a disjunct of D by $\neg A_0$. We shall prove by downward tree induction that if $\tau \colon \vdash_{T^\lambda} D$ and D' is a slice of D, there exists τ' with $|\tau'| \leq |\tau|$ such that $\tau' \colon \vdash_{T^\lambda} D'$. (The proof for $\neg A_1$ is entirely similar.)

If D is an axiom of T^λ it is open but $\neg[A_0 \vee A_1]$ is not, so the only slice of D is D itself. If D is inferred from D_1 (or from D_1 and D_2) by any rule other than the $\neg\vee$-rule, a slice D' of D can be inferred from a slice D_1' of D_1 (or slices D_1' and D_2' of D_1 and D_2) by the same rule.

If D is $\neg[A_2 \vee A_3] \vee C$ and is inferred from $\neg A_2 \vee C$ and $\neg A_3 \vee C$ by the $\neg\vee$-rule, and D' is a slice of D, there are two cases: either D' is $\neg[A_2 \vee A_3] \vee C'$ where C' is a slice of C, or else A_2 is A_0, A_3 is A_1, and D' is $\neg A_0 \vee C$. In the first case D' can be inferred from the slices $\neg A_2 \vee C'$ and $\neg A_3 \vee C'$ of $\neg A_2 \vee C$ and $\neg A_3 \vee C$, and in the second case we already have D'. □

THEOREM G. *Let* $\tau \colon \vdash_{T^{\lambda+1}} B \vee C$ *be a first* $(\lambda+1)$*-cut, with cut formula* $A_0 \vee A_1$*. Then there exists* τ' *with* $|\tau'| \leq 6|\tau|$ *such that we have* $\tau' \colon \vdash_{T^\lambda} B \vee C$*.*

Proof. We use the notation of (1)–(3). By Theorem F there exist τ_2' and τ_2'', with $|\tau_2'| \leq |\tau_2|$ and $|\tau_2''| \leq |\tau_2|$, such that $\tau_2' \colon \vdash_{T^\lambda} \neg A_0 \vee C$ and $\tau_2'' \colon \vdash_{T^\lambda} \neg A_1 \vee C$. Let τ' be constructed as follows:

$[A_0 \vee A_1] \vee B$	(τ_1)
$A_0 \vee [A_1 \vee B]$	($\vee$-rule)
$\neg A_0 \vee C$	(τ_2')
$[A_1 \vee B] \vee C$	(λ-cut rule, since $\Lambda(A_0) \leq \lambda$)
$A_1 \vee [B \vee C]$	($\vee$-rule)
$\neg A_1 \vee C$	(τ_2'')
$[B \vee C] \vee C$	(λ-cut rule, since $\Lambda(A_1) \leq \lambda$)
$B \vee C$	($\vee$-rule).

Then crudely estimating the expression lengths of the extra formulas by the expression lengths of their proofs, we obtain $|\tau'| \leq 6|\tau|$. □

7. Instantiations

Case (iii) is the unkindest cut of all. It is the source of the explosion in expression length. If the cut formula is $\exists \mathrm{x} \mathrm{A}_0$, its negation is $\neg\exists \mathrm{x} \mathrm{A}_0$. We prove two lemmas. The first, Theorem H, shows that $\tau_2 \colon \vdash_{\mathrm{T}^\lambda} \neg\exists \mathrm{x} \mathrm{A}_0 \vee \mathrm{C}$ can be replaced by $\tau_2' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{A}_{0\mathrm{x}}(\mathrm{x}') \vee \mathrm{C}$ where x' is a new variable. The second, Theorem I, shows that we can replace x' in τ_2' by any term a. Then in Theorem J we eliminate a first $(\lambda+1)$-cut using many λ-cuts with cut formulas that are instances of A_0. But we do so at the cost of additional grafts onto the tree and complicating the ancestry with new and longer formulas that may produce new instances in further cut eliminations. Semantically, an instantiation is the infinite disjunction of all instances. What Gentzen demonstrated is that in any proof, cuts by instantiations can be replaced by finitely many cuts by instances, but these may be instances that are only *implicit* in the proof.

THEOREM H. *Let* D' *be obtained from* D *by replacing an occurrence of a disjunct* $\neg\exists \mathrm{x} \mathrm{A}_0$ *of* D *by* $\neg \mathrm{A}_{0\mathrm{x}}(\mathrm{x}')$*, where* x' *is not free in* D*. If* $\tau \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$*, then there exists* τ' *with* $|\tau'| \leq |\tau|$ *such that* $\tau' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}'$*.*

Proof. We assume that x' occurs nowhere in τ (if it does, replace it everywhere by another variable), and we assume by Theorem C that τ is adjusted. For the purposes of this proof, a *slice* of a formula D is a formula obtained by replacing zero or one occurrences of $\neg\exists \mathrm{x} \mathrm{A}_0$ as a disjunct of D by $\neg \mathrm{A}_{0\mathrm{x}}(\mathrm{x}')$. We shall prove by downward tree induction that if $\tau \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}$ and D' is a slice of D, there exists τ' with $|\tau'| \leq |\tau|$ such that $\tau' \colon \vdash_{\mathrm{T}^\lambda} \mathrm{D}'$.

