Direttore Responsabile: Ruggero Ferro
Iscrizione al Registro Stampa del Tribunale di Padova n. 1235 del 26.9.1990

Pubblicato con il contributo di:

Cassa di Risparmio di Padova e Rovigo

Stampa Veronese (Padova)
Laser Fotocomposizioni (Padova)

MECHANISMS AND SEARCH

ASPECTS OF PROOF THEORY

Wilfried Sieg
Department of Philosophy
Carnegie Mellon University
Pittsburgh, Pennsylvania

PREFACE.

The material of these notes was presented in lectures I gave in Milano in May 1992. Some of the material in the first, third, and fourth lectures had been developed for courses in Siena and München in the Spring of 1988, but the remainder is based on papers and manuscripts written during the last three years in Pittsburgh.

My reasons for selecting the material are elaborated in the Introduction. Here I simply say that I attempted to give a *partial snapshot* of proof theory *from one particular perspective* by describing three themes that hang together quite intimately: foundational reduction, computational information, and (heuristics in the) automated search for proofs. These are themes that were emphasized in the twenties, but have been developed more distinctively only since the fifties. *Technically* the themes are held together by the possibility of normalizing proofs and thus, in the case of first order logic, of bounding the logical complexity of formulas occurring in them. But these themes are also held together *conceptually*: That is the rationale for including an unusual amount of philosophical and historical material.

Thanks are due to Wilfried Buchholz and Helmut Schwichtenberg who made my sabbatical stay in München possible (during the academic year 1987-88), to Franco Montagna who invited me to Siena, and – most of all – to Daniele Mundici who was my host in Milano. Mundici encouraged me during my wonderfully intense and vibrant visit in Milano to complete these notes. I am also grateful to the students in Siena, München, and Milano who attended and criticized my lectures. Mario Chiari helped me to bring the early material into proper shape; Connie Bartusis proofread all lectures with a keen eye and critical pen, improving their style considerably. Finally, I want to acknowledge the financial support of CNR, DFG, and Carnegie Mellon University.

Pittsburgh, July 1, 1992

TABLE OF CONTENTS.

INTRODUCTION

If one tries to characterize what is distinctive about logic in our century one clearly has to point to its close *association with mathematics*: Logic has been using mathematical tools in its presentation and critical self-examination, and mathematics has been logic's primary field of application and source of problems. Yet underneath the mathematical shell, the *philosophical origins* of fundamental issues have been preserved to a great extent. It is the glory of logic that it complements formal mathematical work by informal rigorous reflection. Here are three prime examples: (1) the analysis of "logical consequence" (in its semantic and syntactic guise from Aristotle to Frege, Hilbert, Gödel, and Tarski); (2) the analysis of "set" (from Cantor and Dedekind through Zermelo's cumulative hierarchy to constructible sets – in both Gödel's technical sense and the informal sense); (3) the analysis of "formality" (from the quasi-normative requirements in Leibniz to Turing's Thesis and subsequent generalizations). These examples are not isolated from the rest of logic, but actually constitute its core of permanent contributions; they are not isolated either from each other, but are deeply connected through questions concerning the nature of mathematical experience and, ultimately, the nature of the human mind. It was the concern with these general philosophical questions that led, in the very first place, to the methodological emphasis on constructivity in mathematics and on effectiveness in metamathematics. Not surprisingly, this has led to developments that are of increasing significance in *computer science*.

With respect to all of these issues Hilbert had a directing influence in the twenties and even earlier. As to (1), he formulated most clearly the completeness problem; as to (2), he emphasized that the axiomatic method should be applied to the notion of "set" and inspired Zermelo, but also von Neumann and Bernays; finally, as to (3), he formulated sharply the decision problem for predicate logic and viewed it as a fundamental problem. I want to emphasize this in contradistinction to the conventional view that ties Hilbert's foundational work exclusively to his PROGRAM. Clearly, Hilbert's desire to settle foundational problems in mathematics by finitist consistency proofs was important and, indeed, it was for the purpose of this program that he quite literally invented a new subject, namely PROOF THEORY. In my view, one can discern *three main themes* of proof theoretic research: **(1)** the reductive foundational, (2) the informative mathematical (computational), and (3) the cognitive psychological.

The first theme has been the dominant one, because of the *foundational aims* of Hilbert's program. In the early twenties Hilbert set himself the task of securing the instrumental usefulness of *all* of classical mathematics. He hoped to achieve that aim by *reducing* analysis and even set theory to a fixed, absolutely fundamental part of arithmetic, so-called finitist mathematics. The specific proposal of how to achieve such a reduction is the mathematical centerpiece of Hilbert's program: A finitist consistency proof would allow the transformation of "classical", set theoretic proofs of finitist statements into finitist proofs. Note that this takes on a methodologically most important problem Dirichlet had posed through his use of analytic methods in number theory! The program was refuted fortunately or unfortunately by Gödel's Incompleteness Theorems. A *general reductive program* has been pursued, and significant progress has been made in its pursuit. The aim of obtaining an "absolute reduction" for *all* of classical to finitist mathematics had to be replaced, however, by the more modest task of establishing the consistency of theories for parts of classical mathematics relative to suitable constructive theories. As that amounts to establishing partial conservativeness, the *main question* really is: **What more than finitist mathematics do we have to know to recognize the (partial) correctness of a strong classical theory?**

The theme concerning mathematical and, later, computational information was at first completely subsidiary to the first theme: One had to see that certain formal theories could serve as frames for mathematical practice. Whitehead and Russell's *Principia Mathematica* and the set theoretic developments in mathematics were taken by Hilbert as evidence that type theory or set theory could serve for that purpose. Gödel's First Incompleteness Theorem established that arithmetical truth cannot be fully captured by derivability in formal theories and it removed one of the crucial (implicit) assumptions of Hilbert. However, by exposing the limitations of formal methods Gödel's result opened the possibility of exploiting *formal* proofs to obtain "information" beyond establishing the truth of a theorem. If the *proper* inclusion of provability in truth is to be exploited, it seems that it is best to use weak theories that are nevertheless adequate for the formalization of mathematical practice. As a matter of fact, the presentation

of analysis given by Hilbert (during the early twenties in second order arithmetic) can be viewed in this light as an important first step. Refinements during the subsequent fifty years have made clear that *all* of classical analysis can be carried out in theories that are reducible to elementary arithmetic; *parts* of analysis and also of algebra can be carried out in even weaker theories. Joining such quasi-empirical investigations with proof theoretic work allows then the in-principle-extraction of detailed "computational information". That comes under the heading of *provably recursive (or provably total) functions*; i.e., one determines exactly the class of those recursive functions whose termination can be proved in the formal theory at hand. Such results give (in general, crude) bounds from proofs of Π^0_2-theorems and, turning the table, are used to prove the independence of such theorems. In any event, here we have one way of answering the *main question* : **What more than its truth do we know, if we have proved a theorem in a weak formal theory ?**

The third theme is intimately connected with the mechanical modelling of reasoning in the tradition of Leibniz, and, to a certain extent, Frege. This theme was definitely taken up by Hilbert himself; in "Über das Unendliche" he claimed:

> The formula game that Brouwer so deprecates has, besides its mathematical value, an important general philosophical significance. For this formula game is carried out according to certain definite rules, in which the technique of our thinking is expressed. These rules form a closed system that can be discovered and definitively stated. The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds.

If anything is an early formulation of goals for contemporary cognitive psychology, this is. The claims were made (somewhat) plausible only by Gentzen's development of the calculi of natural deduction. In German they are called "Kalküle des natürlichen Schließens" emphasizing that they (are to) *correspond* to an argumentative practice that comes naturally. Strangely enough (and it is indeed surprising, even if one takes into account the variety of different aims that are being pursued), this tradition has hardly influenced the *(automated) theorem proving systems* of today; for them a different tradition in proof theory has been more important, namely one that is reflected in Herbrand's theorem and related results. Here the *main question* is (or rather should be): **What more than the formal rules of a calculus should a computer know, when searching for a proof of a statement?**

In my lectures I want to illustrate these *central* themes paradigmatically and discuss some of the answers in technical detail – clearly, in areas where I feel competent and where I can add to what is in the literature. These considerations underlie my selection of topics, and I apologize at the outset for the wholesale omission of, e.g., *systems of ordinal notations* or Π^1_2-*Logic* or *Linear Logic*. For each of the questions I do consider, answers can be obtained by investigating suitable formalisms in a variety of ways. It turns out that one approach to the original programmatic consistency problem is particularly successful. It is due to Gentzen and involves the representation of reasoning in *special calculi*, that is, sequent calculi and natural deduction calculi; for both kinds of calculi crucial NORMAL FORM THEOREMs can be established.

Sequent calculi were used by Gentzen to give consistency proofs for (parts of) arithmetic, and they have been used ever since for stronger and stronger subsystems of analysis in the pursuit of theme **(1)** (e.g., Schütte, Takeuti, Tait, Feferman, Buchholz, Pohlers, Sieg). They have also been employed in work on theme **(2)**, as witnessed by Schwichtenberg's beautiful reformulation of Kreisel's early work on the characterization of the provably recursive functions of Peano arithmetic and by the more recent work (e.g., of Buss) on the proof theoretic characterization of complexity classes. In relation to theme **(3)**, I point to Hao Wang's work in automated theorem proving; he exploited already in the early sixties a particular way of establishing the completeness of sequent calculi *without cut*. Let me note that the single most fundamental fact (and most useful for applications) is the subformula property of normal derivations; it is a direct consequence of the *normal form theorem*. This property guarantees the crucial bounding (of the logical complexity) of formulas that may occur in derivations.[1]

[1] References to the literature will be given throughout these notes; let me mention some pertinent sources: for (1) [Kreisel 1958A & 1968] and [Sieg 1988 & 1990]; for (2) [Kreisel 1951 & 1958B], [Parsons 1970 & 1972], [Sieg 1985 & 1991], and [Buss 1986]; for (3) [Gallier 1986] and [Fitting 1991].

PART A. BACKGROUND.

1. Proof theoretic perspectives. After depicting themes and surveying topics, let me start out with some historical remarks on the context in which Hilbert's program arose, because it is still widely and deeply misunderstood as an ad hoc weapon against the growing influence of Brouwer's intuitionism.

Reductive programs. The problems that motivated Hilbert's program can be traced back to the central foundational issue in 19th century mathematics, namely securing a basis for analysis. A possible resolution was indicated by the slogan "Arithmetize analysis!" That direction was given already by Gauss, and its meaning can be fathomed from Dirichlet's claim that any theorem of analysis can be formulated as a theorem concerning the natural numbers. For some the arithmetization of analysis was accomplished by the work of Cantor, Dedekind, and Weierstrass; for others, e.g., Kronecker, a stricter arithmetization was required, one which would base the whole content of all mathematical disciplines (with the exception of geometry and mechanics) on "the concept of number taken in its most narrow sense, and thus to strip away the modifications and extensions of this concept, which have been brought about in most cases by applications in geometry and mechanics" ([Kronecker 1887], p. 253). In a footnote, Kronecker makes clear that he has in mind "in particular the addition of the irrational and continuous magnitudes". Kronecker strongly opposed Cantor's and Dedekind's free use of set theoretic notions, as it violated methodological restrictions on "legitimate" mathematical concepts and arguments.

Having been informed (by Cantor in 1897) about the problematic character of some set theoretic considerations and the inconsistency of Dedekind's "Was sind und was sollen die Zahlen", Hilbert addressed the issues directly in his paper "Über den Zahlbegriff" and again in his Paris lectures of 1900. His goal was to establish by a consistency proof the existence of the set of natural and real numbers and of the Cantorian alephs; but he gave only a *very* rough indication, how such a proof could be carried out: Provide models for an axiomatic characterization of the reals and the alephs. In his Heidelberg address of 1904 Hilbert gave up this first attempt at circumventing the Cantorian problems in set theory as far as they affected analysis.[2] The then recently discovered elementary contradictions of Zermelo and Russell had changed his outlook on these problems. Bernays is quoted in Reid's biography of Hilbert as saying:

> Under the influence of the discovery of the antinomies in set theory, Hilbert temporarily thought that Kronecker had probably been right there. [I.e., right in insisting on restricted methods.] But soon he changed his mind. Now it became his goal, one might say, to do battle with Kronecker with his own weapons of finiteness by means of a modified conception of mathematics. .. ([Reid 1970], p. 173)

The key question was, how might that be done? The radicalization of the axiomatic method, high lighted in his own *Grundlagen der Geometrie*, and the fresh developments in logic due to Frege and Peano provided the basic background for Hilbert's way of answering this question. The ultimate goal of his proposal in 1904 was the same as the one he had formulated earlier. But now Hilbert indicated a possibility of giving consistency proofs without presupposing set-theoretic notions. He proposed a simultaneous *formal* development of logic and arithmetic, so that proofs could be viewed as *finite* mathematical stuctures. The new task was to show by mathematical means that such formal proofs could not lead to a contradiction. But neither the formal logical apparatus was clearly specified, nor was there an explicit concern about the mathematical means needed to prove such facts.

This formulation foreshadowed aspects of the proof-theoretic program Hilbert pursued in the twenties together with, e.g., Bernays, Ackermann, von Neumann, Herbrand. There was, however, a crucial and sophisticated shift in what a consistency proof was to establish and how it was to be given. To bring this out , let **P** be a formal theory in which mathematical practice can be represented and let **F** be a theory formulating principles of finitist mathematics. Under weak assumptions on **P** (satisfied by the usual formal theories) the consistency statement for **P** is equivalent to the reflection principle

$$\mathrm{Pr}(a,\overline{\sigma(\psi)}) \rightarrow \psi$$

Pr is the canonical proof predicate for **P**, $\sigma(\psi)$ indicates (the Gödel-number of) the translation of the **F**-statement ψ into the language of **P**, and $\overline{\sigma(\psi)}$ is the corresponding numeral in that language. Proving the reflection principle in **F** amounts to recognizing – from the restricted standpoint of **F** – the truth of the **F**-statements whose translations have been derived in **P**. As a matter of fact, the proof would yield a method of turning any **P**-proof of $\sigma(\psi)$ into an **F**-proof of ψ. Finitist mathematics was viewed as a fixed part of elementary arithmetic and its philosophical justification seemed to be unproblematic. Thus Hilbert thought that the consistency proof for **P** would solve the foundational problems "once and for all" by mathematical considerations. Bernays emphasized in 1922: "This is precisely the great advantage of Hilbert's proposal, that the problems and difficulties arising in the foundations of mathematics are transferred from the epistemological-philosophical to the genuinely mathematical domain".

