THE JOURNAL OF SYMBOLIC LOGIC
Volume 44, Number 1, March 1979

THE RELATIVE EFFICIENCY OF PROPOSITIONAL PROOF SYSTEMS

STEPHEN A. COOK AND ROBERT A. RECKHOW

§1. **Introduction.** We are interested in studying the length of the shortest proof of a propositional tautology in various proof systems as a function of the length of the tautology. The smallest upper bound known for this function is exponential, no matter what the proof system. A question we would like to answer (but have not been able to) is whether this function has a polynomial bound for some proof system. (This question is motivated below.) Our results here are relative results.

In §§2 and 3 we indicate that all standard Hilbert type systems (or Frege systems, as we call them) and natural deduction systems are equivalent, up to application of a polynomial, as far as minimum proof length goes. In §4 we introduce *extended Frege* systems, which allow introduction of abbreviations for formulas. Since these abbreviations can be iterated, they eliminate the need for a possible exponential growth in formula length in a proof, as is illustrated by an example (the pigeon-hole principle). In fact, Theorem 4.6 (which is a variation of a theorem of Statman) states that with a penalty of at most a linear increase in the number of lines of a proof in an extended Frege system, no line in the proof need be more than a constant times the length of the formula proved. The most difficult result is Theorem 4.5, which states that all extended Frege systems, regardless of which set of connectives they use, are about equivalent, as far as minimum proof length goes. Finally, in §5 we discuss the substitution rule, and show that Frege systems with this rule can simulate extended Frege systems.

Some of our results here appeared earlier in the conference proceedings [1], and Reckhow's Ph. D. thesis [2]. (These two papers also establish and report non-polynomial lower bounds on some proof systems more restricted than the ones mentioned above.)

To motivate the study of propositional proof systems, let us briefly review some of the theory of $\mathscr{P}$ and $\mathscr{NP}$ (see [3], [4], and Chapter 10 of [5]). By convention, $\mathscr{P}$ denotes the class of sets of strings recognizable by a deterministic Turing machine in time bounded by a polynomial in the length of the input. $\mathscr{NP}$ is the same for nondeterministic Turing machines. If we let TAUT denote the set of tautologies over any fixed adequate set of connectives, then the main theorem in [3] implies that $\mathscr{P} = \mathscr{NP}$ if and only if TAUT is in $\mathscr{P}$. Now $\mathscr{P} = \mathscr{NP}$ not only would imply the existence of relatively fast algorithms for many interesting and apparently unfeasible combinatorial algorithms in $\mathscr{NP}$ (see [4]), it would also have an interesting philosophical consequence for mathematicians. If $\mathscr{P} = \mathscr{NP}$, then there is a polynomial p and an algorithm $\mathscr{A}$ with the following property. Given any

Received May 24, 1977.

proposition S of set theory and any integer n, $\mathscr{A}$ determines within only $p(n)$ steps whether S has a proof of length n or less in (say) Zermelo-Fraenkel set theory. To see that the existence of $\mathscr{A}$ follows from $\mathscr{P} = \mathscr{NP}$, observe that the problem solved by $\mathscr{A}$ is in $\mathscr{NP}$. In fact, a nondeterministic Turing machine can write any string of length n on its tape and then verify that the string is a proof of the given proposition. For any reasonable logical theory, this verification can be performed within time bounded by some polynomial in n.

Hence the importance of showing $\mathscr{P} \neq \mathscr{NP}$ (or $\mathscr{P} = \mathscr{NP}$?). A related important question is whether $\mathscr{NP}$ is closed under complementation, i.e. $\Sigma^* - L$ is in $\mathscr{NP}$ whenever L is in $\mathscr{NP}$. (Here we use the notation Σ^* for the set of all finite strings over the finite alphabet Σ under consideration, and the assumption $L \subseteq \Sigma^*$. This notation will be used throughout.) If $\mathscr{NP}$ is not closed under complementation, then of course $\mathscr{P} \neq \mathscr{NP}$. On the other hand, if $\mathscr{NP}$ is closed under complementation, this would have interesting consequences for each of the combinatorial problems in [4]. Hence the following result is important.

1.1. PROPOSITION. *$\mathscr{NP}$ is closed under complementation if and only if* TAUT *is in $\mathscr{NP}$.*

1.2. *Notation.* $\mathscr{L}$ is the set of functions $f: \Sigma_1^* \to \Sigma_2^*$, Σ_1, Σ_2 any finite alphabets, such that f can be computed by a deterministic Turing machine in time bounded by a polynomial in the length of the input.

PROOF OF 1.1. The complement of the set of tautologies is in $\mathscr{NP}$, since to verify that a formula is not a tautology one can guess at a truth assignment and verify that it falsifies the formula. Conversely, suppose the set of tautologies is in $\mathscr{NP}$. By the proof of the main theorem in [3], every set L in $\mathscr{NP}$ is reducible to the complement of the tautologies in the sense that there is a function f in $\mathscr{L}$ such that for all strings x, $x \in L$ iff $f(x)$ is not a tautology. Hence a nondeterministic procedure for accepting the complement of L is: on input x, compute $f(x)$, and accept x if $f(x)$ is a tautology, using the nondeterministic procedure for tautologies assumed above. Hence the complement of L is in $\mathscr{NP}$. $\square$

The question of whether TAUT is in $\mathscr{NP}$ is equivalent to whether there is a propositional proof system in which every tautology has a short proof, provided "proof system" and "short" are properly defined.

1.3. DEFINITIONS. If $L \subseteq \Sigma^*$, a *proof system* for L is a function $f: \Sigma_1^* \to L$ for some alphabet Σ_1 and f in $\mathscr{L}$ such that f is onto. We say that the proof system is *polynomially bounded* iff there is a polynomial $p(n)$ such that for all $y \in L$ there is $x \in \Sigma_1^*$ such that $y = f(x)$ and $|x| \leq p(|y|)$, where $|z|$ denotes the length of a string z.

If $y = f(x)$, then we will say that x is a *proof* of y, and x is a *short* proof of y if in addition $|x| \leq p(|y|)$. Thus a proof system f is polynomially bounded iff there is a bounding polynomial $p(n)$ with respect to which every $y \in L$ has a short proof.

1.4. PROPOSITION. *A set L is in $\mathscr{NP}$ iff $L = \varnothing$ or L has a polynomially bounded proof system.*

The analogous statement for recursive function theory is that L is recursively enumerable iff $L = \varnothing$ or L is the range of a recursive function. The proof of the present proposition is straightforward. If $L \in \mathscr{NP}$, then some nondeterministic Turing machine M accepts L in polynomial time. If $L \neq \varnothing$, we define f such that if x codes a computation of M which accepts y, then $f(x) = y$. If x does not code an accepting computation, then $f(x) = y_0$ for some fixed $y_0 \in L$. Then f is clearly a polynomially bounded proof system for L. Conversely, if f is a polynomially bounded proof system for L, then a fast nondeterministic algorithm for accepting L is, on input y, guess a short proof x of y and verify $f(x) = y$. □

Putting Propositions 1.1 and 1.4 together we see that $\mathscr{NP}$ is closed under complementation if and only if TAUT has a polynomially bounded proof system, in the general sense of Definition 1.3. It is easy to see (and is argued below) that any conventional proof system for tautologies can naturally be made to fit the definition of proof system in 1.3. Although it is doubtful that every general proof system for TAUT is natural, nevertheless this general framework helps explain the motivating question of this paper: Are any conventional propositional proof systems polynomially bounded?

We cannot answer that question directly (except negatively for certain restricted systems: see [1] and [2], and also [8]), but at least we can put different proof systems into equivalence classes such that the answer is the same for equivalent systems. We conjecture that the answer is always no.

1.5. Definition. If $f_1: \Sigma_1^* \to L$ and $f_2: \Sigma_2^* \to L$ are proof systems for L, then f_2 *p-simulates* f_1 provided there is a function $g: \Sigma_1^* \to \Sigma_2^*$ such that g is in $\mathscr{L}$, and $f_2(g(x)) = f_1(x)$ for all x.