If D is an axiom of T^λ it is open, so the only slice of D is D itself. If D is inferred from D_1 (or from D_1 and D_2) by any rule other than the $\neg\exists$-rule, a slice D' of D can be inferred from a slice D_1' of D_1 (or slices D_1' and D_2' of D_1 and D_2) by the same rule.

If D is $\neg\exists \mathrm{x}_1 \mathrm{A}_1 \vee \mathrm{C}$ and is inferred by the $\neg\exists$-rule from $\neg \mathrm{A}_{1\mathrm{x}_1}(\mathrm{x}'') \vee \mathrm{C}$ where x'' is not free in D, and D' is a slice of D, there are two cases: either D' is $\neg\exists \mathrm{x}_1 \mathrm{A}_1 \vee \mathrm{C}'$ where C' is a slice of C, or else x_1 is x, A_1 is A_0, and D' is $\mathrm{A}_{0\mathrm{x}}(\mathrm{x}'') \vee \mathrm{C}$ where the restricted variable x'' is not free in D. In the first case, D' can be inferred by the $\neg\exists$-rule from the slice $\neg \mathrm{A}_{0\mathrm{x}}(\mathrm{x}') \vee \mathrm{C}'$ of $\neg \mathrm{A}_{0\mathrm{x}}(\mathrm{x}') \vee \mathrm{C}$. In the second case, since τ is adjusted we can replace x'' everywhere by x'; then we already have D'. □

THEOREM I. If $\tau\colon \vdash_{T^\lambda} D$ then there exists τ' with $|\tau'| \leq |\tau||a|$ such that $\tau'\colon \vdash_{T^\lambda} D_x(a)$.

Proof. By Theorem C we assume that τ is adjusted for the variables in a. Construct τ' by replacing each C in τ by $C_x(a)$; since there are at most $|\tau|$ occurrences of x in τ, the bound on $|\tau'|$ is clear.

We argue by downward tree induction. If C is an axiom of T^λ, so is $C_x(a)$. If C is inferred from C_1 (or from C_1 and C_2) by any rule, then $C_x(a)$ is inferred from $C_{1x}(a)$ $\big($or from $C_{1x}(a)$ and $C_{2x}(a)\big)$ by the same rule—for the $\neg\exists$-rule, this holds because τ is adjusted for the variables in a. □

THEOREM J. *Let* $\tau\colon \vdash_{T^{\lambda+1}} B \vee C$ *be a first* $(\lambda+1)$*-cut, with cut formula* $\exists x A_0$. *There exists* τ' *with* $|\tau'| \leq |\tau|^2$ *such that* $\tau'\colon \vdash_{T^\lambda} B \vee C$.

Proof. We use the notation of (1)–(3). Thus we have

(1) $\tau\colon \vdash_{T^{\lambda+1}} B \vee C$

(2) $\tau_1\colon \vdash_{T^\lambda} \exists x A_0 \vee B$

(3) $\tau_2\colon \vdash_{T^\lambda} \neg\exists x A_0 \vee C$

where $\Lambda(\exists x A_0) = \lambda + 1$, τ_1 is the restriction of τ to $\exists x A_0 \vee B$ and its ancestry, and τ_2 is the restriction of τ to $\neg\exists x A_0 \vee C$ and its ancestry. Hence

(4) $|\tau| = |\tau_1| + |\tau_2| + |B \vee C|$.

To simplify the arithmetic of estimating expression lengths, let $\hat{T}^\lambda$ be T^λ with an additional rule of inference, the *second form of the* λ*-cut rule*: infer $C \vee B$ from $A \vee B$ and $\neg A \vee C$ provided that $\Lambda(A) \leq \lambda$. Clearly this rule follows from the λ-cut rule and the $\vee$-rule, and if $\hat{\tau}\colon \vdash_{\hat{T}^\lambda} D$ then there exists τ with $|\tau| \leq 2|\hat{\tau}|$ such that $\tau\colon \vdash_{T^\lambda} D$.

For the purposes of this proof, a *slice* D' of a formula D is a formula obtained by replacing zero or one occurrences of $\exists x A_0$ as a disjunct of D by C.

I claim that if $\sigma\colon \vdash_{T^\lambda} D$ and D' is a slice of D, there exists σ' with $|\sigma'| \leq 2(|\sigma||C| + |\tau_2||\sigma|)$ such that $\sigma'\colon \vdash_{T^\lambda} D'$. The proof is by downward tree induction.

If D is an axiom of T^λ it is open, so the only slice of D is D itself. If D is inferred from D_1 (or from D_1 and D_2) by any rule other than the $\exists$-rule, a slice D' of D can be inferred from a slice D'_1 of D_1 (or slices D'_1 and D'_2 of D_1 and D_2) by the same rule. Notice that each replacement of a formula D by a slice D' adds at most |C| to the expression length, so these replacements add at most $|\sigma||C|$ to the expression length.