The radical foundational aims of Hilbert's program had to be abandoned on account of Gödel's Incompleteness Theorems. A "generalization" of the program was developed in response to Gödel's results, and it has been pursued with great vigor and mathematical success for parts of analysis.[3] The basic task of the generalized reductive program can be seen as follows: Find for a significant part of classical mathematical practice, formalized in a theory **P***, an appropriate constructive theory **F***, such that **F*** proves the partial reflection principle for **P***. That is, **F*** proves for any **P***-derivation D

$$\mathrm{Pr}^*(\overline{\mathrm{D}},\overline{\sigma(\psi)}) \rightarrow \psi;$$

and ψ is in a class Λ of **F***-statements. It follows immediately that **P*** is conservative over **F*** with respect to the statements in Λ; consequently, **P*** is consistent relative to **F***. (I made the assumption satisfied by the theories discussed below, that **F*** is easily seen to be contained in **P***. If this is not the case, reductions in both directions have to be established.) The Gödel Gentzen reduction of classical elementary arithmetic (**Z**) to its intuitionistic version (**HA**) is the early paradigm of a successful contribution to the generalized program. Clearly, (**Z**) is taken as **P***, (**HA**) as **F***, and Λ consists of all negative arithmetic and Π^0_2-sentences. It was incidentally this result that showed to the Hilbert school that intuitionistic and finitist reasoning did *not* coincide, "contrary to the prevailing views at the time" as Bernays put it[4]. In addition, it gave an important positive impetus to(wards) the generalized program.

[1] For details about these early considerations, see [Sieg 1990).

[3] Bernays and Kreisel were highly influential in this development; for relatively recent and polished formulations see [Bernays 1970], pp.186-187 and [Kreisel 1968], pp.321-323.

[4] [Bernays 1967], p. 502.

> It thus became apparent that the "finite Standpunkt" is not the only alternative to classical ways of reasoning and is not necessarily implied by the idea of proof theory. An enlarging of the methods of proof theory was therefore suggested: instead of a restriction to finitist methods of reasoning, it was required only that the arguments be of a constructive character, allowing us to deal with more general forms of inferences. [5]

The questions that had sweeping general answers in the original Hilbert program had to be addressed anew, indeed in a much more subtle way. Which parts of classical mathematical practice can be represented in a certain theory **P***? What are (the grounds for) the principles in the "corresponding" constructive **F***? Briefly put, if a metamathematical conservation result has been obtained, it has to be complemented by additional mathematical and philosophical work establishing its foundational interest by answering these questions. Classical analysis was viewed as *decisive* for the generalized program, and its basic notions and results were presented carefully and in detail by Hilbert and Bernays in Supplement IV of their *Grundlagen der Mathematik II* .

Analysis or second order arithmetic. The extremely elegant formalism used by Hilbert and Bernays involves the ε-calculus and is, essentialy, equivalent to the theory (**AC**) described below. I will give now a description of the standard formal frame for second order arithmetic; its basic structure is the septuple

$$< \mathbf{N}, \mathbf{N}^{\mathbf{N}}, \ 0, ', \ <,> , ()_0 , ()_1 >$$

Thus we consider natural numbers and unary functions from **N** to **N** as the basic objects of the theory. Alternatively, one can consider sets of natural numbers as the second order entities; the latter can be represented in our framework by their characteristic functions. <, > is a pairing function; $(\ )_0$ and $(\ )_1$ are the corresponding projection functions. For convenience we add a standard enumeration $<f_j>_{j\varepsilon \mathbf{N}}$ of the unary primitive recursive functions, turning the septuple into an octuple. The language $\mathcal{L}^2$, appropriate for this structure, contains the language $\mathcal{L}$ of elementary number theory: x,y,z,... are used as individual variables; a,b,c,... as individual parameters; 0, ', < , >, $(\ )_0$, $(\ )_1$, f_j as constants. *Terms* are built up in the usual way: Using s,t,.... as syntactic variables over terms, we call *numerical equations* expressions of the form s=t. *Formulas* are obtained from numerical equations and inequalities by closing under ∧,∨,∃,∀. The connectives →,↔, and the negation of complex formulas are definable. To expand $\mathcal{L}$ to $\mathcal{L}^2$ we add second order variables f,g,h, ... , parameters u,v,w, ... , and second order quantification.

[5] [Bernays, 1967], p. 502.

The basic theory (**BT**) contains the familiar axioms for 0,', pairing, and projections, the recursion equations for all primitive recursive function(al)s, and the schema for explicit definition of functions in the form

$$(\exists f)(\forall x)\ f(x)=t_a[x]$$

or, upon changing the language a little, in the form

$$(\forall x)\ \lambda x.t(x)=t_a[x]$$

If the term t contains second order parameters, they are considered to be universally quantified in these principles of explicit definition. The theory contains also the induction schema **IA** for quantifier-free formulas ϕ of $\mathcal{L}^2$:

$$\phi 0\ \&\ (\forall x)(\phi x \rightarrow \phi x') \rightarrow (\forall x)\phi x$$

where ϕ may contain second order parameters. *Full second order arithmetic* or *classical analysis* (**CA**) extends (**BT**) by the second order induction axiom

$$(\forall f)[f(0)=1\ \&\ (\forall x)(f(x)=1 \rightarrow f(x')=1) \ \rightarrow\ (\forall x)(f(x)=1)]$$

and by the comprehension principle **CA**

$$(\exists f)(\forall x)\ [f(x)=1 \leftrightarrow \phi x]$$

where ϕ is any formula of $\mathcal{L}^2$; if ϕ contains parameters, they are taken to be universally quantified in **CA**. There are a number of other function existence principles that are important for mathematical practice and which yield – over (**BT**) – proof theoretically equivalent formalizations of classical analysis. I will just consider some choice principles:

$$\mathbf{AC_0}\text{:}\quad (\forall x)(\exists y)\phi xy \ \rightarrow\ (\exists f)(\forall x)\phi x f(x)$$

$$\mathbf{AC}\text{:}\quad (\forall x)(\exists f)\phi x f \ \rightarrow\ (\exists g)(\forall x)\phi x (g)_x$$

where $(g)_x$ is the function with $(g)_x(z) = g(<x,z>)$ for all z;

$$\mathbf{DC}\text{:}\quad (\forall g)(\exists h)\phi gh \ \rightarrow\ (\forall g)(\exists f)(\forall x)[(f)_0(x)=g(x)\ \&\ \phi(f)_x(f)_{x+1}]$$

Theorem. (**CA**)≡(**AC**$_0$)⊆(**AC**)⊆(**DC**).

Proof (of (**CA**)≡(**AC**$_0$)). Assume $(\exists f)(\forall x)(f(x)=1 \leftrightarrow \psi(x))$ and $(\forall x)(\exists y)\chi(x,y)$; show: $(\exists f)(\forall x)\chi x f(x)$. Proof: $g(<x_0,x_1>)=1 \leftrightarrow \chi(x_0,x_1) \wedge (\forall z)(z<x_0 \rightarrow \neg\chi(x_0,z))$. Then g exists according to **CA**; define $f(x)=y \leftrightarrow g(<x,y>)=1$. – Consider ψ for **CA**; then build up $\chi(x,y)$ as $(\psi(x)\wedge y=1)\vee(\neg\psi(x)\wedge y=0)$. By $\mathbf{AC}_0$ there is an f, such that $\chi(x,f(x))$ and clearly $f(x)=1\leftrightarrow\psi(x)$. **Q.E.D.**

These are interesting stability results for the axiomatic characterization of classical analysis, but there are also some most important relations to parts of set theory: Zermelo-Fraenkel set theory [with the axiom of choice but] without the powerset axiom is of the same proof theoretic strength as **(CA)** [repectively, **(AC)**].

But we are far from being able to treat full analysis for purposes of the reductive program; thus the focus has to shift to *subsystems of analysis*. They are principally distinguished by their restricted function existence principles including, for example, the comprehension principle or forms of the axiom of choice for classes of formulas, like Π^0_n, Π^0_∞, Π^1_n. There is a second important distinguishing feature that comes to the fore when the function existence principles are restricted. This feature concerns the *induction principle;* it can be formulated either as a second order axiom or as a schema (for all formulas of the language). In the former case, the principle is available only for functions that can be proved to exist in the theory. For example, $(\Pi^0_\infty$-**CA**$)$ denotes the theory obtained from **(BT)** by adding the comprehension principle for all formulas in Π^0_∞ and the full induction schema; $(\Pi^0_\infty$-**CA**$)\upharpoonright$ or "restricted-$(\Pi^0_\infty$-**CA**$)$" is the corresponding theory with the induction axiom. Clearly, $(\Pi^0_\infty$-**CA**$)\upharpoonright$ is equivalent to the theory obtained from **(BT)** by just adding the arithmetic comprehension principle. The resulting theories are of remarkably different strength: $(\Pi^0_\infty$-**CA**$)\upharpoonright$ is a conservative extension of elementary number theory **(Z)**, whereas $(\Pi^0_\infty$-**CA**$)$ proves the consistency of **(Z)**.

There is one very weak system we shall consider: It was introduced by Friedman and is labelled $(\mathbf{WKL}_0)$. An equivalent formulation is this:

$$(\mathbf{F}){:}=(\mathbf{BT} + \Sigma^0_1\text{–AC}_0 + \Sigma^0_1\text{–IA} + \text{WKL}).$$

The principle WKL is König's infinity lemma for trees of 0-1 sequences. In our framework it can be formulated as follows:

$$(\forall f)[T(f) \wedge (\forall x)(\exists y)(lh(y)=x \wedge f(y)=1) \rightarrow (\exists g)(\forall x)\, f(\overline{g}(x))=1];$$

T(f) expresses that f is (the characteristic function of) a tree of 0-1 sequences; lh is the length-function for sequences of numbers. T(f) is the purely universal formula

$$(\forall x)(\forall y)\ [(f(x{*}y)=1 \rightarrow f(x)=1) \wedge (f(x{*}{<}y{>})=1 \rightarrow y\leq 1)]$$

This theory is surprisingly strong for mathematical work, but metamathematically it is weak: **(F)** is conservative over **(PRA)** for Π^0_2-sentences. That is the reason **(F)** can be taken as the starting-point for *computational* reductions: if **(F)** proves $(\forall x)(\exists y)Rxy$, then there is a primitive recursive function f and a proof in PRA of $(\forall x)Rxf(x)$.

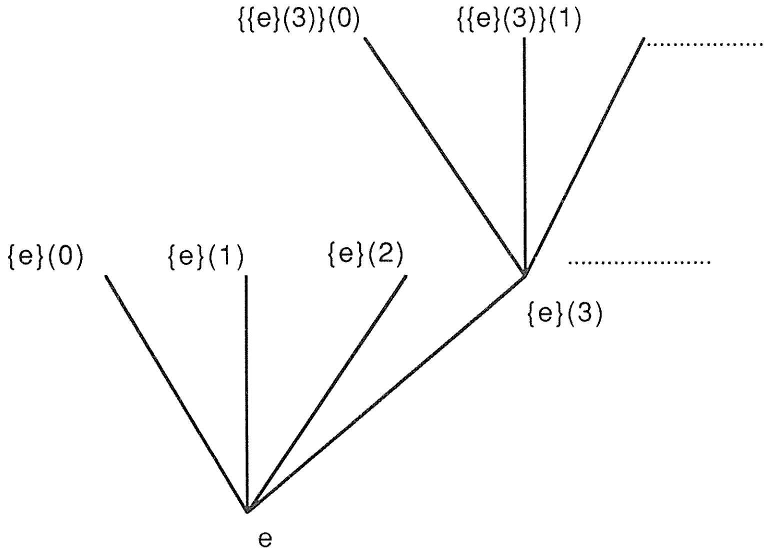
Foundational reductions. Recall that the goal is to reduce certain **P*** in which parts of mathematical practice can be developed to theories **F*** that are distinguished for philosophical, foundational reasons. Examples are the reductions of **(F)** to **(PRA)** and of $(\Pi^0_\infty$-**CA**$)\upharpoonright$ to **(HA)**. But there are foundationally satisfying reductions for much stronger theories, indeed for theories like $(\Pi^1_1$-**CA**$)\upharpoonright$ and $(\Sigma^1_2$-**AC**$)$ that are for the actual practice of analysis far too strong. The **F***'s to which they are reducible are justified from an intuitionistic point of view. Let me describe these reductive results.

Generalized inductive definitions play a central role here, both technically and conceptually. Classes given by inductive definitions, **i.d. classes** for short, have been used in constructive mathematics ever since Brouwer. Two familiar examples of such classes are well-founded trees of finite sequences of natural numbers (the "unsecured sequences" of Brouwer) and Borel sets. The former were employed in Brouwer 's justification of bar-induction; the latter in Bishop's original development of measure theory. In spite of the fact that i.d. classes can be avoided in the current practice of constructive analysis, particular ones are of intrinsic mathematical and foundational interest. The constructive (well-founded) trees form such a distinguished class, called **O**. It is given by two inductive clauses, namely:

(1) if e is 0, then e is in **O**, and

(2) if e is (the Gödel number of) a recursive function enumerating elements of **O**, then e is in **O**.