Thus g translates a proof x of y in the system f_1 into a proof $g(x)$ of y in f_2. It is easy to see, using the fact that $\mathscr{L}$ is closed under composition, that p-simulation is a transitive reflexive relation, so that its symmetric closure is an equivalence relation.

1.6. Proposition. *If a proof system f_2 for L p-simulates a polynomially bounded proof system f_1 for L*, then f_2 *is also polynomially bounded.*

This is an immediate consequence of the definitions of "proof system" and "polynomially bounded", and the fact that every function in $\mathscr{L}$ is bounded in length by a polynomial in the length of its argument. □

We close this section by establishing some notation and terminology specific for propositional proof systems which will be used in the rest of this paper. The letter κ will always stand for an adequate set of propositional connectives which are binary, unary, or nullary (have two, one, or zero arguments). *Adequate* here means that every truth function can be expressed by formulas built up from members of κ. A *formula* refers to a propositional formula built up in the usual way from atoms (propositional variables) and connectives from some set κ, using infix notation. (We speak of a formula *over* κ if its connectives are from κ.) If $A_1, \ldots, A_n, B$ are formulas, then we write $A_1, \ldots, A_n \models B$ if B is a logical consequence of $A_1, \ldots, A_n$ (i.e. every truth assignment satisfying $A_1, \ldots, A_n$ satisfies B). Each of our propositional proof systems will be defined relative to some connective set κ, and will be capable of proving all tautologies over κ by proofs using formulas over κ. A *derivation* (from zero or more lines called *hypotheses*) in such a system is a finite sequence of *lines*, ending in the line proved. A line is always a formula, except in the case of natural deduction systems (§3). Each line must either be a hypothesis, or *follow* from earlier lines by a rule of inference. (In case the rule itself has no hypothesis, the rule is an *axiom scheme*.) If the derivation has no hypothesis, it is called a *proof.*

Thus to specify a propositional proof system for our purposes, it is only necessary to specify κ, the definition of a line, and a finite set of rules of inference. To make this notion of proof system be an instance of our abstract Definition 1.3, we note first of all that formulas can be naturally regarded as strings over a finite alphabet. The only problem is that an atom itself must be regarded as a string (say the letter P followed by a string over $\{0, 1\}$) in order that there be an unlimited supply of atoms. Then a proof π in the propositional system which is, say, a sequence of formulas, can naturally be regarded as a string over a finite alphabet which includes the comma as a separator symbol, as well as the symbols necessary to specify the formulas. The function f which abstractly specifies the system would be given by $f(\pi) = A$ if π proves A, and $f(\pi) = A_0$ for some fixed tautology A_0 if π is a string not corresponding to a proof in the system.

The notation $A_1, \ldots, A_n \vdash_{\mathscr{S}}^{\pi} B$ means that π is a derivation of B from hypotheses $A_1, \ldots, A_n$ in the proof system $\mathscr{S}$. (The notation $\vdash_{\mathscr{S}}$ means that there is some derivation π in the system $\mathscr{S}$.) We use the following notation for various length measures:

$l(A)$ is the number of occurrences of atoms and nullary connectives in a formula (or sequence) A.

$\lambda(\pi)$ is the number of lines in a derivation π.

$\rho(\pi) = \max_i l(A_i)$, if π is $(A_1, \ldots, A_n)$.

$|\pi|$ or $|A|$ is the length of π or A as a string.

§2. **Frege systems.** In the most usual propositional proof systems the rules of inference are formula schemes, and an instance of the scheme is obtained by applying a substitution to the scheme. We shall call such systems *Frege systems*, after Frege [6].

Throughout this section we assume that all formulas are over some fixed adequate connective set κ. The following terms are defined relative to κ.

2.1. DEFINITIONS. If $D_1, \ldots, D_k$ are formulas and $P_1, \ldots, P_k$ are distinct atoms, then $\sigma = (D_1, \ldots, D_k)/(P_1, \ldots, P_k)$ is a *substitution*, and σA is the formula which results by simultaneously replacing P_i by D_i, $i = 1, \ldots, k$, in formula A. A *Frege rule* is a system of formulas $(C_1, \ldots, C_n)/D$, where $C_1, \ldots, C_n \models D$. If $n = 0$, the rule is an *axiom scheme*. For any substitution σ we say that *σD follows from* $\sigma C_1, \ldots, \sigma C_n$ by the rule $(C_1, \ldots, C_n)/D$. An *inference system* $\mathscr{F}$ is a finite set of Frege rules. The notions of *derivation* and the symbol $\vdash$ for $\mathscr{F}$ are defined as in the end of §1, where now a *line* in a derivation is a formula. By our condition on the definition of Frege rule, it is clear that if $A_1, \ldots, A_n \vdash_{\mathscr{F}} B$ then $A_1, \ldots, A_n \models B$.

2.2. DEFINITIONS. An inference system $\mathscr{F}$ is *implicationally complete* if $A_1, \ldots, A_n \vdash_{\mathscr{F}} B$ whenever $A_1, \ldots A_n \models B$. A *Frege system* is an implicationally complete inference system.

In fact, Frege's original system in [6] does not fit the above definition, because it has axioms instead of axiom schemes, and tacitly includes the substitution rule (see §5). According to Church [12, p. 158], the idea of axiom schemes used to replace the substitution rule is due to von Neumann [13]. If we modify Frege's system to be a Frege system, the result has connectives $\kappa = \{\neg, \supset\}$, and the rule

$$\frac{A, A \supset B}{B}$$

and the six axiom schemes

$$A \supset (B \supset A), \qquad (C \supset (B \supset A)) \supset ((C \supset B) \supset (C \supset A)),$$

$$(D \supset (B \supset A)) \supset (B \supset (D \supset A)), \qquad (B \supset A) \supset (\neg A \supset \neg B),$$

$$\neg\neg A \supset A, \qquad A \supset \neg\neg A.$$

2.3. THEOREM. *For any two Frege systems $\mathscr{F}_1$ and $\mathscr{F}_2$ over κ there is a function f in $\mathscr{L}$ and constant c such that for all formulas $A_1, \ldots, A_n$, B and derivations π, if $A_1, \ldots, A_n \vdash^{\pi}_{\mathscr{F}_1} B$ then $A_1, \ldots, A_n \vdash^{f(\pi)}_{\mathscr{F}_2} B$, and $\lambda(f(\pi)) \leq c\lambda(\pi)$ and $\rho(f(\pi)) \leq c\rho(\pi)$. (See the end of* §1 *for notation.)*

2.4. COROLLARY. *Any two Frege systems over κ p-simulate each other. Hence one Frege system over κ is polynomially bounded iff all Frege systems over κ are.*

The corollary is an immediate consequence of the theorem and Proposition 1.6. Reckhow [2] proves a generalization of the corollary to cover the case of Frege systems with different connective sets simulating each other, even when some of the connectives have arity greater than two. His proof is much more complicated than our proof of Theorem 2.3 given below, largely because of the difficulty of simulating systems using the connectives $\equiv$ and $\not\equiv$ by systems without these connectives. Fortunately, Corollary 4.6. below, concerning extended Frege systems, makes Corollary 2.4 and Reckhow's generalization less important than they might appear at first, since extended Frege systems seem to be more natural than Frege systems when measuring proof lengths.

The lemma below is used in the proof of Theorem 2.3. (The notation $\sigma(\pi)$ means $\sigma A_1, \ldots, \sigma A_k$, if π is a derivation $A_1, \ldots, A_k$.)