If D is $\exists x_1 A_1 \vee B_1$ and is inferred by the $\exists$-rule from $A_{x_1}(a) \vee B_1$, and D' is a slice of D, there are two cases: either D' is $\exists x_1 A_1 \vee B_1'$ where B_1' is a slice of B_1, or else x_1 is x and A_1 is A_0, and D' is $C \vee B_1$. In the first case, D' can be inferred by the $\exists$-rule from the slice $A_{x_1}(a) \vee B_1'$ of $A_{x_1}(a) \vee B_1$, again at the cost of adding at most $|C|$ to the expression length.

In the second case, where D is $A_{0x}(a) \vee B_1$, we argue as follows. By Theorem H, there exists τ_2' with $|\tau_2'| \leq |\tau_2|$ such that we have $\tau_2' \colon \vdash_{T^\lambda} \neg A_{0x}(x') \vee C$ where x' is not free in $\neg \exists x A_0 \vee C$. By Theorem I, there exists τ_2'' with $|\tau_2''| \leq |\tau_2||a|$ such that $\tau_2'' \colon \vdash_{T^\lambda} \neg A_{0x}(a) \vee C$. Since $\Lambda\big(A_{0x}(a)\big) = \lambda$, we infer $C \vee B_1$ (i.e. D') by the second form of the λ-cut rule. This adds at most $|\tau_2||a|$ to the length of the proof, and since the total length of such occurrences of terms a is at most $|\sigma|$, we have a proof σ' of length at most $2(|\sigma||C| + |\tau_2||\sigma|)$, as claimed. (The factor of 2 comes from our having used the second form of the λ-cut rule.)

Apply this to the case that σ is τ_1. Then we have τ_1' such that $\tau_1 \colon \vdash_{T^\lambda} C \vee B$. Let τ' be τ_1' followed by $B \vee C$ ($\vee$-rule). Then we have $|\tau'| \leq 2(|\tau_1||C| + |\tau_2||\tau_1|) + |B \vee C| \leq |\tau|^2$ by (4). □

8. Cut elimination

THEOREM K. *Suppose that* $\tau \colon \vdash_{T^{\lambda+1}} D$. *Then there exists* τ' *with* $|\tau'| \leq |\tau|^{2^{|\tau|}}$ *such that* $\tau' \colon \vdash_{T^\lambda} D$.

Proof. By Theorems E, G, and J, each first inference by the $(\lambda+1)$-cut rule in τ can be replaced by λ-cuts, at the cost of squaring the length of the proof. But there are at most $|\tau|$ such inferences. □

Let $\Sigma(\lambda) = \lambda^{2^\lambda}$, let Σ^α be the α-fold composition of Σ with itself, and let $\Phi(\lambda) = \Sigma^\lambda(\lambda)$.

THEOREM L. *Suppose that* $\pi \vdash_T D$. *Then there exists* τ *such that* $|\tau| \leq \Phi(2^{4|\pi|^2})$ *and* $\tau \colon \vdash_{T^0} D$.

Proof. Let $\lambda = 2^{4|\pi|^2}$. By Theorems A and B, there exists τ_0 with $|\tau_0| \leq \lambda$ such that $\tau_0 \colon \vdash_{T^\omega} D$. The maximum level of a formula in τ_0 is at most λ, so $\tau_0 \colon \vdash_{T^\lambda} D$. Apply Theorem K λ times to obtain a tree proof τ such that $\tau \colon \vdash_{T^0} D$. □

9. Consequences of cut elimination

We say that A is a *subformula* of B in case A occurs in B. (This terminology differs from the usage in [Kl §78].) An occurrence of a subformula is *outer* in case it is not in the scope of any quantifier. We

say that A is *subordinate* to B in case A is obtained from B by applying the following steps zero or more times: replace an outer occurrence of a subformula by a subformula of it; replace an outer occurrence of $\exists$xC by an instance of C. Finally, we say that A is *dominated* by B (or that B *dominates* A) in case every non-open disjunct of A is subordinate to B.

THEOREM M. *If* τ: $\vdash_{T^0}$ D *then every formula in* τ *is dominated by* D.

Proof. Three preliminary observations:

(1) if A is subordinate to B and B is subordinate to C, then A is subordinate to C;

(2) if A is a subformula of B, then A is subordinate to B;

(3) if A is subordinate to B, then A is dominated by B.

The property (1) is immediate from the definition. For (2), replace the occurrence of B in itself by A. For (3), observe that every disjunct of A is a subformula of A, so it is subordinate to A by (2) and so is subordinate to B by (1).

We argue by upward tree induction. We need to show for each rule of T^0 that the conclusion dominates the premise or premises.

If B is inferred from A by the $\vee$-rule, then every disjunct of A is a subformula of B and so is subordinate to it by (2). Hence A is dominated by B.

If $\neg\neg$A $\vee$ B is inferred from A $\vee$ B by the $\neg\neg$-rule, then A $\vee$ B is subordinate to $\neg\neg$A $\vee$ B (replace the occurrence of $\neg\neg$A by A), so it is dominated by it, by (3).

If $\neg$[A $\vee$ B] $\vee$ C is inferred from $\neg$A $\vee$ C and $\neg$B $\vee$ C by the $\neg\vee$-rule, then $\neg$A $\vee$ C is subordinate to $\neg$[A $\vee$ B] $\vee$ C (replace the occurrence of A $\vee$ B by A), and so is $\neg$B $\vee$ C (replace the occurrence of A $\vee$ B by B), so they are dominated by it, by (3).