The elements of **O** are thus generated by joining recursively given sequences of previously generated elements of **O** and can be pictured as infinite, well-founded trees. Locally the structure of such a tree can be visualized as follows:



Higher tree classes are obtained by a suitable iteration of this definition along a given recursive well-ordering of the natural numbers. Suitable means here that branchings in the trees are not only taken over the natural numbers but also over already given lower tree classes. Constructive theories for **O** have been formulated as extensions of intuitionistic arithmetic with the following principles:

O. 1. $(\forall x)(A(\mathbf{O},x) \to \mathbf{O}x)$

O. 2. $(\forall x)(A(\Psi,x) \to \Psi x) \to (\forall x)(\mathbf{O}x \to \Psi x)$

where $A(\mathbf{O},x)$ is the disjunction of the antecedents of the generating clauses for **O**; it is obviously arithmetic in **O** (indeed, just Π^0_1 in **O**). $A(\Psi,x)$ is obtained from $A(\mathbf{O},x)$ by replacing all occurrences of $\mathbf{O}z$ with Ψz. **O**.1 may be called a *definition principle* making explicit that applications of the defining clauses to elements of **O** yield elements of **O**. **O**.2 is a schematic *proof principle* by induction on **O** for any formula Ψz of the language. The resulting theory is called $ID_1(\mathbf{O})$. For the higher tree classes the definition and proof principles can be formulated in a similar, though more complicated manner. The theory is denoted by $ID_{<\lambda}(\mathbf{O})$, when the iteration proceeds along arbitrary initial segments of the given well-ordering of type λ.

The theories (for higher tree classes) remain meaningful from a classical point of view, even when more general defining clauses are considered. Let P be a unary predicate variable and consider formulas $A(P,x)$ that contain P only positively[6]. Each such formula determines an i.d. class P^A that is definable in the standard way by an impredicative instance of the comprehension principle. The theories obtained from classical number theory by adding the definition and proof principles for all i.d. classes P^A given such A is denoted by ID^c_1. As above, one can consider iterations of such definitions and obtain theories $ID^c_{<\lambda}$. Feferman (1970) and Friedman (1970) established that, for example, $(\Pi^1_1\text{-}\mathbf{CA})\restriction$, $(\Delta^1_2\text{-}\mathbf{CR})$, and $(\Delta^1_2\text{-}\mathbf{CA})$ are equivalent to $ID^c_{<\omega}$, $ID^c_{<\omega^\omega}$, and $ID^c_{<\varepsilon_0}$ respectively. Thus, it is sufficient – for ıeducing these impredicative subsystems – to reduce the classical theories of inductive definitions to suitable intuitionistic theories. That was achieved (among other things) by Buchholz, Pohlers, and myself in 1977; we did this in different ways and for different intuitionistic theories, see [Buchholz e.a.]. This is the reduction I achieved.

Theorem. For any primitive recursive well-ordering of limit characteristic λ, $ID^c_{<\lambda}$ is conservative over $ID_{<\lambda}(\mathbf{O})$ for all negative arithmetic and Π^0_2-formulas.

Reductions to theories of tree classes are most satisfactory from an intuitionistic point of view. The question is whether still stronger (that is, "stronger" in the syntactic classification schema of the Π^1_n) parts of (**CA**) can be reduced in this way. Unfortunately the answer is "no". It is a well-known result of Addison and Kleene that the iteration of the hyperjump (and thus of inductive definitions) along recursive ordinals leads only to Δ^1_2-sets. Consequently, theories for i.d. classes cannot be used for reductions of $(\Pi^1_n$-**CA**) with n≥2. Here is a major conceptual problem, namely, to find a broad notion of "constructive mathematical object" and suitable principles for it that can serve as a starting point for foundational reductions of parts of analysis beyond $(\Delta^1_2\text{-}\mathbf{CA})$. There are results for $(\Delta^1_2\text{-}\mathbf{CA}+\mathbf{BI})$: Jäger and Pohlers determined the proof theoretic ordinal of the theory, and Jäger reduced it to

[6] The class of positive (in P) formulas can be given inductively - together with that of negative ones - as follows: (i) any formula of the language of arithmetic is in POS and NEG; Pt is in POS for any term t; (ii) if ϕ and φ are in POS (NEG), then their conjunction, disjunction, universal and existential quantification are in POS (NEG); (iii) if ϕ is in POS (NEG) and φ is in NEG (POS), then $(\phi \text{ -> } \varphi)$ is in NEG (POS) and $\neg\phi$ is in NEG (POS).

Feferman's constructive theory T_0 (thus establishing with earlier work of Feferman the equivalence of these theories). The system of notations used by Jäger and Pohlers was based on work by Buchholz who recast that work in a most perspicuous way in his (1986). The system of notations used by Jäger and Pohlers actually is more extensive than needed for the ordinal-theoretic analysis of the theory (Δ^1_2-**CA+BI**), but it presumably falls far short of the ordinals needed for (Π^1_2-**CA**). Significant new work is due to Rathjen (e.g., 1991) and Weiermann (1991). Good presentations of some of this work are in [Jäger 1986], [Buchholz and Schütte 1988], and [Pohlers 1989].

For me logic, and proof theory in particular, still have the fascination that arises from the combination of detailed, rigorous work with open, wide-ranging reflections. The possibility and, indeed, need for the latter is sometimes hidden, alluded to in brief remarks, delegated to Postscripta, or (sup-) pressed into footnotes. From the discussion of the foundational aims of proof theory it should be quite clear that mathematical reductive results have to be complemented by analyses of the philosophical distinctiveness of the constructive theory to which a classical one has been reduced. That is very much in the open, but there is also the more subtle (and pervasive) assumption, namely, that we are dealing with **formal theories**! The focus on formal theories, i.e., theories whose axioms and rules are somehow effectively presentable, is required so that our considerations satisfy epistemological, normative demands. How these demands were "transformed" into precise mathematical definitions will be the main concern of the next lecture.

2. Effectiveness and provability. This lecture is concerned with the analysis of *effective calculability* in the thirties and is roughly divided into five sections. I shall discuss the logical **decision problem** first and describe briefly connections to issues of decidability in mathematics. Then I'll analyze under the heading **step-by-step to absoluteness** the connection between work of Church & Kleene, Gödel, and Hilbert & Bernays. I will argue – against accepted wisdom – that their work focused on *one* central informal notion, namely, "computability in a symbolic calculus", and that in each case a serious stumbling-block to a convincing analysis emerged; a stumbling-block that was overcome only by Turing. Turing's solution is discussed in the third section entitled **determinacy & finiteness**. The fourth section focuses on **Gödel's** concept of a **general recursive function** that is related to Herbrand's proposal of generalizing the concept of a primitive recursive function. Finally, an elaboration of the difference between Gödel's and Herbrand's proposals will lead to the notion of **provably total function**. And that forms a natural stepping-stone to the next part of my lectures, in which I address the question of how to extract computational information from formal proofs.

Decidability. In some respects, the issues I alluded to go back to Greek mathematics and philosophy; they concern, on the one hand, the axiomatic presentation of geometry (Euclid) and, on the other hand, the formalization of logical reasoning (Aristotle). But it was only Frege who provided, with his *Begriffsschrift*, a sufficiently expressive formal language and a sufficiently strong logical calculus that allowed the realization of the earlier intentions with respect to mathematics. Frege required that (i) all assumptions be explicitly formulated in the formal language, and that (ii) each step in a proof be taken in accord with one of the antecedently specified rules of the logical calculus. He considered the second requirement as his way of sharpening the axiomatic method he explicitly traced back to Euclid. With this sharpening Frege pursued the aim of recognizing the "epistemological nature" of theorems. In the introduction to *Grundgesetze der Arithmetik* he wrote:

> By insisting that the chains of inference do not have any gaps we succeed in bringing to light every axiom, assumption, hypothesis or whatever else you want to call it on which a proof rests; in this way we obtain a basis for judging the epistemological nature of the theorem.

But that can be done, Frege realized, only if inferences do not require contentual knowledge: their applications have to be recognizable as correct on account of the syntactic form of the sentences occurring in them. Indeed, Frege claimed that in his logical system "inference is conducted like a calculation" and continued:

> I do not mean this in a narrow sense, as if it were subject to an algorithm the same as ... ordinary addition and multiplication, but only in the sense that there is an algorithm at all, i.e., a totality of rules which governs the transition from one sentence or from two sentences to a new one in such a way that nothing happens except in conformity with these rules.[7]

Almost fifty years later, in 1933, Gödel referred back to Frege and Peano when he formulated "the outstanding feature of the rules of inference" in a formal mathematical system. The rules, Gödel said, "refer only to the outward structure of the formulas, not to their meaning, so that they can be applied by someone who knew nothing about mathematics, or by a machine."[8] Frege did not consider the possibility of mechanically drawing inferences to be among the *logically* significant achievements of his *Begriffsschrift*. But Hilbert grasped the potential of this aspect, radicalized it, and exploited it in his formulation and pursuit of the consistency problem. In doing so he believed to have found the basis for mediating between Kronecker's foundational position and the ever more strongly set theoretic practice of mathematics: The restrictive demands of Kronecker were accepted for metamathematics; set theory was to be formulated in a strictly formal way; and within that formal framework mathematics could be freely developed – assuming satisfaction of the minimal requirement, i.e., consistency. It is in this way that I understand Bernays' remark quoted earlier, "... it became his goal, one might say, to do battle with Kronecker with his own weapons of finiteness by means of a modified conception of mathematics." And over the years the strict formalization of mathematics seemed to open up also new ways of solving mathematical problems (through calculation). In Hilbert and Ackermann's book this is called the "rechnerische Behandlung von Problemen", i.e., the calculatory treatment of problems!

The most famous problem among these was the so-called *Entscheidungsproblem* or decision problem. It is closely related to the consistency problem and was pursued by some (e.g., Herbrand) on account of this connection. Its classical formulation in terms of validity and satisfiability is found in Hilbert and Ackermann's book:

[7] [Frege 1984], p. 237. But he was careful to emphasize (in other writings) that all of thinking "can never be carried out by a machine or be replaced by a purely mechanical activity" [Frege 1969], p. 39.

[8] [Gödel 1933], p. 1.

> The Entscheidungsproblem is solved if one knows a procedure that permits the decision concerning the validity, respectively, satisfiability of a given logical expression by a finite number of operations.[9]

Hilbert and Ackermann emphasized the fundamental importance ("grundsätzliche Wichtigkeit") of a solution to the decision problem. Researchers in the Hilbert school realized full well that a positive solution for predicate logic – together with the assumption of the finite axiomatizability of theories and the quasi-empirical completeness of *Principia Mathematica*[10] – would allow the decision concerning the provability (truth) of any mathematical statement. For some that was sufficient reason to expect a negative solution; von Neumann, for example, expressed his views as follows.

> .. it appears that there is no way of finding the general criterion for deciding whether or not a well-formed formula *a* is provable. (We cannot at the moment establish this. Indeed, we have no clue as to how such a proof of undecidability would go.) ... the undecidability is even the *conditio sine qua non* for the contemporary practice of mathematics, using as it does heuristic methods, to make any sense. The very day on which the undecidability does not obtain any more mathematics as we now understand it would cease to exist; it would be replaced by an absolutely mechanical prescription ("eine absolut mechanische Vorschrift"), by means of which anyone could decide the provability or unprovability of any given sentence.
> Thus we have to take the position: it is generally undecidable, whether a given well-formed formula is provable or not.[11]

When claiming that we have no clue as to how a proof of undecidability would go, von Neumann pointed to *the* conceptual problem. After all, there were well-known proofs for the unsolvability of mathematical problems. But note, all such impossibility results were given relative to a determinate class of admissible means, e.g., doubling the cube by using only ruler and compass. And exactly here lies the problem: A negative solution to the Entscheidungsproblem required a mathematically precise answer to the question "What are *absolut mechanische Vorschriften*?" According to the conventional view, we were given an answer to this question by the work of Church, Turing, and others (e.g., Gödel, Kleene, Post, Hilbert, Bernays): there *is* a precise mathematical description of mechanical procedures. Furthermore, Church and Turing proved that there are no recursive (Turing-machine computable) functions providing a positive solution to the decision

[9] [Hilbert and Ackermann], pp. 72 - 73.

[10] That was already explicit in [Löwenheim], see [van Heijenoort], p. 246. Cp. also [Herbrand 1930a], p. 207, where Herbrand speaks of an "experimental certainty" that *Principia Mathematica* allows the representation of all mathematical statements and arguments.

[11] [von Neumann 1927], pp. 11-12.

problem. These results seemed to confirm von Neumann's hunch that heuristic methods will continue to be needed in mathematics; that is, proofs have to be given, new principles have to be recognized, important new notions have to be introduced! That need had already been made most plausible, though not proved, by Gödel's Incompleteness Theorems; after all, they were formulated in Gödel's 1931 paper only for particular theories. A convincing analysis of effective computability was thus required in order to give a negative solution to the decision problem and to come to a proper understanding of the generality of the incompleteness theorems. The question for us is: What are the grounds for accepting the various (equivalent) notions as actually constituting a precise mathematical description of mechanical procedures?

Step-by-step to absoluteness. In his 1934 Lectures at Princeton Gödel strove to make the incompleteness results less dependent on particular formalisms[12], but he did not succeed in resolving the conceptual issue of giving a general notion of "formal theory". He viewed the primitive recursive definability of formulas and proofs as a "precise condition which *in practice* suffices" to describe particular formal systems, but he was clearly looking for a condition that would suffice *in principle*. But in what direction could one search? – Gödel considered it as an "important property" that, for any argument, the value of a primitive recursive function can be computed by a "finite procedure" and he added in footnote 3:

> The converse seems to be true if, besides recursions according to the scheme (2) [of primitive recursion], recursions of other forms ... are admitted. This cannot be proved, since the notion of finite computation is not defined, but it can serve as a heuristic principle.

In the last section of the Lecture Notes Gödel described "general recursive functions" (to be discussed in greater detail below); they are obtained as unique solutions of certain functional equations, and their values must be computable in an "equational calculus". For Gödel, the crucial point of his proposal was the specification of *mechanical* rules for the computation of function values. Though the footnote I just quoted may seem to express a form of Church's Thesis, Gödel emphasized in a 1965 letter to Martin Davis that no formulation of Church's Thesis was intended. He wrote:

[12] The theory Gödel considered is actually second order arithmetic!