2.5. LEMMA. *If π is a derivation of A from $B_1, \ldots, B_k$ in a Frege system $\mathscr{F}$, then $\sigma(\pi)$ is a derivation of σA from $\sigma B_1, \ldots, \sigma B_k$ in $\mathscr{F}$, for any substitution σ.*

The proof is an easy induction on the length of π. □

To prove Theorem 2.3, assume $\mathscr{F}_1$ and $\mathscr{F}_2$ are Frege systems over κ. For each rule $R = (C_1, \ldots, C_m)/D$ in $\mathscr{F}_1$, let π_R be a derivation of D from $C_1, \ldots, C_m$ in $\mathscr{F}_2$. Now suppose π is a derivation of B from $A_1, \ldots, A_n$ in $\mathscr{F}_1$, and suppose $\pi = (B_1, \ldots, B_k)$. To construct the $\mathscr{F}_2$-derivation $f(\pi)$ from π, if B_i follows from earlier B_j's by the $\mathscr{F}_1$-rule R_i and substitution σ_i, simply replace B_i by the derivation $\sigma_i(\pi_{R_i})$ (with hypotheses deleted). According to Lemma 2.5, $\sigma_i(\pi_{R_i})$ is a derivation of B_i from the same earlier B_j's. Clearly $\lambda(f(\pi)) \leq c_1\lambda(\pi)$, where c_1 is the number of lines in the longest derivation π_R, as R ranges over the finite set of rules of $\mathscr{F}_1$. Finally, $\rho(f(\pi)) \leq c_2\rho(\pi)$, where c_2 is an upper bound on $l(A)$ as A ranges over all formulas in all the derivations π_R, R a rule of $\mathscr{F}_1$. □

§3. **Natural deduction systems.** The purpose of this section is to indicate the sense in which natural deduction systems are equivalent to Frege systems. Rather than presenting a specific natural deduction system, such as one appearing in Prawitz [7], we shall introduce a general definition analogous to our general notion of Frege system. To make the classical proposition system of Prawitz fit our definition, it is necessary to allow Prawitz's notion of proof to be a more general directed acyclic graph, rather than a tree. That is, once a formula is derived from a set of assumptions, we do not require that it be derived again if it is used twice. Alternatively, we could stick to Prawitz's tree proofs, provided that if a formula occurred several times in a proof with the same assumptions, it be counted only once in measuring the length of the proof. In fact, we shall present our natural deduction proofs as sequences of lines, and each line will have the form $A_1, \ldots, A_n \to A$, where $A_1, \ldots A_n$ are assumptions which imply A. Thus our proofs require repeating the assumptions for a formula with each step, which makes them a little longer and harder to write down, but easier to analyze. For convenience, we allow only the right-most formula A_n to be discharged. Reckhow [2] gives a more general treatment of natural deduction systems, as well as Gentzen's sequent systems.

Part of the appeal of a natural deduction system is that it allows the "deduction theorem" to be used as a rule. According to the deduction theorem, from a derivation π in a Frege system $\mathscr{F}$ showing $A_1, \ldots, A_m \vdash B$ we can construct a derivation π' in $\mathscr{F}$ showing $A_1, \ldots, A_{m-1} \vdash A_m \supset B$. The trouble is that π' may be twice as long as π, so that if a natural deduction derivation has m nested uses of this deduction rule and they are eliminated sequentially to obtain a Frege derivation, the result might be longer by a factor of 2^m than the original derivation. Fortunately, they can be eliminated simultaneously, as shown by the construction fr($\mathscr{N}$) below.

The following definitions are relative to a given adequate connective set κ.

3.1. *Notation.* Even if $\neg$ or $\vee$ is not in κ, formulas $N(P)$ and $O(P, Q)$ over κ can always be found such that $N(P)$ and $O(P, Q)$ are equivalent to $\neg P$ and $P \vee Q$, respectively, and such that P and Q each has at most one occurrence in each of $N(P)$ and $O(P, Q)$. A fixed "dummy" atom P_0 may occur several times, however. For example, if κ is $\{\not\equiv, \supset\}$ then $N(P)$ could be $(P \not\equiv (P_0 \supset P_0))$ and $O(P, Q)$ could be $((P \not\equiv (P_0 \supset P_0)) \supset Q)$. (See §5. 3.1.1 of [2] for an argument showing how this can be done in general.) Thus we will take $\neg A$ or $A \vee B$ to mean $N(A)$ or $O(A, B)$, respectively, if $\neg$ or $\vee$ is not in κ. We use $\bigvee(A_1, \ldots, A_m)$ to stand for $(\cdots(A_1 \vee A_2) \cdots \vee A_m)$ (association to the left), and $\bigvee'(A_1, \ldots, A_m)$ to stand for $(A_1 \vee \cdots (A_{m-1} \vee A_m) \cdots)$ (association to the right).

3.2. DEFINITIONS. A *naturnal deduction line* (or just *line*) is a pair $\Gamma \to A$, where Γ is any finite sequence of formulas, and A is a formula. If Γ is empty, the line is written simply $\to A$. Associated with a line $L = (A_1, \ldots, A_m) \to A$ are two equivalent formulas $L^* = \bigvee(\neg A_1, \ldots, \neg A_m, A)$ and $L^\# = \bigvee'(\neg A_1, \ldots, \neg A_m, A)$. (If $m = 0$, the $L^* = L^\# = A$.) The line L takes on the same truth value under a truth assignment as formulas L^* and $L^\#$, so that the concepts of validity, logical consequence, etc. are well defined for lines. If Δ is a sequence $B_1, \ldots, B_n$ of formulas and L is the line $(A_1, \ldots, A_m) \to A$, then ΔL is the line $(B_1, \ldots, B_n, A_1, \ldots, A_m) \to A$. If Λ is a set of lines, L is a line, Δ is a sequence of formulas, and σ is a substitution, then $\Lambda \models L$ implies that $\Delta\sigma(\Lambda) \models \Delta\sigma(L)$, where the operations Δ and σ are extended to sets of lines in the natural way. If Λ is a finite set of lines and L is a line such that $\Lambda \models L$, then the system $R = \Lambda/L$ is a *natural deduction rule.* Line L' *follows* from Λ' by rule R provided for some substitution σ and sequence Δ, $\Lambda' = \Delta\sigma(\Lambda)$, and $L' = \Delta\sigma(L)$. A *natural deduction system* is a finite set of natural deduction rules which is implicationally complete (implicationally complete being defined in a manner analogous to that for Frege systems). A formula A is represented in a natural deduction system $\mathscr{N}$ by the line $\to A$. This convention allows us to speak of proofs of formulas and derivations of a formula from formulas in $\mathscr{N}$, and thus write for example $A_1, \ldots, A_n \vdash^{\pi}_{\mathscr{N}} B$ instead of $\to A_1, \ldots, \to A_n \vdash^{\pi}_{\mathscr{N}} \to B$.

If $L = (A_1, \ldots, A_k) \to A$ is a line, then $l(L) = l(A_1) + \cdots + l(A_k) + l(A)$. If π is a derivation, then $\lambda(\pi)$ is the number of lines in π, and $\rho(\pi)$ is the maximum of $l(L)$, for all L in π.

An example of a natural deduction rule, which embodies the deduction theorem, is $R_1 = (P \to Q)/(\to \neg P \vee Q)$. This rule together with its converse $R_2 = (\to \neg P \vee Q)/(P \to Q)$ can turn any Frege system $\mathscr{F}$ into a natural deduction sytem nd($\mathscr{F}$), provided we reinterpret every rule $R = (C_1, \ldots, C_n)/D$ of the Frege system to be $R' = (\to C_1, \ldots, \to C_n)/ \to D$. In fact, if $\Lambda \models L$, then to deduce L from Λ in nd($\mathscr{F}$), we first observe that every hypothesis M in Λ can be changed to $\to M^\sharp$ by repeated use of the rule R_1. By the implicational completeness of $\mathscr{F}$, we can derive $\to L^\sharp$ in nd($\mathscr{F}$) from these lines $\to M^\sharp$. Now L can be derived from $\to L^\sharp$ by repeated use of the rule R_2.