If $\exists$xA $\vee$ B is inferred from $A_x(a)$ $\vee$ B by the $\exists$-rule, then $A_x(a)$ is subordinate to $\exists$xA $\vee$ B $\big($replace it by $\exists$xA and then by $A_x(a)\big)$, and so is B. Any disjunct of $A_x(a)$ $\vee$ B is a disjunct of one of these formulas, so is subordinate to it by (2), and so to $\exists$xA $\vee$ B by (1). Hence $A_x(a)$ $\vee$ B is dominated by $\exists$xA $\vee$ B.

If $\neg\exists$xA $\vee$ B is inferred from $\neg A_x(x')$ $\vee$ B (where x$'$ is not free in $\neg\exists$xA$\vee$B, though that is not relevant here) by the $\neg\exists$-rule, then $\neg A_x(x')$ is subordinate to $\neg\exists$xA $\vee$ B (replace it by $\neg\exists$xA and then replace the occurrence of $\exists$xA by $A_x(x')$), and so is B, so as before the premise is dominated by the conclusion.

If B $\vee$ C is inferred from A $\vee$ B and $\neg$A $\vee$ C by the 0-cut rule, then every non-open disjunct of A$\vee$B is a disjunct of B and so is subordinate

to the conclusion, and similarly for $\neg A \vee C$. Hence both premises are dominated by the conclusion. □

A corollary is that if $\tau\colon \vdash_{T^0} D$ and D is open, then every formula in τ is open. The Hilbert-Ackermann consistency theorem [Sh §4.3] follows immediately, as we now show.

A *quasi-tautology* is a tautological consequence of instances of identity and equality axioms. We say that A is a *quasi-tautological consequence* of $B_1, \ldots, B_\nu$ in case $B_1 \to \cdots \to B_\nu \to A$ is a quasi-tautology.

THEOREM N. *Let* T *be an open theory,* D *an open formula of* T. *Then* $\vdash_T D$ *if and only if* D *is a quasi-tautological consequence of instances of nonlogical axioms of* T. *The theory* T *is inconsistent if and only if there is a quasi-tautology that is a disjunction of instances of negations of nonlogical axioms.*

Proof. Suppose that $\vdash_T D$. By Theorems A and L there exists τ such that $\tau\colon \vdash_{T^0} D$, and by Theorem M every formula in τ is open. Hence inferences by the $\exists$-rule and the $\neg\exists$-rule do not occur. The other rules are special cases of tautological consequence. Therefore D is a tautological consequence of instances of nonlogical axioms and of identity and equality axioms. (Propositional axioms are tautologies.) This proves the first statement, since the converse is trivial. Now T is inconsistent if and only if $\vdash_T x \neq x$, so the second statement follows. □

THEOREM O. *Let* $\tau\colon \vdash_{T^0} D$ *where* D *is an existential prenex formula,* $\exists x_1 \ldots \exists x_\mu B$ *with* B *open. Then every formula in* τ *is a disjunction of formulas of the form*

$$\exists x_\alpha \ldots \exists x_\mu B'$$

where B′ *is open and* $1 \leq \alpha \leq \mu + 1$.

Proof. Here $\alpha = \mu + 1$ means, of course, that the disjunct is open. The result follows from Theorem M since every subordinate formula is of the displayed form. □

THEOREM P. *Let* T *be an open theory and let* D *be an existential prenex formula. Then* $\vdash_T D$ *if and only if there is a disjunction of instances of the matrix of* D *that is a quasi-tautological consequence of instances of nonlogical axioms.*

Proof. Suppose that $\vdash_T D$. By Theorems A and L there exists τ such that $\tau\colon \vdash_{T^0} D$. By Theorem O there is no inference by the $\neg\exists$-rule in τ. Then the result holds by downward tree induction. The converse is trivial. □

Let D be a prenex formula [Sh §3.5] of a theory T. We construct an existential prenex formula $\mathrm{D_H}$, the *Herbrand form* of D, as follows. Let $\mathrm{x}_1, \ldots, \mathrm{x}_\mu$ be the existentially quantified variables and $\mathrm{y}_1, \ldots, \mathrm{y}_\nu$ the universally quantified variables, each sequence in the order of occurrence in D. For each β with $1 \leq \beta \leq \nu$, let $\mathrm{x}_1, \ldots, \mathrm{x}_{\iota(\beta)}$ be the x-variables preceding y_β in D and let f_β be a new function symbol of index $\iota(\beta)$. Let T′ be T with the new function symbols adjoined and let $\mathrm{D_H}$ be

$$\exists \mathrm{x}_1 \ldots \exists \mathrm{x}_\mu \mathrm{D}^\circ_{\mathrm{y}_1,\ldots,\mathrm{y}_\nu}\big(\mathrm{f}_1(\mathrm{x}_1,\ldots,\mathrm{x}_{\iota(1)}),\ldots,\mathrm{f}_\nu(\mathrm{x}_1,\ldots,\mathrm{x}_{\iota(\nu)})\big)$$

where D° is the matrix of D.

Next we prove a lemma that has nothing to do with cut elimination.