> The conjecture stated there only refers to the equivalence of "finite (computation) procedure" and "recursive procedure". However, I was, at the time of these lectures, not at all convinced that my concept of recursion comprises all possible recursions;

At the time, Gödel was equally unconvinced by Church's proposal to identify effective calculability with λ-definability. In conversation with Church in early 1934, he called that proposal "thoroughly unsatisfactory".[13] Nevertheless, Church announced his "thesis" in a talk he contributed to the meeting of the American Mathematical Society on April 19, 1935; but he formulated it in terms of recursiveness, not λ-definability. In the subsequent famous 1936 paper *An unsolvable problem of elementary number theory* Church wrote:

> The purpose of the present paper is to propose a definition of effective calculability which is thought to correspond satisfactorily to the somewhat vague intuitive notion in terms of which problems of this class are often stated, and to show, by means of an example, that not every problem of this class is solvable.

Church proposed again to identify effective calculability with recursiveness. The fact that λ-definability was known to be an equivalent concept simply added for Church "... to the strength of the reasons adduced below for believing that they [these precise concepts] constitute as general a characterization of this notion [i.e., effective calculability] as is consistent with the usual intuitive understanding of it." To give a deeper analysis Church pointed out, in section 7 of his paper, that two methods suggest themselves to characterize effective calculability of number theoretic functions. The first of these methods uses the notion of algorithm, and the second employs the notion of *calculability in a logic*. He argued that they do not lead to definitions more general than recursiveness. Let me indicate briefly the argument pertaining to the second method. Church considered a logic **L**, i.e., a system of symbolic logic whose language contains the equality symbol =, a symbol { }( ) for the application of a unary function symbol to its argument, and numerals for the positive integers. For unary functions F he defined:

> F is *effectively calculable* if and only if there is an expression f in the logic **L** such that: {f}(μ)=ν is a theorem of **L** iff F(m)=n; here, μ and ν are expressions that stand for the positive integers m and n.

Church claimed that such F are recursive, *assuming* that **L** satisfies certain conditions; these conditions amount to the recursive enumerability of **L**'s theorem predicate, and the claim follows by an unbounded search. The

[13] Church in a letter to Kleene, dated November 29, 1935, and quoted in [Davis 1982], p. 9.

crucial condition in Church's list requires the steps in derivations of equations to be, well, recursive! Here we hit on a serious stumbling-block for Church's analysis, since an appeal to the thesis when arguing for it is logically circular. And yet, Church's argument achieves something: The general concept of calculability is explicated as derivability in a symbolic logic, and the step-condition is used to sharpen the idea that we operate by effective rules in such a formalism. I suggest the claim that the steps of any effective procedure must be recursive be called **Church's Central Thesis.** Robin Gandy aptly called Church's argument for his thesis the *"step-by-step argument"*: If steps in computations are recursive, then the functions being calculated are recursive. The mathematical essence of these observations is captured by appropriate versions of Kleene's normal form theorem.

The concept of "calculability in a logic" used in Church's argument is an extremely natural and fruitful one. Of course, it is directly related to "Entscheidungsdefinitheit" for relations and classes introduced by Gödel in his 1931paper and to "representability" as used in his Princeton lectures. It was used in other contemporary analyses: Gödel defined that very notion in his 1936 note *On the length of proofs* and emphasized its "type-absoluteness". In his contribution to the Princeton Bicentennial Conference (1946) Gödel reemphasized absoluteness (in a more general sense) and took it as the main reason for the special importance of recursiveness. Here we have, according to Gödel, the first interesting epistemological notion whose definition is not dependent on the chosen formalism. But the *stumbling-block* Church had to face shows up here, too; after all, absoluteness is achieved only relative to the description of *formal* systems.

The more general definition of absoluteness Gödel gave in 1946 is actually derived from work of Hilbert and Bernays in Supplement 2 of the second volume of *Grundlagen der Mathematik.* They called a number-theoretic function "regelrecht auswertbar" if it is computable in some "deductive formalism" and they formulated three "Rekursivitätsbedingungen" for deductive formalisms. Then they showed: (i) a function that is computable in a deductive formalism satisfying their "recursiveness" conditions can be computed in a very restricted number theoretic formalism, and (ii) the functions computable in the latter formalism are exactly the recursive functions.

Hilbert and Bernays' analysis is in my view a natural and satisfactory capping of the development from Entscheidungsdefinitheit to an "absolute" notion of computability. But their analysis does not overcome the major stumbling-block; rather, it puts the stumbling-block in plain view through the recursiveness conditions that deductive formalisms must satisfy. The crucial condition requires the proof predicate for such formalisms to be primitive recursive! Now I want to show you, how Turing got around the fundamental difficulty.

Determinacy & finiteness. Turing's classical paper *On computable numbers* opens with a description of what is ostensibly its subject, namely, "computable numbers" or "real numbers whose expressions as a decimal are calculable by finite means". Turing is quick to point out that the fundamental problem of explicating "calculable by finite means" is the same when considering, e.g., computable functions of an integral variable. Thus it suffices to address the question: *What does it mean for a real number to be calculable by finite means?* In §9 he argues that the operations of his machines "include all those which are used in the computation of a number". But he does not try to establish the claim directly; he rather attempts to answer "the real question at issue", i.e., *What are the possible processes which can be carried out* (by a human computor) *in computing a number?*

Turing imagines a mechanical computor writing symbols on paper that is divided into squares "like a child's arithmetic book". As the two-dimensional character of this computing space is taken not to be essential, Turing takes a one-dimensional tape divided into squares as the basic computing space and formulates one important restriction. That restriction is motivated by definite limits of our sensory apparatus to distinguish - at one glance - between symbolic configurations of sufficient complexity. It states that *only finitely many distinct symbols can be written on a square.* Turing suggests as a reason that "If we were to allow an infinity of symbols, then there would be symbols differing to an arbitrarily small extent" and we would not be able to distinguish at one glance between them. A second (and related) way of arguing the point uses a finite number of symbols and strings of such symbols: for example, Arabic numerals like 17 or 9999999 are distinguishable at one glance; however, it is not possible for us to determine at one glance whether 98899954967899998769 is identical with 98899954967899998769 or whether they are different.

Now let us turn to the question: *What determines the steps of the computor, and what kind of elementary operations can he carry out?* The behavior is *uniquely* determined at any moment by two factors: (i) the symbols or symbolic configuration he observes, and (ii) his "state of mind" or his "internal state". This uniqueness requirement may be called the **determinacy condition (D)**; it guarantees that computations are deterministic. Internal states are introduced to have the computor's behavior depend possibly on earlier observations, i.e., to reflect his experience. Since Turing wants to isolate operations of the computor that are "so elementary that it is not easy to imagine them further divided", it is crucial that symbolic configurations relevant for fixing the circumstances for the actions of a computor are *immediately recognizable.* So we are led to postulate that a computor has to satisfy two **finiteness conditions:**

(F.1) *there is a fixed finite number of symbolic configurations a computor can immediately recognize;*
(F.2) *there is a fixed finite number of states of mind that need be taken into account.*

For a given computor there are consequently only finitely many different relevant combinations of symbolic configurations and internal states. Since the computor's behavior is – according to (D) – uniquely determined by such combinations and associated operations, the computor can carry out at most finitely many different operations. These operations are restricted as follows:

(O.1) *only elements of observed symbolic configurations can be changed;*
(O.2) *the distribution of observed squares can be changed, but each of the new observed squares must be within a bounded distance L of an immediately previously observed square.*

Turing emphasizes that "the new observed squares must be immediately recognisable by the computer", and that means that the distributions of the new observed squares arising from changes according to (O.2) must be among the finitely many ones of (F.1). Clearly, the same must hold for the symbolic configurations resulting from changes according to (O.1). Since some of the operations may involve a change of state of mind, Turing concludes:

> The most general single operation must therefore be taken to be one of the following: (A) A possible change (a) of symbol [as in (O.1)] together with a possible change of state of mind. (B) A possible change (b) of observed squares [as in (O.2)] together with a possible change of state of mind.

With this restrictive analysis of the possible steps of a computor, the proposition that his computations can be carried out by a Turing machine is established rather easily. Indeed, Turing first "constructs" machines that mimic the work of computors directly and then observes:

> The machines just described do not differ very essentially from computing machines as defined in § 2, and corresponding to any machine of this type a computing machine can be constructed to compute the same sequence, that is to say the sequence computed by the computer [in my terminology: comput*or*].

Thus we have **Turing's Theorem:** Any number theoretic function F that can be computed by a computor, satisfying the determinacy condition (D) and the conditions (F) and (O), can be computed by a Turing machine.

Turing's analysis and his theorem can be generalized by making an observation concerning the determinacy condition: (D) is not needed to guarantee the Turing computability of F in the theorem. Computors that do not satisfy (D) can be mimicked by non-deterministic Turing machines and thus, exploiting the reducibility of non-deterministic to deterministic machines, by deterministic Turing machines. And that allows us to connect Turing's considerations with those of Church we discussed earlier. Consider, for that purpose, an effectively calculable function F and a (non-deterministic) computor who calculates the value of F in a logic **L**. Using the generalized form of Turing's Theorem and the fact that Turing computable functions are recursive, F is recursive. This argument for F's recursiveness does no longer appeal to Church's Thesis; rather, such an appeal is replaced by the assumption that the calculation in the logic is done by a computor satisfying the conditions (F) and (O). If that assumption is to be discharged, then a substantive thesis is needed again. And it is this thesis I want to call **Turing's Central Thesis.** It expresses the fact that a mechanical computor indeed satisfies the finiteness conditions (F), and that the elementary operations he can carry out are restricted as conditions (O) require.

Church wrote in his review of Turing's paper when comparing Turing computability, recursiveness, and λ-definability: "Of these, the first has the advantage of making the identification with effectiveness in the ordinary (not explicitly defined) sense evident immediately ..." For Gödel, Turing's work provided "a precise and unquestionably adequate definition of the general concept of formal system". In the historical and systematic context Turing found himself, he asked exactly the right question: *What are the possible processes a human computor can carry out in computing a number?* The general problematic *required* an analysis of the idealized capabilities of a mechanical computor. Let me emphasize that the separation between conceptual analysis (leading to the axiomatic conditions) and rigorous proof (establishing Turing's Theorem) is essential for clarifying on what the correctness of his general thesis rests; namely, on recognizing that the axiomatic conditions are true for computors who proceed mechanically. We have to remember that quite clearly when moving to methodological discussions in artificial intelligence and cognitive science. Even Gödel got it wrong, when he claimed that Turing's argument in his 1936 paper was intended to show that "mental processes cannot go beyond mechanical procedures".

Gödel's recursive functions. Another proposal Gödel got thoroughly wrong was Herbrand's! Recall that in the last section of his Princeton Lecture Notes Gödel addressed the question *What other recursions beyond primitive ones might be admitted in defining functions whose values can still be determined by a finite computation?* This is discussed under the heading "general recursive functions", and Gödel gave a definition of a general notion of recursive function that (he thought) had been suggested to him by Herbrand in a private communication, as we know now, of April 7, 1931:

> If ϕ denotes an unknown function, and $\psi_1, \ldots, \psi_k$ are known functions, and if the ψ's and ϕ are substituted in one another in the most general fashions and certain pairs of resulting expressions are equated, then, if the resulting set of functional equations has one and only one solution for ϕ, ϕ is a recursive function.[14]

Gödel went on to make two restrictions on this definition and required, first of all, that the left-hand sides of the functional equations be in a standard form with ϕ being the outermost symbol and, secondly, that "for each set of natural numbers $k_1, \ldots, k_l$ there shall be exactly one and only one m such that $\phi(k_1, \ldots, k_l)=m$ is a derived equation". The rules that were allowed in giving derivations are of a very simple character: Variables in any derived equation can be replaced by numerals, and if the equation $\phi(k_1, \ldots, k_l)=m$ has been obtained, then occurrences of $\phi(k_1, \ldots, k_l)$ on the right-hand side of a derived equation can be replaced by m. So much about this proposal; it was taken up for a systematic development in [Kleene 1936].

[14] [Gödel I], p. 368.

What was important about Gödel's modifications? For Gödel himself the crucial point was the precise specification of *mechanical* rules for deriving equations or, to put it differently, for carrying out computations. That point of view was also expressed by Kleene who wrote with respect to the definition of "general recursive function of natural numbers":

> It consists in specifying the form of the equations and the nature of the steps admissible in the computation of the values, and in requiring that for each given set of arguments the computation yield a unique number as value.[15]

In a letter to van Heijenoort, dated 14 August 1964, Gödel asserted that "it was exactly by specifying the rules of computation that a mathematically workable and fruitful concept was obtained".[16] When making this claim Gödel took for granted what he had expressed in an earlier letter to van Heijenoort, namely, that Herbrand's suggestion had been "formulated *exactly* as on page 26 of my lecture notes, i.e. without reference to computability."[17] But Gödel had been unable to find Herbrand's letter among his papers and had to rely on his recollection which, he said, "is very distinct and was still very fresh in 1934". However, the letter from Herbrand was found by John W. Dawson in Gödel's Nachlass, reads like a preliminary version of parts of [Herbrand 1931c], and on the evidence of that letter it is clear that Gödel misremembered. Herbrand as a matter of fact wrote – describing a system of arithmetic and the introduction of recursively defined functions *into that system* with intuitionistic, i.e., finitist, justification –

> In arithmetic we have other functions as well, for example functions defined by recursion, which I will define by means of the following axioms. Let us assume that we want to define all the functions $\phi_n(x_1, x_2, \ldots, x_{pn})$ of a certain finite or infinite set F. Each $\phi_n(x_1, \ldots)$ will have certain defining axioms; I will call these axioms (3F). These axioms will satisfy the following conditions:
>
> (i) The defining axioms for ϕ_n contain, besides ϕ_n, only functions of lesser index.
>
> (ii) These axioms contain only constants and free variables.
>
> (iii) We must be able to show, by means of intuitionistic proofs, that with these axioms it is possible to compute the value of the functions univocally for each specified system of values of their arguments.