Notice that every derivation in $\mathscr{F}$, of say B from $A_1, \ldots, A_n$, can be turned into a derivation of B from $A_1, \ldots, A_n$ in nd($\mathscr{F}$) simply by adding the symbol $\to$ to the left of every formula in the derivation.

Conversely, every natural deduction system $\mathscr{N}$ can be turned into a Frege system fr($\mathscr{N}$), where the rules of fr($\mathscr{N}$) consist of the two rules R' and R'' for every rule R of $\mathscr{N}$. To explain R' and R'' we need to recall the notation M^* for $\bigvee(\neg A_1, \ldots, \neg A_m, A)$ and introduce the notation $(PM)^*$ for $\bigvee(P, \neg A_1, \ldots, \neg A_m, A)$, where M is a line $(A_1, \ldots, A_m) \to A$ and P is an atom. If $R = \Lambda/L$, then $R' = \Lambda^*/L^*$ and $R'' = (P\Lambda)^*/(PL)^*$, where P is some atom not occurring in Λ or L, and we have extended the * notation to sets Λ of lines in the obvious manner. It is easy to see that the rules R' and R'' are sound if R is sound.

Now if $\pi = L_1, \ldots, L_n$ is any derivation in $\mathscr{N}$, then we claim that $\pi^* = L_1^*, \ldots, L_n^*$ is a derivation in fr($\mathscr{N}$). For suppose L_i follows from earlier L_j's by the the rule $R = \Lambda/L$ in $\mathscr{N}$. Then for some substitution σ and sequence Δ, L_i is $\Delta\sigma(L)$ and the earlier L_j's comprise the set $\Delta\sigma(\Lambda)$. If Δ is empty, the L_i^* follows from earlier L_j^*'s by the Frege rule $R' = \Lambda^*/L^*$ by σ, since for any line M, $(\sigma(M))^* = \sigma(M^*)$. If Δ is not empty, then L_i^* follows from earlier L_j^*'s by the Frege rule $R'' = (P\Lambda)^*/(PL)^*$ and substitution σ', where σ' is the substitution obtained by simultaneously applying the substitution σ and $\bigvee(\neg A_1, \ldots, \neg A_k)/P$, where Δ is $(A_1, \ldots, A_k)$. We need the fact that for any line M with no occurrence of P, $\sigma'((PM)^*) = (\Delta\sigma(M))^*$.

Thus π^* is a derivation in fr($\mathscr{N}$) for every derivation π in $\mathscr{N}$. Notice that since $(\to A)^* = A$, if π is a derivation in $\mathscr{N}$ of B from $A_1, \ldots, A_l$, then π^* is a derivation in fr($\mathscr{N}$) of B from $A_1, \ldots, A_l$. Further, notice that $\lambda(\pi^*) = \lambda(\pi)$ and $\rho(\pi^*) \leq c\rho(\pi)$, where the constant c depends only on the underlying connective set κ.

Although the constructions above allow us to translate back and forth between Frege and natural deduction systems, the following result still needs a separate proof.

3.3. THEOREM. *Given natural deduction systems $\mathscr{N}_1$ and $\mathscr{N}_2$ over κ there is a function f in $\mathscr{L}$ and a constant c such that for all lines $L_1, \ldots, L_n$, L and derivations π, if $L_1, \ldots, L_n \vdash^{\pi}_{\mathscr{N}_1} L$, then $L_1, \ldots, L_n \vdash^{f(\pi)}_{\mathscr{N}_2} L$, and $\lambda(f(\pi)) \leq c\lambda(\pi)$ and $\rho(f(\pi)) \leq c\rho(\pi)$.*

The proof is very similar to the proof of Theorem 2.3. Lemma 2.5 is replaced by the statement that if π is a derivation in $\mathscr{N}$ of line M from lines $M_1, \ldots, M_k$, then $\Delta\sigma(\pi)$ is a derivation of $\Delta\sigma(M)$ from $\Delta\sigma(M_1), \ldots, \Delta\sigma(M_k)$. □

3.4. COROLLARY. *Let κ be any adequate set of connectives. All Frege and natural deduction systems over κ p-simulate all other Frege and natural deduction systems over κ. Hence one such system over κ is polynomially bounded if and if all such systems over κ are polynomially bounded.*

The corollary follows immediately from Theorems 2.3 and 3.3, together with the constructions nd($\mathscr{F}$) and fr($\mathscr{N}$) given above. □

Reckhow [2] treats a kind of natural deduction system in which Γ in a line $\Gamma \to A$ is regarded as a set of formulas rather than a sequence of formulas. Such a system might allow for shorter proofs, since in effect there are implicit rules which allow Γ to be reordered. In [2] it is shown that the above corollary holds for this system, and that the second part holds even when the systems have different connective sets.

The corollary also holds for Gentzen systems with cut, provided a Gentzen proof is considered to be a sequence of sequents, so that a given occurrence of a sequent can be used more than once in a proof, as opposed to the more usual definition that a Gentzen proof is a tree of sequents. When a Gentzen proof is defined to be a tree, an exponential lower bound for the number of sequents in a minimum cut-free proof of a formula follows from an unpublished result of Statman. More recently, Cook and Rackoff have an unpublished result showing an exponential lower bound for Gentzen proofs considered as sequences, provided both the cut and thinning rules are disallowed.

§4. **Extended Frege systems.** The previous sections have indicated that certain standard proof systems for the propositional calculus are about equally powerful. We now look for natural extensions of these systems which might be more powerful, in the sense that they yield shorter proofs. To motivate this search, we try to use Frege systems to simulate an informal proof of the "pigeon-hole principle".

One statement of the pigeon-hole principle is that no injective function maps $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n-1\}$, $n \geq 2$. For each value of n, this statement may be formalized in the propositional calculus as follows. Let P_{ij}, $1 \leq i \leq n$, $1 \leq j \leq n-1$, be a set of atoms, whose intended meaning is "i is mapped to j". Let $\mathscr{S}_n$ be the set (or sometimes the conjunction of the formulas in the set) $\{P_{i1} \vee \cdots \vee P_{i,n-1} \mid 1 \leq i \leq n\} \cup \{\neg P_{ik} \vee \neg P_{jk} \mid 1 \leq i < j \leq n, 1 \leq k \leq n-1\}$. If a truth assignment were given for which each formula in $\mathscr{S}_n$ is true then one could define a function f which by the first set of disjunctions is from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n-1\}$ and which by the second set is injective. Thus the formula $A_n = \neg\mathscr{S}_n$ is a tautology.

An informal proof of the pigeon-hole principle proceeds by induction on n. It is obvious for $n = 2$. In general, if f: $\{1, \ldots, n\} \to \{1, \ldots, n-1\}$, then let f': $\{1, \ldots, n-1\} \to \{1, \ldots, n-2\}$ be defined by $f'(i) = f(i)$ if $f(i) \neq n-1$; otherwise $f'(i) = f(n)$. If f is injective, it is easy to see that f' is also, contradicting the induction hypothesis.