THEOREM Q, *Let* D *be a prenex formula of a theory* T *and let* T′ *and* $\mathrm{D_H}$ *be as above. Then* $\vdash_{\mathrm{T}}$ D *if and only if* $\vdash_{\mathrm{T}'}$ $\mathrm{D_H}$.

Proof. Suppose that $\vdash_{\mathrm{T}}$ D. Then certainly $\vdash_{\mathrm{T}'}$ D. Working from right to left with repeated applications of the substitution theorem and the distribution rule [Sh §3.2] we have $\vdash_{\mathrm{T}'}$ $\mathrm{D} \to \mathrm{D_H}$ and consequently $\vdash_{\mathrm{T}'}$ $\mathrm{D_H}$.

Conversely, suppose that $\vdash_{\mathrm{T}'}$ $\mathrm{D_H}$. By the theorem on constants [Sh §3.3] we assume without loss of generality that D is closed. We work in the theory $\mathrm{T_c}$ [Sh §4.2] with special constants and special axioms adjoined. By the *special constant for a closed generalization* $\forall \mathrm{y}\mathrm{C}$ we mean the special constant s for $\exists \mathrm{y} \neg \mathrm{C}$; then the special axiom $\exists \mathrm{y} \neg \mathrm{C} \to \neg \mathrm{C}_\mathrm{y}(\mathrm{s})$ is equivalent to $\mathrm{C}_\mathrm{y}(\mathrm{s}) \to \forall \mathrm{y}\mathrm{C}$. Form D′, a formula of $\mathrm{T_c}$, by working from left to right: delete the first quantifier and replace its variable by the corresponding special constant, and repeat. Then

(1) $\qquad \vdash_{\mathrm{T_c}} [\mathrm{D} \leftrightarrow \mathrm{D}']$.

Also, let D″ be

$$\exists \mathrm{x}_1 \ldots \exists \mathrm{x}_\mu \mathrm{D}^\circ_{\mathrm{y}_1,\ldots,\mathrm{y}_\nu}(\mathrm{s}_1,\ldots,\mathrm{s}_\nu)$$

where the s_β are the special constants just introduced corresponding to the $\forall \mathrm{y}_\beta$. Then

(2) $\qquad \vdash_{\mathrm{T_c}} [\mathrm{D}' \leftrightarrow \mathrm{D}'']$

by μ applications of the distribution rule. Since $\mathrm{T_c}$ is a conservative extension of T it suffices, by (1) and (2), to show that $\vdash_{\mathrm{T_c}}$ D″. But this is easy: in the proof of $\mathrm{D_H}$ in T′, replace each occurrence of a term beginning with f_β by s_β, for all β with $1 \leq \beta \leq \nu$. By induction on theorems [Sh §3.1] this gives a proof of D″ in $\mathrm{T_c}$ (without, in fact, using special axioms). □

Now we have Herbrand's theorem [Sh §4.4].

THEOREM R. *Let* D *be a prenex formula in an open theory* T. *Then* $\vdash_{\mathrm{T}}$ D *if and only if there is a disjunction of instances of the matrix of* $\mathrm{D_H}$ *that is a quasi-tautological consequence of instances of nonlogical axioms.*

Proof. By Theorems Q and P. □

10. Infeasibility of cut elimination

Let Z_0 be the theory whose nonlogical symbols are the constant 0 and the unary function symbol S, and whose nonlogical axioms are

(1) $\mathrm{S}x \neq 0$
(2) $\mathrm{S}x = \mathrm{S}y \to x = y$.

Then Z_0 has a model σ, the *standard model*, with universe $\mathbb{N}$, $\sigma(0) = 0$, and $\sigma(\mathrm{S}) = \mathrm{S}$ (the successor function). Hence Z_0 is consistent. This is an infinitary consistency proof; notice that there is no finite model of Z_0.

The only variable-free terms of Z_0 are *numerals* S...S0. Hence the only variable-free atomic formulas are equalities a = b where a and b are numerals. Define a truth valuation $\mathcal{V}$ [Sh §3.1] on the elementary formulas by setting $\mathcal{V}(\mathrm{A}) = \mathsf{T}$ if and only if A is a = b where a and b are identical numerals. Extend $\mathcal{V}$ to all formulas in the usual way. Notice that if A is a variable-free instance of (1), (2), an identity axiom x = x, or an equality axiom x = y → Sx = Sy, then $\mathcal{V}(\mathrm{A}) = \mathsf{T}$. Now suppose that Z_0 is inconsistent. By Theorem N there is a tautology B that is a disjunction of negations of instances of such axioms. Substituting 0 for all variables we may suppose that B is variable-free. But this is impossible since $\mathcal{V}(\mathrm{B}) = \mathsf{F}$. This is a finitary consistency proof for Z_0.

For ν in $\mathbb{N}$, let $\bar{\nu}$ be the numeral with ν occurrences of S. Let T be a consistent open extension of Z_0 and let b be a variable-free term of T. If $\vdash_{\mathrm{T}}$ b = $\bar{\nu}$ we say that b *denotes* ν in T; then ν is unique by (1), (2), and the consistency of T. Let $\hat{\mathrm{T}}$ be obtained by adjoining a unary predicate symbol φ and the axiom

(0) $\varphi(0)$ & $[\varphi(x) \to \varphi(\mathrm{S}x)]$.