It is most plausible that Herbrand admitted, in addition to the (intuitionistically interpreted) axioms, substitution rules of the sort

[15] [Kleene 1936], p. 727.

[16] [van Heijenoort 1985], p. 115.

[17] In a letter to van Heijenoort of 23 April 1963, excerpted in the introductory note to [Herbrand 1931c], see [Herbrand 1971], p. 283. (Gödel refers to his 1934 lectures.) The background for and the content of the Herbrand-Gödel correspondence is described in [Dawson 1991].

formulated by Gödel as rules of computation. Indeed, he asserted in his paper [1931c] – as he had done in his letter to Gödel – that all intuitionistic computations can be carried out, e.g., in the formal system **P** of *Principia Mathematica*. This is not to suggest that Gödel was wrong in his assessment, but rather to point to the most important step he had taken, namely, *to disassociate recursive functions from an epistemologically restricted notion of proof*. Later on, Gödel even dropped the regularity condition that guaranteed the totality of calculable functions. He emphasized then[18] "that the precise notion of mechanical procedures is brought out clearly by Turing machines producing partial rather than general recursive functions." However, at *this earlier* historical juncture, the explicit introduction of an equational calculus with purely formal, mechanical rules for computing was important for the mathematical development of recursion theory and also for the conceptual analysis. After all, it brought out clearly what, according to Gödel, Herbrand had failed to see, namely, "that the computation (for all computable functions) proceeds by exactly the same rules."[19]

Herbrand's provably total functions. I want to make some additional remarks on Herbrand's proposal(s) and to analyze in particular the restrictive conditions he imposed. A careful description and thoughtful interpretation of the proposal(s) can be found in [van Heijenoort 1985]. It should be noted, however, that this paper was written before Dawson's discovery of the Gödel-Herbrand correspondence. Van Heijenoort had thus to rely on Gödel's reports concerning not only the details of Herbrand's suggestion to him, but also its very framing as an attempt to give a general characterization of effective calculability.

In any event, van Heijenoort distinguished three different occasions in 1931 on which Herbrand "proposed ... to introduce a class of computable functions that would be more general than that of primitive recursive functions". The first proposal is found in Herbrand's [1931a] on page 273, where Herbrand described the restricted means allowed in metamathematical arguments and required, in particular, that "all the functions introduced must be actually calculable for all values of their arguments by means of

[18] [Wang 1974], p. 84. The very notion of partial recursive function, of course, had been introduced in [Kleene 1938].

[19] in [vanH 1985], page 115

operations described wholly beforehand." The second proposal is the one reported in Gödel's Princeton Lectures (without making reference to computability), and the third suggestion was made in Herbrand's [1931c] on pages 290 and 291. It is formulated as follows, again in the context of a system for arithmetic:

> We can also introduce any number of functions $f_i(x_1, x_2, ..., x_{ni})$ together with hypotheses such that
> (a) The hypotheses contain no apparent variables;
> (b) Considered intuitionistically, they make the actual computation of the $f_i(x_1, x_2, ..., x_{pn})$ possible for every given set of numbers, and it is possible to prove intuitionistically that we obtain a well-determined result.

With van Heijenoort I assume that, here too, Herbrand used "intuitionistic" as synonymous with "finitist".[20] This third proposal is identical with the one made by Herbrand in his letter to Gödel quoted above except for clause (i) from the earlier definition; but that clause is implicitly assumed, as is clear from the examples Herbrand discusses. I view the first formulation on the one hand as a preliminary, not fully elaborated version of the second and third formulation; on the other hand, I view it as a more explicit indication of the Kroneckerian element in metamathematics I pointed to earlier on. Thus, we can see the evolution of essentially one formulation!

This is (prima facie) not in conflict with the interpretations Gödel considered[21], e.g., that Herbrand envisioned "unformalized and perhaps unformalizable computation methods" and refused "to confine himself to formal rules of computation"; but, as we will see, it is in conflict with Gödel's understanding that Herbrand's proposal leads to a class of functions larger than that of general recursive functions. So let us distinguish two features of Herbrand's schema, namely, (1) the defining axioms (plus suitable rules) must make the actual intuitionistic computation of function values possible, and (2) the termination of computations has to be provable intuitionistically. That is, in modern terminology, we are dealing with "intuitionistically provably total (or provably recursive) functions", where provability is not a formal notion. However, a connection to a formal notion of provability is given in the fourth section of [1931c], where Gödel's Incompleteness Theorems for the system **P** of *Principia Mathematica* is discussed. Herbrand

[20] A more detailed description of intuitionistic arguments is given in note 5 of Herbrand's [1931c], pp. 288-289.

[21] in [vanH 1985], pp.115-117.

asserts there that any intuitionistic computation can be carried out in **P** and that any intuitionistic argument can be formalized in **P**. He concludes, after sketching Gödel's proof, that **P**'s consistency is not provable by arguments formalizable in **P**, hence not intuitionistically either. What is most interesting is his remark that Gödel's argument does not apply to the system of arithmetic that includes the above schema for introducing functions: The functions that are introducible cannot be described intuitionistically, as we could diagonalize to obtain additional functions. This last observation can be turned around so as to show that the class of provably total functions of a formal theory cannot be enumerated by an element of that class.

What then is the extension of Herbrand's class of functions? According to the discussion reported above, it includes the primitive recursive functions and is included in the class of provably recursive functions of **P**. Indeed, at the end of [1931c] Herbrand asserts that ordinary analysis (I assume that Herbrand means by that full second-order arithmetic) can take the place of **P** in the above claims concerning the formalizability of intuitionistic computations and arguments. Indeed, he conjectures that full first-order arithmetic with recursion equations for only addition and multiplication might already be sufficient. If the latter conjecture were true, Herbrand's class would be included in the class of provably recursive functions of Peano arithmetic. Basic in this discussion is Herbrand's conviction that the system of arithmetic described in his [1931c] (possibly even without the infinitary rule D) allows one to carry out all intuitionistic proofs. The paper [1931c] was dated Göttingen, July 14, 1931; in the letter to Gödel of April 7, 1931, and sent from Berlin, the claim concerning intuitionistic proofs is explicitly stated for the much weaker system with quantifier-free induction only. As a matter of fact, Herbrand claims there also that "... each proof in this arithmetic, which has no bound variables, is intuitionistic - this fact rests on the definition of our functions and can be seen directly." If that were true, Herbrand's class would consist of exactly the primitive recursive functions. In conclusion, it seems that Gödel was right – for stronger reasons than he put forward – when he cautioned that Herbrand had *foreshadowed*, but not *introduced*, the notion of general recursive function.[22]

[22] In a letter to van Heijenoort of August 14, 1964; see [vanH 1985], pp. 115-116.

PART B. PROVABLY TOTAL FUNCTIONS.

In the first part of these lectures I described three main themes of proof theoretic research and their intimate historical and systematic connection with the analysis of effective computability. As to the latter, two distinct approaches emerged. One is connected with Gödel and began with his definition of the class of *general recursive functions* via a suitable equational calculus. The other, pursued by Herbrand, also requires that effectively computable functions be defined as solutions of functional equations, but in addition, their *totality* has to be proved finitistically. It is this notion of *provably total function* that will be prominent in the two lectures of this part. However, we are not using informal finitist proofs, but rather proofs in particular formal theories for proving the totality of simply defined functions. In the fifties, Kreisel asked the question: Given a formal theory $\mathfrak{T}$, can we find a natural class $\mathfrak{F}$ of recursive functions, such that the $\mathfrak{T}$-provably total functions are exactly the elements of $\mathfrak{F}$? During the last few years the question has been turned around for small classes of recursive functions (complexity classes): Given a class $\mathfrak{F}$ of recursive functions, can we find a natural theory $\mathfrak{T}$, such that the elements of $\mathfrak{F}$ are exactly the $\mathfrak{T}$-provably total functions? The hope has been that relationships between formal theories might reveal relationships between the corresponding classes of functions.

1. Sequent calculi and normal derivations. A variety of technical tools have been employed in proof theory; for example the ε-calculus, the no-counter-example interpretation, the *Dialectica* interpretation. However, the tools most directly useful and most perspicuous in my view are finitary and infinitary *sequent calculi* for which normalization theorems can be established. The reductive results I mentioned in A.1 have been proved by use of such calculi and an associated lucid *method* that is also due to Gentzen. This will be illustrated now by considering the first consistency result that was (properly) obtained in the Hilbert school; its strongest version is due to Herbrand. Then I will discuss the cut elimination theorem and some of its extensions in detail.

A consistency proof. The classical sequent calculi we are considering are presented in the style of Tait (1968); i.e., finite *sets* of formulas are proved and *negation* is directly available only for atomic formulas. Thus, the basic logical symbols are $\wedge$, $\vee$, $\exists$, $\forall$. The rules of the calculi are included among the following ones, where Γ is used as a syntactic variable ranging over finite sets of formulas and Γ, ϕ stands for the union of Γ and singleton ϕ:

$$\text{LA:} \quad \Gamma, \varphi, \neg\varphi, \qquad \varphi \text{ atomic}$$

$$\underline{\wedge}\text{:} \quad \frac{\Gamma,\varphi_0 \qquad \Gamma,\varphi_1}{\Gamma,\varphi_0\wedge\varphi_1}$$

$$\underline{\vee}_{\underline{i}}\text{:} \quad \frac{\Gamma,\varphi_i}{\Gamma, \varphi_0\vee\varphi_1}, \qquad i = 0,1$$

$$\underline{C}\text{:} \quad \frac{\Gamma,\varphi \qquad \Gamma,\neg\varphi}{\Gamma}$$

When appropriate, rules for quantifiers will be available. For example, when considering (extensions of) number theory we have the rules:

$$\underline{\forall}\text{:} \quad \frac{\Gamma,\varphi a}{\Gamma, (\forall x)\varphi x} \qquad a \notin P(\Gamma)$$

$$\underline{\exists}\text{:} \quad \frac{\Gamma,\varphi t}{\Gamma, (\exists x)\varphi x}$$

Here $a \in P(\Gamma)$ means that the parameter a occurs in one of the formulas in Γ. The rules for function quantifiers are analogous. The axioms for identity will be discussed later; we certainly should have as derived rules:

$$\frac{\Gamma,\varphi t}{s\neq t ,\Gamma, \varphi s} \qquad\qquad \frac{\Gamma,\varphi t}{t\neq s ,\Gamma, \varphi s}$$

and as a theorem Γ, a=a. Derivations are built up in tree form as usual; let me use D,E,... as syntactic variables ranging over derivations. Gentzen's *Hauptsatz* is the fundamental fact concerning these finitary calculi: Every derivation can be transformed into a normal one, i.e., a derivation in which the cut-rule is not applied. By inspecting the rules one see immediately that all formulas occurring in a normal derivation of Δ are subformulas of elements of Δ.

Let me explain, through an example, how the subformula property (and $\underline{\forall}$–inversion) can be exploited in the "canonical" proof of the reflection principle; the idea is simple, pervasive, and elegant. Consider a fragment of arithmetic, say (N); it has the usual axioms for zero and successor, defining

equations for finitely many primitive recursive functions, and the induction schema for quantifier-free formulas. Consequently, all of the axioms can be taken to be in quantifier-free form. Now assume that (N) proves a Π^0_1-statement and, thus, by $\underline{\forall}$–inversion a quantifier-free statement ψ. A normal derivation of Δ,ψ can be obtained, where Δ contains only negations of (N)-axioms. These considerations can actually be carried out in (**PRA**), i.e.,

$$(\mathbf{PRA}) \vdash \mathrm{Pf}_N\,(\overline{\psi}) \rightarrow \mathrm{Pf}^n(\overline{\Delta,\psi}).$$

Pf_N and Pf^n express that there is a derivation in (N) and, respectively, that there is a normal derivation in the sequent calculus. A normal derivation of Δ,ψ contains only subformulas of elements in its endsequent. So one can use an adequate, quantifier-free truth definition Tr (for quantifier-free formulas of bounded complexity) to show that

$$(\mathbf{PRA}) \vdash \mathrm{Pf}^n(\overline{\Delta,\psi}) \rightarrow \mathrm{Tr}(\overline{\Sigma(\Delta,\psi)}),$$

where $\Sigma(\Delta,\psi)$ is the disjunction of the formulas in Δ,ψ. This is possible because the language of (N) contains only finitely many symbols for primitive recursive functions; we can easily define a primitive recursive valuation function for all terms built up from them. More generally, but for the same reason, (N) could contain all functions of a fixed segment of the Grzegorczyk hierarchy.[23] As Tr is provably adequate we have

$$(\mathbf{PRA}) \vdash \mathrm{Tr}(\overline{\Sigma(\Delta,\psi)}) \rightarrow \Sigma(\Delta,\psi),$$

and, thus,

$$(\mathbf{PRA}) \vdash \mathrm{Pf}_N\,(\overline{\psi}) \rightarrow \psi.$$

This last step can be taken, since the axioms of (N) *are* axioms of (**PRA**). The subformula property is obviously the crucial feature of normal derivations; this feature makes it possible to use truth definitions for formulas of restricted syntactic complexity and to establish *formally* the truth of theorems. I tried to explain this feature through a particular example; now let me proceed directly to the general and detailed proof theoretic work.

[23] For details concerning the standard material for truth definitions, see [Schwichtenberg 1977], pp. 893-894.

Sequent calculus. The form of the classical sequent calculi we are considering was presented above in a very rough way. To establish Gentzen's Hauptsatz we need a more precise metamathematical presentation of the calculus. To that end we start out by defining the notions of *principal, minor,* and *side formula* for inferences and specifying the general form of those inferences that were introduced earlier. (My presentation follows [Schwichtenberg].)