To mimic this proof in a Frege system, we try to deduce $\mathscr{S}_{n-1}$ from $\mathscr{S}_n$. For each i, j, we introduce a formula B_{ij} which means $f'(i) = j$. $B_{ij} = P_{ij} \vee (P_{i,n-1} \;\&\; P_{nj})$, $1 \leq i \leq n-1$, $1 \leq j \leq n-2$. Let σ_{n-1} be the substitution B_{ij}/P_{ij} $(1 \leq i \leq n-1,\ 1 \leq j \leq n-2)$. The argument that f injective implies f' injective shows $\mathscr{S}_n \models \sigma_{n-1}(\mathscr{S}_{n-1})$. By completeness, $\mathscr{S}_n \vdash \sigma_{n-1}(\mathscr{S}_{n-1})$. Similarly, $\mathscr{S}_{n-1} \vdash \sigma_{n-2}(\mathscr{S}_{n-2})$, so by Lemma 2.5, there is a derivation of the same number of lines showing $\sigma_{n-1}(\mathscr{S}_{n-1}) \vdash \sigma_{n-1}\,\sigma_{n-2}(\mathscr{S}_{n-2})$, so $\mathscr{S}_n \vdash \sigma_{n-1}\,\sigma_{n-2}(\mathscr{S}_{n-2})$. Proceeding this way, we finally obtain a derivation showing $\mathscr{S}_n \vdash \sigma_{n-1} \cdots \sigma_2(\mathscr{S}_2)$. But $\mathscr{S}_2$ is $\{P_{11}, P_{21}, \neg P_{11} \vee \neg P_{21}\}$, from which a contradiction is easily derived, so by the deduction theorem, $\vdash \neg\mathscr{S}_n$; i. e. $\vdash A_n$

It is not hard to see that by choosing the rules of our Frege system conveniently, the derivation of $\sigma_{n-1}(\mathscr{S}_{n-1})$ from $\mathscr{S}_n$ has $O(n^3)$ lines. Hence the entire proof of A_n has $O(n^4) = O(N^{4/3})$ lines, where N is $|A_n|$. On the other hand, each application of a substitution σ_i triples the length of a formula, so the longest formulas in the proof of A_n grow exponentially in n.

A simple device to reduce the formula length in the above proof is to introduce new atoms which abbreviate the formulas B_{ij}. Thus the atom Q^1_{ij} has a defining formula $Q^1_{ij} \equiv (P_{ij} \vee (P_{i,n-1} \,\&\, P_{nj}))$, $1 \le i \le n-1$, $1 \le j \le n-2$. From these defining formulas and the formulas $\mathscr{S}_n$, the formulas $\tau_{n-1}(\mathscr{S}_{n-1})$ are easily derived, where τ_{n-1} is the substitution Q^1_{ij}/P_{ij} $(1 \le i \le n-1, 1 \le j \le n-2)$. In general, a new atom Q^{k+1}_{ij} is introduced for $\sigma_{n-1} \cdots \sigma_{n-k}(B_{ij})$ with defining formula $Q^{k+1}_{ij} \equiv (Q^k_{ij} \vee (Q^k_{i,n-k-1} \,\&\, Q^k_{n-k,j}))$, and the formulas $\tau_{n-k-1}(\mathscr{S}_{n-k-1})$ are easily derived from these defining formulas and the formulas $\tau_{n-k}(\mathscr{S}_{n-k})$ where τ_{n-k} is the substitution Q^k_{ij}/P_{ij} $(1 \le i \le n-k, 1 \le j \le n-k-1)$. In this way, a contradiction is derived from $\mathscr{S}_n$ in $O(n^4)$ lines, where now each formula has length only $O(n)$. Hence A_n has a proof of length $O(n^5)$ in this framework. This kind of proof system can be formalized as follows:

4.1. Definition. An *extended Frege system* over a connective set κ is a proof system which consists of a Frege system $\mathscr{F}$ over κ together with the *extension rule* which allows formulas of the form $P \equiv A$ to be added to a derivation, where A is any formula over κ, and P is any "new" atom. (P must not occur in A, in any lines preceding $P \equiv A$, or in any hypotheses to the derivation. P can occur in later lines, but not in the last line.) We say P is a *defined* atom and $P \equiv A$ is its *defining formula.* If $\equiv$ is not in κ, we choose some short formula $P \sim Q$ over κ which is equivalent to $P \equiv Q$, and let $P \sim A$ be the defining formula for P. The extended Frege system based on $\mathscr{F}$ is denoted by $e\mathscr{F}$.

(The extension rule was first suggested by Tseitin [8], in the context of resolution proofs.)

4.2. Proposition (Soundness of $e\mathscr{F}$). *If $A_1, \ldots, A_n \vdash_{e\mathscr{F}} B$, then $A_1, \ldots, A_n \models B$.*

Proof. Let τ be any truth assignment to the atoms of $A_1, \ldots, A_n$ and B which satisfies $A_1, \ldots, A_n$. Then τ can be extended to make each line in the derivation true. In particular, if $P \equiv A$ is a defining formula, then P has not occurred earlier in the derivation, so we are free to extend τ so $\tau(P) = \tau(A)$. Hence $\tau(B)$ is true, since B is the last line of the derivation. □

Although the extension rule apparently allows the lengths of formulas in a derivation to be greatly reduced, the following result shows the number of lines in a proof cannot be much reduced.

4.3. *Proposition. If π is a derivation of B from $A_1, \ldots, A_n$ in $e\mathscr{F}$, then there is a derivation π' of B from $A_1, \ldots, A_n$ in $\mathscr{F}$ with $\lambda(\pi') \le \lambda(\pi) + cm$ where c depends only on $\mathscr{F}$, and m is the number of defining formulas in π.*

Proof. Suppose $P_i \sim C_i$, $1 \le i \le m$, are the defining formulas in π (given in the order in which they occur in π). Then π is a derivation in $\mathscr{F}$ of B from $A_1, \ldots, A_n, P_1 \sim C_1, \ldots, P_m \sim C_m$. Now let σ be the composed substitution

$$\frac{C_m}{P_m} \circ \frac{C_{m-1}}{P_{m-1}} \circ \cdots \circ \frac{C_1}{P_1}.$$

By Lemma 2.5, $\sigma(\pi)$ is a derivation of σB from $\sigma A_1, \ldots, \sigma A_n, \sigma(P_1 \sim C_1), \ldots,$

$\sigma(P_m \sim C_m)$. By the restrictions on the defined atoms P_i, $\sigma(\pi)$ is a derivation in $\mathscr{F}$ of B from $A_1, \ldots, A_n, \sigma C_1 \sim \sigma C_1, \ldots, \sigma C_m \sim \sigma C_m$. But $Q \sim Q$ has some fixed proof in $\mathscr{F}$ of some number of lines (say c lines), so by Lemma 2.5, each $\sigma C_i \sim \sigma C_i$ has a proof in $\mathscr{F}$ of c lines. Also $\lambda(\sigma(\pi)) = \lambda(\pi)$. Hence we construct π' from $\sigma(\pi)$ together with these m proofs, and the proposition follows. □

Of course the formulas of π' can grow exponentially in m, even if the formulas of π are short, as shown by the pigeon-hole example at the beginning of this section.

We mentioned that Reckhow [2] strengthened Theorem 2.3 to cover the case of different connective sets, but the proof was complicated by the difficulties of finding a short translation for a formula containing $\equiv$ into one containing, say, just &, $\vee$, and $\neg$. In the case of extended Frege systems, this difficulty can be circumvented. Theorem 4.5 below states that if the number of lines in the shortest proof of a tautology A is bounded by some function $L(l(A))$ in some extended Frege system, then essentially the same is true of any extended Frege system over any connective set, and furthermore the lengths of the formulas in a proof need not be much longer than the formula proved. (The latter is in sharp contrast to the apparent situation for Frege proofs without extension.)

4.5. THEOREM. *Suppose $e\mathscr{F}$ and $e\mathscr{F}'$ are extended Frege systems over κ and κ', respectively, and suppose $L(n) \geq n$ is a natural number function such that every tautology A over κ has a proof π in $e\mathscr{F}$ with $\lambda(\pi) \leq L(l(A))$. Then every tautology A' over κ' has a proof π' in $e\mathscr{F}'$ such that $\lambda(\pi') \leq cL(cl(A'))$ and $\rho(\pi') \leq cl(A')$, where the constant c depends only on $\mathscr{F}$ and $\mathscr{F}'$.*

4.6. THEOREM (STATMAN)[1]. *For any extended Frege system $e\mathscr{F}$ and tautology A, if π is a proof of A in $e\mathscr{F}$, then there is a proof π' of A in $e\mathscr{F}$ such that $\lambda(\pi') \leq c(\lambda(\pi) + l(A))$ and $\rho(\pi') \leq cl(A)$, where the constant c depends only on $\mathscr{F}$.*

4.7. COROLLARY (TO THEOREM 4.5). *A given extended Frege system is polynomially bounded if and only if all extended Frege systems over all connective sets are polynomially bounded. Also, an extended Frege system $e\mathscr{F}$ is polynomially bounded if and only if there is a polynomial bound on the number of lines in proofs in $e\mathscr{F}$. Hence, if $\mathscr{P} \neq \mathscr{NP}$, then there is no polynomial bound on the number of lines in proofs in extended Frege systems, Frege systems, or (by §3) natural deduction systems.*

Propositions 4.5, 4.6, and 4.7 are evidence that the extended Frege systems are a very natural class of proof system. Further evidence is provided by results in [11], which show that extended Frege system proofs can simulate the proof of any theorem of a certain number theory system PV. ("Simulate" here means something similar to the way in which extended Frege proofs simulate the proof of the pigeon hole principle in the example given at the beginning of this section.) The same paper [11] shows that extended Frege systems are the most efficient systems whose soundness is provable in PV.