Then $\hat{\mathrm{T}}$ is also consistent: in any proof replace each φ(a) by a = a.

THEOREM S. *Let* T *be a consistent open extension of* Z_0. *If* b *denotes* ν *and* $\pi \vdash_{\hat{\mathrm{T}}^0} \varphi(\mathrm{b})$ *then* $|\pi| \geq \nu$.

Proof. We do not assume that π is a tree proof. But let π' be obtained from π by repeatedly deleting all occurrences of formulas, except the last, that are not used as a premise in a rule to infer a later formula. Then $|\pi'| \leq |\pi|$ and π' is still a proof of $\varphi(\mathrm{b})$, and it has the same formulas as the corresponding tree proof, so by Theorem M every formula in π' is open. Substituting 0 for all occurrences of variables we may assume that every formula in π' is variable-free. Then inferences by the $\exists$-rule and the $\neg\exists$-rule do not occur, so $\varphi(\mathrm{b})$ is a tautological consequence of variable-free instances $\mathrm{A}_1, \ldots, \mathrm{A}_\alpha$ of (0) and of equality axioms $x = y \to \varphi(x) \to \varphi(y)$, and variable-free instances $\mathrm{B}_1, \ldots, \mathrm{B}_\beta$ of nonlogical axioms of T, identity axioms, and equality axioms for the nonlogical symbols of T. Now there is a truth valuation $\mathcal{V}$ on the formulas of T that assigns T to all the B's and assigns F to $\mathrm{a} = \mathrm{b}$ whenever a and b denote different numbers—otherwise T would be inconsistent.

Now suppose that $|\pi| < \nu$, so $|\pi'| < \nu$. Then there is a $\gamma < \nu$ such that no term occurring in π' denotes γ. Extend $\mathcal{V}$ to the formulas of $\hat{\mathrm{T}}$ by setting $\mathcal{V}(\mathrm{A})$, where A is an elementary formula of $\hat{\mathrm{T}}$ but not T, equal to T if and only if A is $\varphi(\mathrm{a})$ where a denotes a number $< \gamma$. Then the extended $\mathcal{V}$ assigns T to each of the A's but it assigns F to $\varphi(\mathrm{b})$, which is a contradiction. □

Now we shall show how to produce short proofs of $\varphi(\mathrm{b})$ for b denoting quite large numbers.

Let Z_1 be the extension of Z_0 obtained by adjoining binary function symbols $+$ and $\cdot$ and the axioms

$$x + 0 = x \tag{3}$$
$$x + \mathrm{S}y = \mathrm{S}(x + y) \tag{4}$$
$$x \cdot 0 = 0 \tag{5}$$
$$x \cdot \mathrm{S}y = x \cdot y + x \tag{6}$$
$$(x + y) + z = x + (y + z) \tag{7}$$
$$x \cdot (y + z) = x \cdot y + x \cdot z \tag{8}$$
$$(x \cdot y) \cdot z = x \cdot (y \cdot z) \tag{9}$$
$$x + y = y + x \tag{10}$$
$$x \cdot y = y \cdot x. \tag{11}$$

We have adopted the usual convention that $\cdot$ binds more tightly than $+$.

A formula C is called *unary* in case if has one and only one free variable; then we write $\mathrm{C}(\mathrm{a})$ for $\mathrm{C}_{\mathrm{x}}(\mathrm{a})$ where x is the free variable of C. We use the abbreviations:

$$\mathrm{C}^2 \quad \text{for} \quad \forall y[\mathrm{C}(y) \to \mathrm{C}(y + x)] \tag{12}$$
$$\mathrm{C}^3 \quad \text{for} \quad \forall y[\mathrm{C}^2(y) \to \mathrm{C}^2(y \cdot x)]. \tag{13}$$

THEOREM T. *Let* C *be a unary formula of an extension* T *of* Z_1. *Then the following is a theorem of* T:

(14) $\mathrm{C}(0)\ \&\ \forall x[\mathrm{C}(x) \to \mathrm{C}(\mathrm{S}x)] \to [\mathrm{C}^3(x) \to \mathrm{C}(x)]\ \&\ \mathrm{C}^3(0)\ \&\ [\mathrm{C}^3(x) \to \mathrm{C}^3(\mathrm{S}x)]\ \&\ [\mathrm{C}^3(x_1)\ \&\ \mathrm{C}^3(x_2) \to \mathrm{C}(x_1 + x_2)]\ \&\ [\mathrm{C}^3(x_1)\ \&\ \mathrm{C}^3(x_2) \to \mathrm{C}(x_1 \cdot x_2)]$.