Definitions. (i) The *principal formula(s)*, p.f. , of the axioms are ϕ and $\neg\phi$; the p.f. of the inferences are the inferred formulas with the new connective; $\underline{C}$ does not have a p.f.;
(ii) the *minor formula(s)*, m.f., of $\underline{\wedge}$ in the premise Γ,φ_i is φ_i; of $\underline{\vee}_i$ $[\underline{\forall}, \underline{\exists}]$ it is φ_i, $[\varphi a, \varphi t]$; of $\underline{C}$ in the premise Γ,φ it is φ and in the premise $\Gamma,\neg\varphi$ it is $\neg\varphi$;
(iii) the *side formulas*, s.f., of the inferences are the elements of Γ.

Thus all our inferences are of the form

$$\frac{\Gamma,\Delta_i}{\Gamma,\Delta} \qquad \text{for all } i<k,\ 0\leq k\leq 2$$

where Δ consists of the p.f. of the inference; in case of $\underline{C}$, $\Delta=\emptyset$. The Δ_i contain the m.f. of the i-th premise. Consider such an inference and assume that derivations D_i of its premises Γ,Δ_i are given. Then $D = < (D_i)_{i<k}, (\Delta_i)_{i<k}, \Gamma, \Delta>$ is a derivation of Γ, Δ; the latter sequent is the *conclusion* of the inference. The D_i are called *direct subderivations* of D.

Definition. (i) The *length* $|\phi|$ of a formula ϕ is defined inductively by
$|\phi| = |\neg\phi| = 0$ if ϕ is atomic; $|\phi\wedge\psi| = |\phi\vee\psi| = \sup(|\phi|, |\psi|)+1$;
$|(\forall x)\phi x| = |(\exists x)\phi x| = |\phi a|+1$ (Note that $|\neg\phi| = |\phi|$);

(ii) the *length* $|D|$ of a derivation D is defined inductively to be the sup $(|D_i|+1)$ with D_i as the direct subderivations of D;

(iii) the *cut-rank* $\rho(D)$ of a derivation D is also defined inductively: If D_i are the direct subderivations of D then $\rho(D)$ equals either

$\sup(|\varphi|+1, \sup_{i<k}\rho(D_i))$ if the last rule in D is $\underline{C}$ with cut-formula φ
or $\sup_{i<k} \rho(D_i)$
(iv) a derivation is called *normal* or *cut-free* only when $\rho(D)=0$; if $\rho(D)=1$ it is called *quasi-normal*. (The cut-formulas in quasi-normal derivations are all atomic.)

Now I formulate some lemmata that are easily established by induction on derivations. For the formulation of the first we need the operation $D\Rightarrow D,\Gamma$ that adds Γ to the side formulas of all the inferences; for the formulation of the second, we need the operation $D(a)\Rightarrow D(s)$ that replaces all occurrences of a by s. Clearly, one wants to replace only occurrences of a that are "connected" to occurrences of a in an element of the endsequent and, in addition, one has to insure that the side condition on the universal quantifier rule is not violated: to do this we assume, without loss of generality, that with each such inference there is associated a unique eigenvariable and that these eigenvariables are distinct from parameters occurring in Γ and s, respectively.

Weakening lemma. If D is a derivation of Δ, then D,Γ is a derivation of Δ,Γ; $|D,\Gamma| = |D|$ and $\rho(D,\Gamma) = \rho(D)$.

This lemma allows us, most importantly, to consider a more general formulation of the cut rule, namely,

$$\frac{\Gamma_0,\varphi \qquad \Gamma_1,\neg\varphi}{\Gamma_0, \Gamma_1}$$

Substitution lemma. If D(a) is a derivation of $\Delta(a)$, then D(s) is a derivation of $\Delta(s)$; $|D(a)|=|D(s)|$; $\rho(D(a))=\rho(D(s))$.

$\underline{\wedge}$-Inversion. If D is a derivation of $\Delta,\psi_0\wedge\psi_1$, then there are derivations D_i of Δ,ψ_i, $i=0,1$; $|D_i|\leq|D|$ and $\rho(D_i)\leq\rho(D)$.

$\underline{\forall}$-Inversion. If D is a derivation of $\Delta,(\forall x)\psi(x)$ then there is a derivation D_0 of $\Delta,\psi(c)$; $|D_0|\leq|D|$ and $\rho(D_0)\leq\rho(D)$.(c is new for D).

Eliminating cuts. The proof of the *fact that every derivation can be transformed into a normal one* proceeds by induction on the length* of D[24] and within it by induction on the cut-rank $\rho(D)$; cf. Gentzen's proof or its presentation in Kleene's "Introduction to Metamathematics". But instead of carrying out this argument with inner and outer induction, we separate matters into three steps and formulate three distinct propositions: (1) the reduction lemma, (2) the cut elimination theorem, and (3) the normalization theorem. The reduction lemma is the essential fact for the proof of cut-elimination.

[24] Length* is defined like length, but in the case of $\underline{C}$ one adds the length of the immediate subderivations.

Reduction lemma. Let D_0 and D_1 be derivations of Γ_0,φ and $\Gamma_1,\neg\varphi$ respectively; both derivations have cut-rank $\rho(D_i)\leq|\varphi|$. Then there is a derivation D of Γ_0,Γ_1 with $|D|\leq|D_0|+|D_1|$ and $\rho(D)\leq|\varphi|$.

Proof (by induction on $|D_0|+|D_1|$). The lemma is symmetric w.r.t. D_0 and D_1, as $\neg\neg\varphi\equiv\varphi$ and $|\neg\varphi|=|\varphi|$.

Case 1. Either φ or $\neg\varphi$ is *not* the principal formula of the last inference in D_0, D_1 respectively. Assume the former; then the last inference of D_0 is of the form

$$\frac{\Lambda,\varphi,\Delta_i}{\Lambda,\varphi,\Delta} \qquad \text{for all } i<k$$

with Δ_i containing the m.f. of the inference, Δ the p.f., and Λ,φ the s.f. Clearly, $\Gamma_0=\Lambda,\Delta$. By induction hypothesis the sequences

$$\Lambda,\Delta_i,\Gamma_1$$

are provable for all i<k with derivations of length $<|D_0|+|D_1|$ and cut-rank $\leq|\varphi|$. The conclusion is obtained, as $\Gamma_0=\Lambda,\Delta$, by the inference

$$\frac{\Lambda,\Delta_i,\Gamma_1}{\Lambda,\Delta,\Gamma_1} \qquad \text{for all } i<k$$

Case 2. φ and $\neg\varphi$ are the p.f. of the last inference in D_0, respectively D_1.

Case 2.1. φ or $\neg\varphi$ is atomic. Then the last and only inferences in D_0 and D_1 must be instances of (logical) axioms; consequently, Γ_0,Γ_1 is also an instance of an axiom.

Case 2.2. φ or $\neg\varphi$ is a disjunction $\psi_0\vee\psi_1$. By symmetry we can assume the former. So $\neg\varphi\equiv\neg\psi_0\wedge\neg\psi_1$. We can also assume that φ is a s.f. of the last inference in D_0, replacing D_0 by D_0,φ if necessary. So the last inference is of the form

$$\frac{\Gamma_0,\varphi,\psi_i}{\Gamma_0,\varphi}$$

By induction hypothesis we have a derivation D_0' of

$$\Gamma_0,\psi_i,\Gamma_1$$

of length $<|D_0|+|D_1|$ and cut-rank $\leq|\varphi|$. By $\underline{\wedge}$-inversion we obtain from D_1 a derivation D_1' of

$$\Gamma_1,\neg\psi_i$$

of length $\leq|D_1|<|D_0|+|D_1|$ and cut-rank $\leq|\varphi|$. Joining D_0' and D_1' by $\underline{C}$ with cut-formula ψ_i we obtain a derivation of Γ_0,Γ_1; its length is $\leq|D_0|+|D_1|$ and cut-rank $\leq|\varphi|$.

Case 2.3. φ or $\neg\varphi$ is an existential statement $(\exists y)\psi y$. By symmetry we can assume the former. So $\neg\varphi\equiv(\forall y)\neg\psi y$. We assume again that φ is a s.f. of the last inference of D_0. So the last inference is of the form

$$\frac{\Gamma_0,\varphi,\psi t}{\Gamma_0,\varphi}$$

By induction hypothesis we have a derivation D_0' of

$$\Gamma_0,\psi t,\Gamma_1$$

of length $<|D_0|+|D_1|$ and cut-rank $\leq|\varphi|$. By $\underline{\forall}$-inversion and substitution we obtain from D_1 a derivation D_1' of

$$\Gamma_1,\neg\psi t$$

of length $\leq|D_1|<|D_0|+|D_1|$ and cut-rank $\leq|\varphi|$. Joining D_0' and D_1' by $\underline{C}$ with cut-formula ψt we obtain a derivation of Γ_0,Γ_1; its length is $\leq|D_0|+|D_1|$ and its cut-rank $\leq|\varphi|$. **Q.E.D.**

The crucial point, for sure, is that the cut-rank of the derivation D for Γ_0,Γ_1 is bounded by $|\varphi|$; a derivation for Γ_0,Γ_1 with cut-rank bounded by $|\varphi|+1$ is trivially obtained by joining the given derivations D_0 and D_1 by a cut with cut-formula φ.

Cut elimination theorem. If D is a derivation of Γ with $\rho(D)>0$, then we can find a derivation D' of Γ with $\rho(D')<\rho(D)$ and $|D'|\leq 2^{|D|}$.

Proof (by induction on $|D|$). The claim follows by induction hypothesis in all cases except when the last inference in D is $\underline{C}$ with cut-formula φ and $\rho(D)=|\varphi|+1$. In this case we have direct subderivations D_0 of Γ,φ and D_1 of $\Gamma,\neg\varphi$. By induction hypothesis there are derivations D_0' and D_1' of Γ,φ and $\Gamma,\neg\varphi$ such that $D_i'\leq 2^{|D|}$ and $\rho(D_i')<\rho(D_i)\leq|\varphi|+1$. The hypotheses of the reduction lemma are satisfied, and we can obtain a derivation of Γ with length $\leq 2^{|D_0|}+2^{|D_1|}\leq 2^{\sup(|D_0|,|D_1|)+1}\leq 2^{|D|}$ and cut-rank $\leq|\varphi|<|\varphi|+1=\rho(D)$. **Q.E.D.**

Corollary (normalization theorem). If D is a derivation of Γ, then we can find a normal derivation D' of Γ of length $\leq 2^{|D|}_{\rho(D)}$.

The corollary is obtained from the cut-elimination theorem by induction on $\rho(D)$. The function 2^n_m is defined by primitive recursion as follows: $2^n_0=n$ and

$2^n_{m+1}=2^{2^n_m}$. – By not eliminating all cuts, for example not those with atomic cut-formulas, one can obtain partial normalization results; this will be pursued in a variety of ways below. But first we have to draw the crucial consequence from the normalization theorem. It is concerned with the *bounding* of the logical complexity of formulas appearing in a (normal) proof of a sequent Γ: Every formula in D is a subformula of an element in Γ.

Definition. ϕ is a subformula of ψ iff [(ϕ is ψ) or (ψ is $\neg\xi$, ξ is atomic and ϕ is ξ) or (ψ is $\xi_0\wedge\xi_1$ or $\xi_0\vee\xi_1$ and ϕ is a subformula of ξ_0 or ξ_1) or (ψ is $(\forall x)\xi x$ or $(\exists x)\xi x$ and ϕ is ξt or a subformula of ξt for any term t)].

Corollary (subformula property). If D is a normal derivation of Γ, then every formula in D is a subformula of some element in Γ.

Proof (by induction on normal derivations). One just has to notice that all the rules occurring in normal derivations have the property: any formula in its premise(s) is a subformula of a formula in its conclusion. **Q.E.D.**

Remark: There is a different way of proving a normal form theorem for the sequent calculus! The completeness proof for the calculus without the cut-rule shows that to establish all logical truths the cut-rule is not needed; that is, if a sequent can be proved at all, it is (by the soundness of the full calculus) a logical truth, and thus it can be established by a normal proof.

Extensions. The considerations for pure predicate logic can be modified and extended to treat *finitary calculi* with additional, mathematical axioms, additional sorts (e.g., finite type theory), or additional rules (e.g., induction rule), but also to treat *infinitary calculi*. I will consider only finitary calculi. The *first extension* – to treat theories with universal axioms – admits new axioms in addition to the logical ones. The particular way I treat them is modeled after [Girard, 1987], pp.123-126. We start with a definition: Let $\mathcal{T}$ be a set of sequents whose elements are literals (i.e., either atoms or negated atoms); if $\mathcal{T}$ is closed under substitution[25] it is called a *Post System*. Let me describe some examples:

(1) the axioms for equality can be expressed by a Post System :

(EA_1) Γ, t=t

(EA_2) Γ, s≠t, $\neg\phi$s, ϕt for arbitrary terms t,s but only atomic ϕ.

[25] "Closed under substitution" means: if D(a)∈ T then D(t)∈ T for each term t in the language at hand.

(2) the axioms for *elementary arithmetic* can be formulated by a Post System including (EA_1), (EA_2) and

(BA_1) Γ, 0≠s'

(BA_2) Γ, s'≠t', s=t

(BA_3) Γ, s+0=s

Γ, s+t'=(s+t)'

(BA_4) Γ, s.0=0

Γ, s.t'=(s.t)+s

Sometimes it is convenient to have < as a basic symbols with the axioms:

(BA_5) Γ, ¬s<0

Γ, s<t' ↔ (s<t ∨ s=t)

(BA_6) Γ, s≤t ↔ (s<t ∨ s=t)

Clearly, the second sequent of (BA_5) and the one of (BA_6) have to be canonically rewritten. Let me indicate this for the former:

Γ, ¬s<t', s<t, s=t

Γ, ¬s<t, s<t'

Γ, s≠t, s<t'.

(3) The axioms for additional functions in $\mathcal{F}\subseteq\mathcal{PR}$ can be expressed using the Post System that consists of all instances of the defining equations for the elements of $\mathcal{F}$. The resulting theories are (**PRA**) for $\mathcal{F}=\mathcal{PR}$ and (**KEA**) for $\mathcal{F}=\mathcal{E}_3$, the class of Kalmar-elementary functions. In general, I denote the extension of $\mathcal{T}$ by the defining axioms for the functions in $\mathcal{F}$ by $\mathcal{T}(\mathcal{F})$.