The remainder of this section is devoted to proving Theorems 4.5 and 4.6. Let us

[1]After proving a version of Theorem 4.5 without the bound on $\rho(\pi')$ in course notes [9], the first author received an earlier version of Statman [10] and realized the proof in the notes could be strengthened to yield the present Theorems 4.5 and 4.6. Statman's theorem in [10] has a more general setting than 4.6, but a weaker bound on $\lambda(\pi')$. The authors wish to thank Martin Dowd for helpful discussions concerning Theorem 4.6.

start by showing that a bound on proof length in an extended Frege system gives us a bound on derivation length.

4.8. LEMMA. *Suppose $e\mathcal{F}$ and $L(n)$ satisfy the hypotheses of Theorem 4.5. If $A_1, \ldots, A_m, B$ are formulas over κ such that $A_1, \ldots, A_m \models B$, then there is a derivation π in $e\mathcal{F}$ of B from $A_1, \ldots, A_m$ with $\lambda(\pi) \leq cL(cn)$, where $n = l(A_1) + \cdots + l(A_m) + l(B)$, and c depends only on $\mathcal{F}$.*

PROOF. Suppose first that the connective set κ of $\mathcal{F}$ contains $\vee$ and $\neg$. Since $A_1, \ldots, A_m \models B$, we have $\models (\neg A_1(\neg A_2 \vee \cdots \vee(\neg A_m) \vee B) \cdots))$. Hence this formula has a proof π' in $e\mathcal{F}$ with $\lambda(\pi') \leq L(n)$, $n = l(A_1) + \cdots + l(A_m) + l(B)$. If we assume $\mathcal{F}$ has the cut rule

$$\frac{P, \neg P \vee Q}{Q}$$

then by appending m applications of this rule to π', we obtain a derivation π of B from $A_1, \ldots, A_m$ satisfying the lemma, with $\lambda(\pi) \leq 2L(n)$. If the cut rule is not in $\mathcal{F}$, then by Theorem 2.3 the rule can be simulated to produce a derivation π with $\lambda(\pi) \leq cL(n)$.

If $\vee$ or $\neg$ is not in κ, one can check that nevertheless there are formulas $O(P, Q)$ and $N(P)$ over κ equivalent to $P \vee Q$ and $\neg P$, respectively, such that $O(P, Q)$ and $N(P)$ have at most one occurrence each of P and Q (see 3.1). In this case we obtain the bound $\lambda(\pi) \leq cL(cn)$. □

To prove Theorems 4.5 and 4.6 we need the notion of a defining set of formulas def(A) for a formula A. We assume that every formula B (over any connective set) has associated with it an atom P_B such that P_Q is Q for any atom Q, and distinct nonatomic formulas have distinct associated atoms. To be definite, we could let P_B be the string consisting of the letter P followed by the string B, if B is nonatomic. In any case, we shall also assume for convenience later, that there are infinitely many atoms P, called *admissible* atoms, which are *not* of the form P_B for any nonatomic B.

Let us call a formula A *admissible* if all its atoms are admissible. If A is admissible, then every truth assignment τ to the atoms of A has a unique extension τ' to the atoms P_B, B any subformula of A, such that $\tau'(P_B) = \tau(B)$. We shall define def(A) such that any extension τ'' of τ satisfies def(A) iff τ'' agrees with τ' on the atoms P_B. For example, if A is $Q \vee (R \,\&\, S)$, then def(A) might be $\{(P_{(R\&S)} \equiv (R \,\&\, S)), (P_A \equiv Q \vee P_{(R\&S)})\}$. In fact, it is useful to more generally define $\text{def}_\kappa(A)$, where κ is any adequate set of connectives, perhaps different from the set of connectives appearing in A.

4.9. DEFINITION. Let κ_1 and κ_2 be connective sets. Corresponding to each nullary connective (constant) K_1 in κ_1 we associate a fixed formula K_2 over κ_2 equivalent to K_1; corresponding to each unary connective u_1 over κ_1 we associate a fixed formula $u_2 P$ over κ_2 equivalent to $u_1 P$, and corresponding to each binary connective $\circ_1$ in κ_1 we associate a fixed formula $P \circ_2 Q$ over κ_2 equivalent to $P \circ_1 Q$. We assume the formulas $P \sim_1 Q$ over κ_1 and $P \sim_2 Q$ over κ_2 are each equivalent to $P \equiv Q$. For each formula A_1 over κ_1 we associate a set $\text{def}_{\kappa_2}(A_1)$ of formulas over κ_2 defined by induction on the length of A_1 as follows:

$\text{def}_{\kappa_2}(P) = \varnothing$ (the empty set) for each atom P.

$\text{def}_{\kappa_2}(K_1) = \{P_{K_1} \sim_2 K_2\}$ for each constant K_1 in κ_1.

$\mathrm{def}_{\kappa_2}(u_1A) = \mathrm{def}_{\kappa_2}(A) \cup \{P_{u_1A} \sim_2 u_2P_A\}$, for each unary connective u_1 in κ_1.

$\mathrm{def}_{\kappa_2}(A \circ_1 B) = \mathrm{def}_{\kappa_2}(A) \cup \mathrm{def}_{\kappa_2}(B) \cup \{P_{A\circ_1 B} \sim_2 P_A \circ_2 P_B\}$, for each binary connective $\circ_1$ in κ_1.

In case $\kappa_1 = \kappa_2$, we assume $K_1 = K_2$, $u_1 = u_2$, and $\circ_1 = \circ_2$. It is easy to check that the total number of occurrences of atoms in $\mathrm{def}_{\kappa_2}(A)$ is bounded by a linear function of $l(A)$.

4.10. LEMMA. *Suppose $e\mathscr{F}$ is an extended Frege system over κ, A is an admissible formula over κ, and $\mathrm{def}_\kappa(A) \vdash^{\pi}_{e\mathscr{F}} P_A$. Then for some π' we have $\vdash^{\pi'}_{e\mathscr{F}} A$, where $\lambda(\pi') \le \lambda(\pi) + cl(A)$ and $\rho(\pi') \le (\rho(\pi) + c)l(A)$, and c depends only on $\mathscr{F}$.*

PROOF. Let σ be the simultaneous substitution E/P_E for all nonatomic subformulas E of A, so in particular $\sigma P_A = A$. Then every formula in $\sigma(\mathrm{def}_\kappa(A))$ is an instance of $P \sim P$, and each of these instances will have a proof in $\mathscr{F}$ of some fixed number of lines, and a number of atoms bounded by a constant times $l(A)$. These proofs, together with $\sigma(\pi)$, comprise π'. □

4.11. LEMMA. *If $e\mathscr{F}$ and $e\mathscr{F}'$ are extended Frege systems over κ and κ' respectively, A' is an admissible formula over κ', and $\mathrm{def}_\kappa(A') \vdash^{\pi}_{e\mathscr{F}} P_{A'}$, then for some derivation π', $\mathrm{def}_{\kappa'}(A') \vdash^{\pi'}_{e\mathscr{F}'} P_{A'}$, where $\lambda(\pi') \le c\lambda(\pi)$ and $\rho(\pi') \le d$, and the constants c and d depend only on $\mathscr{F}$ and $\mathscr{F}'$.*

PROOF. Suppose π is $B_1, \ldots, B_m$. We may assume, by renaming if necessary, that all atoms of each B_i are admissible, except possible those which occur in the hypotheses or conclusion of π (i.e. except those of the form $P_{C'}$, where C' is a subformula of A'). We shall construct the derivation π' in $e\mathscr{F}'$ by filling out the skeleton derivation $P_{B_1}, \ldots, P_{B_m}$. (Notice that P_{B_m} is $P_{A'}$, since B_m is $P_{A'}$ and in general $P_Q = Q$ for any atom Q.) In fact, we shall show that for some constants c and d depending only on $\mathscr{F}$ and $\mathscr{F}'$, each P_{B_i} can be derived from earlier P_{B_j}'s and $\mathrm{def}_{\kappa'}(A')$ in at most c lines by formulas C with $l(C) \le d$.