Proof. Suppose the hypothesis $\mathrm{C}(0)\ \&\ \forall x[\mathrm{C}(x) \to \mathrm{C}(\mathrm{S}x)]$. I make the following universally quantified claims:

$*1.$ $\mathrm{C}^2(x) \to \mathrm{C}(x)$
$*2.$ $\mathrm{C}^2(0)$
$*3.$ $\mathrm{C}^2(x) \to \mathrm{C}^2(\mathrm{S}x)$
$*4.$ $\mathrm{C}^2(x_1)\ \&\ \mathrm{C}^2(x_2) \to \mathrm{C}^2(x_1 + x_2)$
$*5.$ $\mathrm{C}^3(x) \to \mathrm{C}^2(x)$
$*6.$ $\mathrm{C}^3(0)$
$*7.$ $\mathrm{C}^3(x) \to \mathrm{C}^3(\mathrm{S}x)$
$*8.$ $\mathrm{C}^3(x_1)\ \&\ \mathrm{C}^3(x_2) \to \mathrm{C}^3(x_1 + x_2)$
$*9.$ $\mathrm{C}^3(x_1)\ \&\ \mathrm{C}^3(x_2) \to \mathrm{C}^3(x_1 \cdot x_2)$.

For $*1$, suppose $\mathrm{C}^2(x)$ and let y be 0 in (12). Then we have $\mathrm{C}(0) \to \mathrm{C}(0 + x)$. But $\mathrm{C}(0)$ by hypothesis, and $0 + x = x$ by (10) and (3).

We have $*2$ by (12) and (3).

For $*3$, suppose $\mathrm{C}^2(x)\ \&\ \mathrm{C}(y)$. We need to show $\mathrm{C}(y + \mathrm{S}x)$. We have $\mathrm{C}(y + x)$, and so $\mathrm{C}\big(\mathrm{S}(y + x)\big)$ by hypothesis. But $\mathrm{S}(y + x) = y + \mathrm{S}x$ by (4).

For $*4$, suppose $\mathrm{C}^2(x_1)\ \&\ \mathrm{C}^2(x_2)\ \&\ \mathrm{C}(y)$. We need to show that $\mathrm{C}\big(y + (x_1 + x_2)\big)$. But we have $\mathrm{C}(y + x_1)$ by (12) and so $\mathrm{C}\big((y + x_1) + x_2\big)$, again by (12). But $(y + x_1) + x_2 = y + (x_1 + x_2)$ by (7).

For $*5$, suppose $\mathrm{C}^3(x)$. We have $\mathrm{C}^2(\mathrm{S}0)$ by $*2$ and $*3$ and hence $\mathrm{C}^2(\mathrm{S}0 \cdot x)$ by (13). But $\mathrm{S}0 \cdot x = x \cdot \mathrm{S}0 = x \cdot 0 + x = 0 + x = x + 0 = x$ by (11), (6), (5), (10), and (3).

We have $*6$ by (13), (5), and $*2$.

For $*7$, suppose $\mathrm{C}^3(x)\ \&\ \mathrm{C}^2(y)$. We need to show $\mathrm{C}^2(y \cdot \mathrm{S}x)$. We have $\mathrm{C}^2(y \cdot x)$ by (13) and so $\mathrm{C}^2(y \cdot x + y)$ by $*4$. But $y \cdot x + y = y \cdot \mathrm{S}x$ by (6).

For $*8$, suppose $\mathrm{C}^3(x_1)\ \&\ \mathrm{C}^3(x_2)\ \&\ \mathrm{C}^2(y)$. We need to show $\mathrm{C}^2\big(y \cdot (x_1 + x_2)\big)$. We have $\mathrm{C}^2(y \cdot x_1)$ and $\mathrm{C}^2(y \cdot x_2)$ by (13) and so $\mathrm{C}^2(y \cdot x_1 + y \cdot x_2)$ by $*4$. But $y \cdot x_1 + y \cdot x_2 = y \cdot (x_1 + x_2)$ by (8).

For $*9$, suppose $\mathrm{C}^3(x_1)$ & $\mathrm{C}^3(x_2)$ & $\mathrm{C}^2(y)$. We need to show $\mathrm{C}^2\big(y\cdot(x_1\cdot x_2)\big)$. We have $\mathrm{C}^2(y\cdot x_1)$ by (13) and so $\mathrm{C}^2\big((y\cdot x_1)\cdot x_2\big)$, again by (13). But $(y\cdot x_1)\cdot x_2 = y\cdot(x_1\cdot x_2)$ by (9). □

Of course, what we have just given is not a proof in T in the formal sense of proof that we have been discussing. It uses many derived rules treated in [Sh Ch 3], such as the deduction theorem and the equivalence theorem.

Let Z_2 be the extension of Z_1 obtained by adjoining constants

$$\mathrm{e}_0\ \mathrm{e}_1\ \mathrm{e}_2\ \mathrm{e}_3\ \mathrm{e}_4\ \mathrm{e}_5\ \mathrm{e}_6\ \mathrm{e}_7\ \mathrm{e}_8\ \mathrm{e}_9\ \mathrm{e}_{10}\ \mathrm{e}_{11}\ \mathrm{e}_{12}\ \mathrm{e}_{13}\ \mathrm{e}_{14}\ \mathrm{e}_{15}\ \mathrm{e}_{16}$$