If we add to the logical axioms a Post System $\mathcal{T}(\mathcal{F})$ by admitting its sequents as axioms, then we can readily obtain a generalization of the Normalization Theorem – if we require that $\mathcal{T}(\mathcal{F})$ is also closed under cut. As I think such systems are artificial, this closure won't be required. We will consider instead of normal derivations *quasi-normal* ones. Such derivations allow atomic cuts, in particular with elements from $\mathcal{T}(\mathcal{F})$ as cut-formulas. The standard terminology can readily be extended to explicate the notion of $\mathcal{T}(\mathcal{F})$-derivation and thus of $\mathcal{T}(\mathcal{F})\vdash$D. (Clearly, the principal formulas of axioms Γ,Δ are the elements of Δ.)

Theorem ($\mathcal{T}$-normalization). Let D be a $\mathcal{T}(\mathcal{F})$-derivation of Γ; then there is a quasi-normal $\mathcal{T}(\mathcal{F})$-derivation E of Γ with $|E| \leq 2_m^{|D|}$ and $m=\rho(D)-1$.

Proof. One proceeds as in the proof of the normalization theorem above. It is only the proof of the reduction lemma that has to be modified slightly: in case 2.1. one has to consider the possibility that $\mathcal{T}(\mathcal{F})$-axioms are involved. If one of the axioms is a logical one then Γ_0, Γ_1 must be a $\mathcal{T}(\mathcal{F})$-axiom; if both are $\mathcal{T}(\mathcal{F})$-axioms, then we can infer Γ_0, Γ_1 by a permitted cut. **Q.E.D.**

Here one could require having only atomic cuts whose cut-formulas are p.f.s in some sequent of $\mathcal{T}(\mathcal{F})$. In applications this is unnecessarily restrictive, since only the complexity of formulas is crucial: We do not obtain the full subformula property, but the important bounding of the logical complexity of formulas occurring in the derivation is still achieved.

Corollary. If D is a quasi-normal $\mathcal{T}(\mathcal{F})$-derivation of Γ, then every formula in D is either a subformula of some element in Γ or of some $\mathcal{T}(\mathcal{F})$-axiom.

Remark. Cut-elimination does *not* hold in general for systems with proper axioms. To see that, consider the following example adapted from [Girard 1987][26]: assume that both A and $\neg A$, B are (proper) axioms. Clearly, B is provable from them by one application of the cut-rule, but there is no cut-free derivation.

Now we shall treat a *second extension* – this time not by mathematical axioms of a restricted form, but rather by a rule for induction, called Θ-IA; it is of the form:

$$\frac{\Gamma,\varphi 0 \qquad \Gamma,\neg\varphi a,\ \varphi a'}{\Gamma,\ \varphi t}\ .$$

Here the parameter a is not in $P(\Gamma \cup \{\varphi t\})$, t is any term, and φa is in Θ, a class of formulas like Δ_0, Σ_n^0, Π_n^0. The theory obtained from an extension of elementary arithmetic $\mathcal{T}(\mathcal{F})$ by adding this induction rule is denoted by ($\Theta(\mathcal{F})$-IA). We distinguish now between O-cuts and I-cuts; the latter are those cuts at least one of whose cut-formulas has been inferred by the induction rule. The O-cut-rank, say: $\rho_O(D)$, is the $\sup\{|\psi|+1 \mid \psi$ is the cut-formula of an O-cut$\}$; the I-cut-rank $\rho_I(D)$ is $\sup\{|\psi|+1 \mid \psi$ is the cut-formula of an I-cut$\}$. We call a derivation D *I-normal* iff $\rho_O(D)=1$; in other words, D is I-normal iff all its cuts are either I-cuts or have atomic cut-formulas. Again, the argument for the normalization theorem is readily adapted to allow the transformation of derivations into I-normal ones.

[26] section 2.7.7 on pag.125

Theorem (I-normalization). If D is a $\Theta(\mathcal{F})$-IA-derivation of Γ, then we can find an I-normal $\Theta(\mathcal{F})$-IA-derivation E of Γ; $|E| \leq 2_m^{|D|}$ and $m=\rho_O(D)-1$.

Proof. One proceeds as in the proof of the normalization theorem above. We have only to modify the proof of the reduction lemma. In case 2 we consider now only O-cuts, i.e., the situation when neither of the last inferences in D_i is taken with the induction rule; the argument proceeds as above. We add, however, a third case covering the possibility that at least one of the last inferences is Θ-IA. Then the claim follows immediately from the assumption taking D as E: it satisfies the condition on the cut-rank trivially. **Q.E.D.**

Even in this case we have a significant *bounding* property for I-normal $\Theta(\mathcal{F})$-IA-derivations:

Corollary. If D is an I-normal derivation of Γ using Θ-IA, then every formula occurring in D is either a literal or a subformula of an element in $\Gamma \cup \Theta \cup \neg\Theta$.

I draw one final consequence that will be important for our intended applications. It is a simple instance of Herbrand's theorem and will be generalized significantly in the next lecture.

$\exists$-inversion. Let Δ contain only existential formulas and let ϕa be quantifier-free; if D is a quasi-normal $\mathcal{T}$-derivation of $\Delta,(\exists x)\phi x$, then there is a finite sequence of terms $t_0,\ldots,t_n$ and a quasi-normal $\mathcal{T}$-derivation E of $\Delta,\varphi t_0,\ldots,\varphi t_n$; $|E| \leq |D|$.

This is proved straightforwardly by induction on the length of D. Note that in theories that allow definition by cases the finite sequence of terms can be joined into a single term t by defining:

$$t = t_0 \text{ if } \varphi t_0; \quad = t_1 \text{ if } \neg\varphi t_0 \wedge \varphi t_1; \quad \ldots \quad = t_n \text{ if } \neg\Sigma_{i<n}(\varphi t_i) \wedge \varphi t_n$$

This kind of "term extraction" will be crucial for obtaining computational information from derivations.

2. Herbrand analyses. In this lecture I intend, first, to present techniques for the extraction of computational information and, second, to prove paradigmatically some results relating fragments of arithmetic and weak subsystems of analysis to classes of recursive functions. To get at the computational content of number-theoretic statements of the form $(\forall x)(\exists y)\psi xy$ I will use derivations in sequent calculi. The normal form theorem guarantees bounds on the logical complexity of formulas occurring in derivations; the invertibility of the (rules for the) quantifiers $\forall$ and, with suitable restrictions, $\exists$ yields a functional analysis of the combination $\forall\exists$. After all, statements of the form $(\exists y)\psi y$ do express a functional dependence of the quantified variable y on the parameters occurring in ψ. In case ψ's matrix is quantifier-free, the uniformity of formal proofs D for such ψ provides the basis for Herbrand analyses, i.e., for the extraction of a term t from D and the generation of an associated proof D* of ψt. The extracted term t reflects both the expressiveness of the term language and the formal structure of the given derivation. Furthermore, if the basic terms in D are computable, t represents also a computation of a restricted sort.

Bounding existential quantifiers. The functional analysis of Π^0_2-theorems will be based on suitable forms of Herbrand's Theorem for Σ^0_1-statements; its basic form is this:

Herbrand's Theorem. Let $\Gamma=\{\phi_0, \ldots ,\phi_n\}$ contain only purely existential formulas; if D is a derivation of Γ, then there is a quasi-normal derivation of $\Delta_0, \ldots ,\Delta_n$; $\Delta_j=\{\ \phi_{i,j} : i\leq n_j$ and $\phi_{\iota,j}$ is an instance of the matrix of ϕ_j }, $j\leq n$. The terms occurring in these instances are built up from terms occurring in D.

A most useful corollary can be established (directly by induction on quasi-normal derivations).

Corollary. Let Γ contain only purely existential formulas and let ϕ be quantifier-free; if D is a derivation of

$$\Gamma,\ (\exists x_0) \ldots (\exists x_n)\phi,$$

then there are sequences of terms $t_{0,i}, \ldots ,t_{n,i}$, $i\leq p$, and a quasi-normal derivation of

$$\Gamma,\ \phi(t_{0,1}, \ldots ,t_{n,1}), \ldots ,\phi(t_{0,p}, \ldots ,t_{n,p}).$$

The corollary can be further extended to I-normal derivations; that extension will be given only in a more specialized setting. We are considering theories $\mathcal{T}(\mathcal{F})$ of the form (QF($\mathcal{F}$)-IA) such that $\mathcal{T}(\mathcal{F})$ and $\mathcal{F}$ satisfy the following two conditions:

(H.1) $\mathcal{F}$ is provably closed under explicit definitions and definition by cases (thus under Boolean operations, max, min);

(H.2) $\mathcal{F}$ is provably closed under bounded search, i.e., for any formula ϕ in QF($\mathcal{F}$) there is an h in $\mathcal{F}$ such that $\mathcal{T}(\mathcal{F})$ proves: $(\exists y\leq x)\phi y <\!\!-\!\!> \phi h(x)$.

Theories $\mathcal{T}(\mathcal{F})$ satisfying these two conditions are called **Herbrand Theories.** It is for them that I establish the most suitable form of $\underline{\exists}$-inversion.

$\underline{\exists}$-inversion. Let $\mathcal{T}(\mathcal{F})$ be an Herbrand theory, let Γ contain only purely existential formulas, and let ψ be quantifier-free; if D is a $\mathcal{T}(\mathcal{F})$-derivation of $\Gamma,(\exists x)\psi x$, then there is a term t* and a(n I-normal) $\mathcal{T}(\mathcal{F})$-derivation D* of $\Gamma,\psi t^*$.
Proof (by induction on I-normal $\mathcal{T}(\mathcal{F})$-derivations). I focus on the central step in the argument when the last inference in D is of the form

$$\frac{\Gamma,\phi 0,(\exists x)\psi x \qquad\qquad \Gamma,\neg\phi a,\phi a',(\exists x)\psi x}{\Gamma,\phi t,(\exists x)\psi x}$$

The induction hypothesis, applied to the derivations D_0 and D_a leading to the premises of the inference, yields terms r and s(a). These terms may contain other parameters as well. The induction hypothesis yields also derivations D_0* and D_a* of

(1) $\qquad\Gamma,\phi 0,\psi r$

and of

(2) $\qquad\Gamma,\neg\phi a,\phi a',\psi s(a)$.

$\mathcal{T}(\mathcal{F})$ proves clearly

$$\neg\phi 0,\phi t,(\exists x\leq t)(\psi x \wedge \neg\psi x')$$

and, with condition H.2 and $\wedge$-inversion, both

(3) $\qquad\neg\phi 0,\phi t,\psi h(t)$

and

(4) $\qquad\neg\phi 0,\phi t,\neg\psi h(t)'$.

From (2), replacing the parameter a by the term h(t), one obtains

(5) $\Gamma,\neg\phi h(t),\phi h(t)',\psi s(h(t))$.

Cutting (5) successively with (4), (3), and (1) yields a derivation of

(6) $\Gamma,\phi t,\psi r,\psi s(h(t))$.

Using condition H.1, definition by cases, we can define a function f in $\mathcal{F}$, such that $\mathcal{T}(\mathcal{F})$ proves $\Gamma,\phi t,\psi(f(t))$. Q.E.D.

For Herbrand theories I can give now an absolutely straightforward answer to the question concerning Skolem-functions via the Term Extraction Lemma. Its proof is immediate by $\underline{\forall}$-inversion and subsequent $\underline{\exists}$-inversion.

Term Extraction. Let $\mathcal{T}(\mathcal{F})$ be an Herbrand theory and let ψ be quantifier-free; if $\mathcal{T}(\mathcal{F})$ proves $(\forall x)(\exists y)\psi xy$, then there is a term t(a) in $\mathcal{L}(\mathcal{F})$, such that $\mathcal{T}(\mathcal{F})$ proves $(\forall x)\psi xt(x)$. $\lambda x.t(x)$ denotes a function in $\mathcal{F}$.

Two remarks are in order. First, that results are insensitive to extensions of the theories by purely universal sentences. Thus the corollary I am going to formulate now holds not only for the theories explicitly mentioned, but also for any of their Π^0_1-extensions. Second, all the considerations can be carried out for standard formulations of open theories, when the induction principle for quantifier-free formulas is given by an open axiom schema.

Corollary. (i) The provably total functions of $I\Delta_0$+exp are exactly the Kalmar-elementary functions. (ii) The provably total functions of (QF($\mathcal{PR}$)-IA) are exactly the primitive recursive ones.

For (i) it has to be observed that the Kalmar-elementary functions can be introduced in a definitional extension of $I\Delta_0$+exp; the proof theoretic analysis is then given for this definitional extension. – A fact similar to (i) can be established, as a matter of fact, for all ($\Delta_0(\mathcal{E}_n)$-IA), $3\leq n$, thus giving a proof-theoretic characterization of all classes $\mathcal{E}_n$ in the Grzegorczyk-hierarchy with index greater than two.

These considerations will be expanded in three different directions: First, I'll show how Σ^0_1-Induction can be eliminated (and that result allows us to obtain quite systematically and easily results concerning fragments of arithmetic); second, I will show how these techniques are useful for investigations of extremely weak fragments of arithmetic, e.g., for bounded arithmetic as introduced by Buss; third, I'll investigate second-order extensions of fragments of arithmetic, in particular Friedman's (**F**).

Induction and recursion. The key-word here is *match-up*, that is, match-up between induction and recursion. I will show that the schema of primitive recursion is exactly right for analyzing the Σ^0_1-induction-principle, and that bounded iteration is exactly right for analyzing s-Σ^b_1-induction. As consequences we obtain very neat proofs of two facts: (1) the provably total functions of (Σ^0_1-IA) coincide with the primitive recursive ones (established by Parsons and independently by Mints and Takeuti), and (2) the provably total functions of (Buss's theory) S^1_2 are exactly the polynomial-time computable ones. Let me start out with the considerations for the former result.