To see how to derive P_{B_i} in π' we consider three cases, depending on how B_i was obtained in π. For each of these cases we assume that some of the formulas of $\mathrm{def}_{\kappa'}(B_i)$ are available in π', either because they are among the hypotheses $\mathrm{def}_{\kappa'}(A')$ of π' or because they are introduced at the beginning of π' by the extension rule. The defining formula for P_C, where C is a subformula of B_i, is in $\mathrm{def}_{\kappa'}(A')$ if C is also a subformula of A'. If C is not a subformula of A', then the defining formula for P_C can legally be included in π' by the extension rule.

Case I. B_i is a hypothesis for π, so B_i is in $\mathrm{def}_\kappa(A')$. We may assume B_i has the form $P_{C'} \sim (P_{D'} \circ P_{E'})$, where C', D', E' are subformulas of A', $P \circ Q$ is the fixed formula over κ equivalent to $P \circ' Q$, and C' is $D' \circ' E'$. (The cases of unary and 0-ary connectives are similar.) Then $P_{C'} \sim' (P_{D'} \circ' P_{E'})$ is in $\mathrm{def}_{\kappa'}(A')$, and so is a hypothesis of π'. Let $H(\circ')$ be the formula $P \sim (Q \circ R)$ over κ. Note that $H(\circ')$ depends only on the connective $\circ'$, and not otherwise on B_i. Then the rule

$$R = \frac{P \sim'(Q \circ' R),\ \mathrm{def}_{\kappa'}(H(\circ'))}{P_{H(\circ')}}$$

is sound, so by Theorem 2.3 we may assume it is a rule of $\mathscr{F}'$. Let σ be an extension of the substitution

$$\frac{P_{C'}, P_{D'}, P_{E'}, P_{B_i}}{P,\ \ Q,\ \ R,\ P_{H(\circ')}}$$

such that $\sigma(\mathrm{def}_{\kappa'}(H(\circ'))) = \mathrm{def}_{\kappa'}(B_i)$. Then P_{B_i} follows in one step by R and σ from $\mathrm{def}_{\kappa'}(A')$ and $\mathrm{def}_{\kappa'}(B_i)$.

Case II. B_i is introduced in π by the extension rule. Then B_i has the form $P \sim C$, where P is a new defined atom. The constraints governing the use of the extension rule imply that P does not occur in the hypotheses or conclusion of π, and by our assumption at the beginning of this proof, P is admissible. Therefore, P does not occur in the hypotheses or conclusion of π'. We note that the formula $P \sim' P_C$, together with any subset of the formulas of $\mathrm{def}_{\kappa'}(B_i)$ not introduced earlier could be introduced by the extension rule in π', after any necessary formulas of $\mathrm{def}_{\kappa'}(B_{i-1})$ and before formulas of $\mathrm{def}_{\kappa'}(B_{i+1})$ are introduced. The order of introduction could be $\mathrm{def}_{\kappa'}(C)$, $P \sim' P_C$, followed by one or more formulas whose conjunction is equivalent to $P_{B_i} \sim' (P \sim' P_C)$. This last formula itself will be in $\mathrm{def}_{\kappa'}(B_i)$ if $\equiv$ is in κ, in which case B_i is $P \equiv C$. In this case, it follows from Theorem 2.3 that P_{B_i} can be deduced in a bounded number of bounded steps in π' from $P \sim' P_C$ and $P_{B_i} \sim' (P \sim' P_C)$. If $\equiv$ is not in κ, there are nevertheless a bounded number of formulas in $\mathrm{def}_{\kappa'}(B_i)$ which imply $P_{B_i} \sim' (P \sim' P_C)$, and the number and structure of these formulas depends only on the way $\equiv$ is represented in κ and κ'. Hence again P_{B_i} can be deduced in π' from $\mathrm{def}_{\kappa'}(B_i)$ and $P \sim' P_C$ by a bounded number of bounded formulas.

Case III. B_i follows from earlier formulas in π by a rule $R = (C_1, \ldots, C_k)/D$ in $\mathscr{F}$ by the substitution σ. Then $C_1, \ldots, C_k \models D$, so the rule

$$R' = \frac{\mathrm{def}_{\kappa'}(D), \mathrm{def}_{\kappa'}(C_1), \ldots, \mathrm{def}_{\kappa'}(C_k), P_{C_1}, \ldots, P_{C_k}}{P_D}$$

is sound, and by Theorem 2.3 we may assume it is a rule of $\mathscr{F}'$. We may assume all formulas $C_1, \ldots, C_k, D$ are admissible. Let σ' be the composition of the substitutions $\sigma(E)/P_E$ for all subformulas E of formulas in the set $\{C_1, \ldots, C_k, D\}$. Then $\sigma'(\mathrm{def}_{\kappa'}(C_j)) \subseteq \mathrm{def}_{\kappa'}(\sigma(C_j))$, $1 \le j \le k$, and $\sigma'(\mathrm{def}_{\kappa'}(D)) \subseteq \mathrm{def}_{\kappa'}(\sigma(D))$. Of course each $\sigma(C_j)$ is some B_l, $l < i$, and $\sigma(D)$ is B_i. By the induction hypothesis $P_{\sigma(C_j)}$ occurs earlier in π'. Hence P_{B_i} follows by R' and σ' from earlier formulas π' and a bounded number of formulas from $\mathrm{def}_{\kappa'}(B_l)$, for various B_l.

This completes the proof of Lemma 4.11.

Now assume the hypotheses of Theorem 4.5, and let A' be any valid formula over κ'. We may assume A' is admissible, for if not, we may rename the atoms in A' so that it is admissible, find a suitable proof of the result, and then rename all atoms in the proof to obtain a suitable proof of A'. Then $\mathrm{def}_\kappa(A') \models P_{A'}$, so by hypothesis, the bounds on $l(\mathrm{def}_\kappa(A'))$, and Lemma 4.8, there is a derivation π in $e\mathscr{F}$ of $P_{A'}$ from $\mathrm{def}_\kappa(A')$ such that $\lambda(\pi) \le c_1 L(c_1 l(A'))$. By Lemma 4.11, there is a derivation π' in $e\mathscr{F}'$ of $P_{A'}$ from $\mathrm{def}_{\kappa'}(A')$ such that $\lambda(\pi') \le c_2 L(c_1 l(A'))$ and $\rho(\pi') \le d$. Theorem 4.5 now follows by Lemma 4.10.