and the axiom

(15) $\mathrm{e}_0 = \mathrm{SS0}$ & $\mathrm{e}_1 = \mathrm{e}_0\cdot\mathrm{e}_0$ & $\mathrm{e}_2 = \mathrm{e}_1\cdot\mathrm{e}_1$ & $\mathrm{e}_3 = \mathrm{e}_2\cdot\mathrm{e}_2$ & $\mathrm{e}_4 = \mathrm{e}_3\cdot\mathrm{e}_3$ & $\mathrm{e}_5 = \mathrm{e}_4\cdot\mathrm{e}_4$ & $\mathrm{e}_6 = \mathrm{e}_5\cdot\mathrm{e}_5$ & $\mathrm{e}_7 = \mathrm{e}_6\cdot\mathrm{e}_6$ & $\mathrm{e}_8 = \mathrm{e}_7\cdot\mathrm{e}_7$ & $\mathrm{e}_9 = \mathrm{e}_8\cdot\mathrm{e}_8$ & $\mathrm{e}_{10} = \mathrm{e}_9\cdot\mathrm{e}_9$ & $\mathrm{e}_{11} = \mathrm{e}_{10}\cdot\mathrm{e}_{10}$ & $\mathrm{e}_{12} = \mathrm{e}_{11}\cdot\mathrm{e}_{11}$ & $\mathrm{e}_{13} = \mathrm{e}_{12}\cdot\mathrm{e}_{12}$ & $\mathrm{e}_{14} = \mathrm{e}_{13}\cdot\mathrm{e}_{13}$ & $\mathrm{e}_{15} = \mathrm{e}_{14}\cdot\mathrm{e}_{14}$ & $\mathrm{e}_{16} = \mathrm{e}_{15}\cdot\mathrm{e}_{15}$.

Then e_{16} denotes $2^{2^{16}} = 2^{65536}$ in Z_2. Here is a proof of $\varphi(\mathrm{e}_{16})$. In (14), let C be $\varphi(x)$. By (0), (14), and (15) we have $\varphi^3(0)$, $\varphi^3(\mathrm{S0})$, $\varphi^3(\mathrm{e}_0)$, $\varphi^3(\mathrm{e}_1)$, $\varphi^3(\mathrm{e}_2)$, $\varphi^3(\mathrm{e}_3)$, $\varphi^3(\mathrm{e}_4)$, $\varphi^3(\mathrm{e}_5)$, $\varphi^3(\mathrm{e}_6)$, $\varphi^3(\mathrm{e}_7)$, $\varphi^3(\mathrm{e}_8)$, $\varphi^3(\mathrm{e}_9)$, $\varphi^3(\mathrm{e}_{10})$, $\varphi^3(\mathrm{e}_{11})$, $\varphi^3(\mathrm{e}_{12})$, $\varphi^3(\mathrm{e}_{13})$, $\varphi^3(\mathrm{e}_{14})$, $\varphi^3(\mathrm{e}_{15})$, $\varphi^3(\mathrm{e}_{16})$, and so $\varphi(\mathrm{e}_{16})$ as claimed.

When the unnecessary complications in this proof are unraveled, we obtain a normal form proof that makes no detours—but by Theorem S its expression length is least 2^{65536}.

We can do worse than this. Let Z_3 be the extension of Z_1 obtained by adjoining binary function symbols $\uparrow$ (exponentiation) and $\Uparrow$ (super-exponentiation) and the axioms

(16) $x\uparrow 0 = \mathrm{S0}$
(17) $x\uparrow \mathrm{S}y = (x\uparrow y)\cdot x$
(18) $x\Uparrow 0 = \mathrm{S0}$
(19) $x\Uparrow \mathrm{S}y = x\uparrow(x\Uparrow y)$.

Then $\mathrm{SS0}\Uparrow\bar{\nu}$ denotes

$$2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$$

with ν occurrences of 2. But we can prove $\varphi(\mathrm{SS0}\Uparrow\bar{\nu})$ in ν steps (using this word loosely), as follows.

By ν applications of (19) we have

(20) $\vdash_{\mathrm{Z}_3} \mathrm{SS0}\Uparrow\bar{\nu} = \mathrm{SS0}\uparrow\cdots\uparrow\mathrm{SS0}$

where there are ν occurrences of SS0 on the right. (Infix symbols are associated from right to left.)

Use the abbreviation $\varphi_1(x)$ for $\varphi(x)$ & $\varphi^3(\mathrm{SS0} \uparrow x)$. Then we have $\vdash_{\mathrm{Z}_3} \varphi_1(0)$ & $[\varphi_1(x) \to \varphi_1(\mathrm{S}x)]$, so (14) applies to $\varphi_1(x)$. Now write $\varphi_2(x)$ for $\varphi_1(x)$ & $\varphi_1^3(\mathrm{SS0} \uparrow x), \ldots, \varphi_\nu(x)$ for $\varphi_{\nu-1}(x)$ & $\varphi_{\nu-1}^3(\mathrm{SS0} \uparrow x)$. By (20) and ν applications of (14) we obtain $\varphi^3(\mathrm{SS0} \Uparrow \bar{\nu})$ and consequently $\varphi(\mathrm{SS0} \Uparrow \bar{\nu})$.

Theorem T is a relativization scheme, invented by Solovay in 1976. Relativization schemes are used extensively in my book entitled *Predicative Arithmetic*;[5] see especially Chapters 3 and 18.

[5]Mathematical Notes 32, Princeton University Press, Princeton, NJ, 1986.