To prove Theorem 4.6, we may assume as above that A is admissible. By induction on the length of B, it is easy to see that for every admissible formula B over κ there is a derivation π_B in $\mathscr{F}$ of $P_B \sim B$ from $\mathrm{def}_\kappa(B)$ such that $\lambda(\pi_B) \le c_1 l(B)$ and $\rho(\pi_B) \le c_2 l(B)$, where κ is the connective set of $\mathscr{F}$. By putting together π_A with π in the theorem, we obtain a derivation π_1 of P_A from $\mathrm{def}_\kappa(A)$ such that $\lambda(\pi_1) \le c_3(\lambda(\pi) + l(A))$ and $\rho(\pi_1) \le c_4 l(\pi)$. We now apply Lemma 4.11 with $\kappa' = \kappa$,

$e\mathscr{F}' = e\mathscr{F}$, and $\pi = \pi_1$ to modify π_1 so its formulas have bounded length, and finally apply Lemma 4.10 to the resulting derivation. □

§5. The Substitution Rule. Frege's original propositional proof system [6] tacitly assumed the following:

5.1. *Substitution Rule.* From A conclude σA, for any substitution σ in the notation of the system.

5.2. Definition. A *Frege system with substitution*, $s\mathscr{F}$, is obtained from a Frege system $\mathscr{F}$ by addition of the substitution rule. Hypotheses are not allowed in derivations in $s\mathscr{F}$.

The reason hypotheses are not allowed in $s\mathscr{F}$-derivations is that in general not $A \models \sigma A$. Thus the substitution rule is unsound in this sense. On the other hand, if $\models A$ then $\models \sigma A$, so if $\vdash_{s\mathscr{F}} A$ then $\models A$. In other words, $s\mathscr{F}$ is a sound system for proving tautologies, but not for deriving formulas from hypotheses.

The theorem below shows that Frege systems with substitution can p-simulate extended Frege systems. The converse may be false, however. (We conjecture Frege systems with substitution are not p-verifiable in the sense of [11], whereas extended Frege systems are p-verifiable.)

5.3. Theorem. *Given an extended Frege system $e\mathscr{F}$ there is a function f in $\mathscr{L}$ and constant c such that for all proofs π and formulas A, if $\vdash^{\pi}_{e\mathscr{F}} A$,* then $\vdash^{f(\pi)}_{s\mathscr{F}} A$, and $\lambda(f(\pi)) \le c\lambda(\pi)\rho(\pi)$ and $\rho(f(\pi)) \le c\lambda(\pi)\rho(\pi)$.

Proof. Suppose $P_1 \sim C_1, \ldots, P_k \sim C_k$ are the defining formulas introduced by extension in π. As discussed before Theorem 3.3, $\mathscr{F}$ can be turned into a natural deduction system $\mathscr{N}$ by including the rules

$$R_1 = \frac{P \to Q}{\to \neg P \vee Q} \quad \text{and} \quad R_2 = \frac{\to \neg P \vee Q}{P \to Q}.$$

Let us assume in addition that $\mathscr{N}$ has the rules

$$R_3 = \frac{P \to Q}{P,\, R \to Q} \quad \text{and} \quad R_4 = \frac{P \to Q}{R,\, P \to Q},$$

and the axiom $P \to P$. Then for each i, $1 \le i \le k$, the line $E_1, \ldots, E_k \to E_i$ can be derived from the axiom and $k - 1$ uses of R_3 and R_4, for any formulas $E_1, \ldots, E_k$. The derivations of these k lines, together with π, describe a derivation π_1 in $\mathscr{N}$ of $E_1, \ldots, E_k \to A$, where now E_i is the defining formula $P_i \sim C_i$, and $\lambda(\pi_1) \le \lambda(\pi) + k^2$ and $\rho(\pi_1) \le (k+1)\rho(\pi)$. Now by adding k applications of rule R_1, we obtain a derivation in $\mathscr{N}$ of B, where B is $\neg E_1 \vee (\neg E_2 \vee \cdots \vee (\neg E_k \vee A)\cdots)$. Hence noting $k < \lambda(\pi)$, we have by the proof of corollary 3.4 a derivation π_2 in $\mathscr{F}$ of B, where $\lambda(\pi_2) \le c_1(\lambda(\pi))^2$ and $\rho(\pi_2) \le c_1\lambda(\pi)\rho(\pi)$. Now assume the defining formulas $P_1 \sim C_1, \ldots, P_k \sim C_k$ are numbered in reverse of the order in which they appear in π. Then $P_1 \sim C_1$ appears last, so P_1 has no occurrence in any C_i or in A. By applying the substitution rule to B with the substitution C_1/P_1, and applying the Frege rule $(\neg(P \sim P) \vee Q)/Q$, we can derive $(\neg E_2 \vee \cdots \vee (\neg E_k \vee A)\cdots)$ from B. By $k - 1$ further applications of the substitution rule and this Frege rule, each of the E_i's can be pruned, and we obtain a proof of A in $s\mathscr{F}$ which satisfies the conditions of the theorem. □

By combining the above theorem with Theorem 4.5, we obtain the following.

5.4. COROLLARY. *If there exists a polynomially bounded extended Frege system, then all Frege systems with substitution over all connectives sets are polynomially bounded.* □

A result similar to Theorem 4.5 can be proved for Frege systems with substitution, using the methods in that proof and in the above argument. In particular, one Frege system with substitution is polynomially bounded if and only if all such systems over all connective sets are polynomially bounded. Reckhow [2] proves this result by different methods.

REFERENCES

[1] S. A. COOK and R. A. RECKHOW, *On the lengths of proofs in the propositional calculus, Preliminary version*, ***Proceedings of the Sixth Annual ACM Symposium on the Theory of Computing, May*** **1974**, pp. 135–148. (See also corrections for the above in ***SIGACT News***, vol. 6(1974), pp. 15–22.)

[2] R. A. RECKHOW, *On the lengths of proofs in the propositional calculus*, Ph.D. Thesis, Department of Computer Science, University of Toronto, 1976.

[3] S. A. COOK, *The complexity of theorem proving procedures*, ***Proceedings of the Third Annual ACM Symposium on the Theory of Computing, May*** **1971**, pp. 151–158.

[4] R. M. KARP, *Reducibility among combinatorial problems*, ***Complexity of computer computations*** (R. E. Miller and J. W. Thatcher, Editors), Plenum Press, NY, 1972, pp. 85–103.

[5] A. V. AHO, J.E. HOPCROFT and J. D. ULLMAN, ***The design and analysis of computer algorithms***, Addison-Wesley, Reading, Massachusetts, 1974.

[6] G. FREGE, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Halle, 1879. English translation in ***From Frege to Godel, a source book in mathematical logic*** (J. van Heijenoort, Editor), Harvard University Press, Cambridge, 1967, pp. 1–82.

[7] D. PRAWITZ, ***Natural deduction. A proof-theoretic study***, Stockholm, 1965.

[8] G. S. TSEITIN, *On the complexity of derivations in propositional calculus*, ***Studies in mathematics and mathematical logic, Part*** II (A. O. Slisenko, editor), 1968, pp. 115–125. (Translated from Russian)

[9] S. A. COOK, C. W. RACKOFF ET AL, Lecture notes for CSC2429F, "Topics in the theory of computation", a course presented by the Department of Computer Science, University of Toronto during the fall, 1976.

[10] R. STATMAN, *Complexity of derivations from quantifier-free horn formulae, mechanical introduction of explicit definitions, and refinement of completeness theorems*, ***Proceedings of Logic Colloquium*** **'76** (Gandy and Hyland, Editors), ***Studies in Logic and Foundations of Mathematics***, vol. 87(1977), North-Holland, Amsterdam.

[11] S. A. COOK, *Feasibly constructive proofs and the propositional calculus*, ***Proceedings of the Saventh Annual ACM Conference on the Theory of Computing, May*** **1976**, pp. 83–97.

[12] A. CHURCH, ***Introduction to mathematical logic***, vol. I, Princeton University Press, Princeton, 1956.

[13] J. VON NEUMANN, *Zur Hilbertschen Beweistheorie*, ***Mathematische Zeitschrift***, vol. 26 (1927), pp. 1–46.

DEPARTMENT OF COMPUTER SCIENCE
UNIVERSITY OF TORONTO
TORONTO, CANADA M5S 1A7

DEPARTMENT OF COMPUTING SCIENCE
UNIVERSITY OF ALBERTA
EDMONTON, CANADA T6G 2E1