Lemma. Let Γ contain only Σ^0_1-formulas; if D is an I-normal derivation of Γ in ($\Sigma^0_1(\mathcal{PR})$-IA), then there is an I-normal derivation of Γ in (QF($\mathcal{PR}$)-IA).

Proof. The argument proceeds by induction on the number # of applications of theΣ^0_1-induction rule in D. Clearly, if #=0, the claim is trivial. So assume that #>0 and consider an application of the Σ^0_1-induction rule such that no other application occurs above it in D. The subderivation E determined in this way ends with the inference

$$\frac{\Delta,(\exists x)\psi x0 \qquad\qquad \Delta,\neg(\exists x)\psi xa,(\exists x)\psi xa'}{\Delta,(\exists x)\psi xt}$$

where ψ is quantifier-free. Without loss of generality we can assume that Δ contains only existential statements; by the corollary to the I-normalization Theorem, all formulas in D are contained in Π^0_1 or Σ^0_1; if Δ contained universal formulas, we could use $\underline{\forall}$-inversion first and carry out the subsequent steps with additional parameters – and these paramaters could be removed in the very last step by applying first the rule for $\exists$ and then for $\forall$. After this digression, showing once more the significance of bounding the logical complexity in "normal" derivations, let me continue with the main argument. Let E_0 be the derivation of the left premise and E_a that of the right premise. $\underline{\exists}$-inversion applied to E_0 yields a term $\sigma[0]$ and a derivation in (QF($\mathcal{PR}$)-IA) of

(1) $\Delta,\psi\sigma[0]0$.

The application of $\underline{\forall}$-inversion and then of $\underline{\exists}$-inversion to E_a yields (for a new parameter c) a term $\tau[a,c]$ and a derivation of

(2) $\quad \Delta,\neg\psi ca,\psi\tau[a,c]a'$.

Now we define a function f by primitive recursion

$$f(0) \quad = \sigma[0]$$
$$f(a') \quad = \tau[a,f(a)]\,;$$

one can verify directly, using (1), (2), and quantifier-free induction that there is a (QF($\mathcal{PR}$)-IA)-derivation of

$$\Delta,\psi f(a)a$$

and thus of

$$\Delta,(\exists x)\psi xt.$$

If this derivation is used to replace E in D, the induction hypothesis on # can be employed to infer the claim of the lemma. **Q.E.D.**

How can we use this fact to establish Parsons's Theorem? (Σ^0_1($\mathcal{PR}$)-IA) is a definitional extension of (Σ^0_1-IA); the lemma tells us that the former theory is conservative for Σ^0_1-formulas, and indeed for Π^0_2-formulas, over (QF($\mathcal{PR}$)-IA). But we already saw earlier, as a direct consequence of the Term Extraction Lemma, that $\mathcal{PR}$ is the class of provably recursive functions of (QF($\mathcal{PR}$)-IA).

Theorem. The provably recursive functions of (Σ^0_1-IA) are exactly the primitive recursive functions.

The schema of these considerations can be used to prove Buss's theorem that the provably total functions of the theory S^1_2 of bounded arithmetic are exactly the polynomial time computable functions. Indeed, using suitable (Skolem-) operator theories [Sieg 1991], p. 421, re-obtains the characterization of all classes in the polynomial hierarchy; in [Buchholz and Sieg], analogous arguments are used to show that $\mathcal{P}$ is the class of provably total functions of a certain theory of binary trees introduced by Ferreira. I will give a perspicuous argument for what Buss considered to be the difficult part of the theorem, namely, that the provably total functions of S^1_2 are contained in $\mathcal{P}$; notably, witnessing functions will not be used. $\mathcal{L}(\mathfrak{B})$, the language of bounded arithmetic, is the language of elementary arithmetic expanded by function symbols $|.|$, $\overset{\to}{:}$, and #, where $|a|$ yields the length of the binary representation of a, $\overset{\to}{:}$ is the shift-right-function, and a#b is $2^{|a|\,|b|}$. The language $\mathcal{L}(\mathcal{P})$ is obtained from $\mathcal{L}(\mathfrak{B})$ by adding function symbols for each element of $\mathcal{P}$. The latter class of functions is defined inductively as the smallest class of functions that contains certain initial functions (0, ', $\overset{\to}{:}$, 2. , χ, *choice*[27]) and that is closed under composition and bounded iteration; a function f is said to be defined *by iteration from* g *and* h *with time bound* p *and space bound* q (p and q suitable polynomials[28]) iff the following holds: If τ is defined by

$$\tau(x,0) = g(x)$$
$$\tau(x,y') = h(x,y,\tau(x,y)),$$

then we must have

$$(\forall y{\le}p(|x|))\ |\tau(x,y)| \le q(|x|)$$

and

$$f(x) = \tau(x,p(|x|))\,;$$

x indicates a sequence of variables. – Letting $\mathcal{F}$ stand for $\mathcal{P}$ or $\mathfrak{B}$, the set of quantifier-free formulas in $\mathcal{L}(\mathcal{F})$ is denoted by QF($\mathcal{F}$). The bounded quantifiers $(\forall x{\le}|t|)$ and $(\exists x{\le}|t|)$, understood again as abbreviations, are called *sharply bounded*. $\Delta^b_0(\mathcal{F})$, the class of sharply bounded formulas, is built up from literals in $\mathcal{L}(\mathcal{F})$ using $\wedge$, $\vee$, and sharply bounded quantifiers; if closure under bounded existential quantification is also required, the set of formulas is called $\Sigma^b_1(\mathcal{F})$. A formula of $\mathcal{L}(\mathcal{F})$ is in s-$\Sigma^b_1(\mathcal{F})$ just in case it is of the form $(\exists x{\le}t)\phi$, where ϕ is in QF($\mathcal{F}$). The theories of bounded arithmetic to be investigated contain the basic axioms for the non-logical symbols of $\mathcal{L}(\mathfrak{B})$, the defining equations for the elements of $\mathcal{P}$ in case the theory is formulated in $\mathcal{L}(\mathcal{P})$, and one of the induction principles Φ-PIND or Φ-LIND. The latter are formulated as rules

$$\frac{\Gamma,\ \varphi 0 \qquad\qquad \Gamma,\ \neg\varphi\vec{a},\ \varphi a}{\Gamma,\ \varphi t}$$

and

$$\frac{\Gamma,\ \varphi 0 \qquad\qquad \Gamma,\ \neg\varphi a,\ \varphi a'}{\Gamma,\ \varphi|t|}\quad ;$$

[27] 2. , χ, and *choice* are the shift-left-function, the characteristic function of $\le$, and the definition by cases function, respectively.

[28] A polynomial is called *suitable* if it has only nonnegative integers as coefficients; thus suitable polynomials are monotonically increasing.

where φ is in Φ (and the parameter a must not occur in the lower sequent). The resulting theories are denoted by (Φ-PIND) and (Φ-LIND); the theory ($\Sigma_1^b(\mathcal{B})$-PIND) is S_2^1 and allows - via a delicate boot-strapping - the introduction of all elements of $\mathcal{P}$: ($\Sigma_1^b(\mathcal{P})$-PIND) is a definitional extension of ($\Sigma_1^b(\mathcal{B})$-PIND) and by Theorem 13a in [Buss, p.52] equivalent to ($\Sigma_1^b(\mathcal{P})$-LIND). Let me formulate some facts whose proofs require care, but are standard and will not be given.

Lemma. (i) $\mathcal{P}$ is provably in (QF($\mathcal{P}$)-LIND) closed under explicit definitions and definition by cases.
(ii) $\mathcal{P}$ is provably in (QF($\mathcal{P}$)-LIND) closed under strictly bounded search, i.e., for any ϕ in QF($\mathcal{P}$) there is an h in $\mathcal{P}$, such that (QF($\mathcal{P}$)-LIND) proves: $(\exists y \le |x|)\phi y \leftrightarrow \phi h(|x|)$.

The last part of the lemma asserts that in (QF($\mathcal{P}$)-LIND) every formula in $\Delta_0^b(\mathcal{P})$ is provably equivalent to a quantifier-free formula. By inspecting the proof of Theorem 14 in [Buss, p.53] one can see that QF($\mathcal{P}$)-replacement is provable in (s-$\Sigma_1^b(\mathcal{P})$-LIND); that fact allows us to show that in (s-$\Sigma_1^b(\mathcal{P})$-LIND) every $\Sigma_1^b(\mathcal{P})$-formula is equivalent to one in s-$\Sigma_1^b(\mathcal{P})$. Thus we have:

Lemma. (i) ($\Delta_0^b(\mathcal{P})$-LIND) is equivalent to (QF($\mathcal{P}$)-LIND).
(ii) (s-$\Sigma_1^b(\mathcal{P})$-LIND) is equivalent to ($\Sigma_1^b(\mathcal{P})$-LIND).

This completes the preparation for the central considerations involving the extraction of terms. (QF($\mathcal{P}$)-LIND) is an Herbrand Theory, slightly modified to adjust for strict boundedness. The considerations for the $\underline{\exists}$-Inversion Lemma can be carried through for this theory, and the (modified) Term Extraction Lemma shows then that the provably total functions are exactly the elements of $\mathcal{P}$. As (s-$\Sigma_1^b(\mathcal{P})$-LIND) is equivalent to S_2^1, it is sufficient for a proof of Buss's theorem to show the following:

Theorem. The provably total functions of (s-$\Sigma_1^b(\mathcal{P})$-LIND) are exactly the polynomial time computable functions.

We only have to establish that the theory (s-$\Sigma_1^b(\mathcal{P})$-LIND) is conservative over (QF($\mathcal{P}$)-LIND) for Π_2^0-formulas. That is obtained directly from the next lemma.

Lemma. Let Γ contain only Σ_1^0-formulas; if **D** is an I-normal derivation of Γ in (s-$\Sigma_1^b(\mathcal{P})$-LIND), then there is an I-normal derivation of Γ in (QF($\mathcal{P}$)-LIND).

Proof. The argument proceeds by induction on the number # of applications of the s-$\Sigma_1^b(\mathcal{P})$-induction rule in **D**. The claim is trivial if #=0. So assume that #>0 and consider an application of the s-$\Sigma_1^b(\mathcal{P})$-induction rule, such that no further application occurs above it. The subderivation **E** determined in this way ends with the inference

$$\frac{\Delta,\psi\mathbf{a}0 \qquad\qquad \Delta, \neg\psi\mathbf{a}a, \psi\mathbf{a}a'}{\Delta,\psi\mathbf{a}|s|}\ ;$$

$\psi\mathbf{a}a$ is of the form $(\exists x)(x\le t[\mathbf{a},a] \wedge \psi^*x\mathbf{a}a)$, where ψ^* is in QF($\mathcal{P}$) and **a** indicates the sequence of parameters occurring in Δ,ψ. Let $\mathbf{E}_0$ be the derivation of the left premise and $\mathbf{E}_a$ that of the right premise. $\underline{\exists}$-inversion allows us[29] to extract from $\mathbf{E}_0$ a term $\sigma[\mathbf{a}]$ and a derivation in (QF($\mathcal{P}$)-LIND) of

(1) $\qquad \Delta, \sigma[\mathbf{a}]\le t[\mathbf{a},0] \wedge \psi^*\sigma[\mathbf{a}]\mathbf{a}0\ .$

The application of $\underline{\forall}$-inversion and then of $\underline{\exists}$-inversion to $\mathbf{E}_a$ yields a new parameter c, a term $\tau[\mathbf{a},c,a]$, and a derivation of

(2) $\qquad \Delta, \neg(c\le t[\mathbf{a},a] \wedge \psi^*c\mathbf{a}a),\ \tau[\mathbf{a},c,a]\le t[\mathbf{a},a'] \wedge \psi^*\tau[\mathbf{a},c,a]\mathbf{a}a'\ .$

Now define:

$$\begin{array}{lll} \rho(\mathbf{a},0) & =\sigma[\mathbf{a}] & \\ \rho(\mathbf{a},a') & =\tau[\mathbf{a},\rho(\mathbf{a},a),a] & \text{if } a<|s| \\ \quad\text{and} & =\rho(\mathbf{a},a) & \text{otherwise ;} \end{array}$$

ρ can be shown to be in $\mathcal{P}$. For that note first that the term s contains neither **a** nor c: **a** not due to the restrictive condition on the rule LIND, c not due to the choice in $\underline{\forall}$-inversion. Note also that t does not contain c. Using (1) and (2) we obtain derivations in (QF($\mathcal{P}$)-LIND) of

(3) $\qquad \Delta, \rho(\mathbf{a},0)\le t[\mathbf{a},0] \wedge \psi^*\rho(\mathbf{a},0)\mathbf{a}0$

and

(4) $\quad \Delta, \neg(\rho(\mathbf{a},a)\le t[\mathbf{a},a] \wedge \psi^*\rho(\mathbf{a},a)\mathbf{a}a),\ \rho(\mathbf{a},a')\le t[\mathbf{a},a'] \wedge \psi^*\rho(\mathbf{a},a')\mathbf{a}a'.$

Now we can infer from (3) and (4) by QF($\mathcal{P}$)-LIND

(5) $\qquad \Delta, \rho(\mathbf{a},|s|)\le t[\mathbf{a},|s|] \wedge \psi^*\rho(\mathbf{a},|s|)\mathbf{a}|s|$

[29] As **D** is I-normal we can assume without loss of generality that Δ contains only existentially quantified formulas.

and from (5) by $\underline{\exists}$

(6) $$\Delta, (\exists x)(x \leq t[\mathbf{a}, |s|] \wedge \psi^{*}x\mathbf{a}|s|)\ .$$

But (6) is the endsequent of **E**, established now by a derivation $\mathbf{E}^{*}$ in (QF($\mathcal{P}$)-LIND). The induction hypothesis yields the claim of the lemma, when applied to the derivation obtained from **D** by replacing **E** through $\mathbf{E}^{*}$. Q.E.D.

Subsystems of analysis. The considerations concerning the elimination of Σ^0_1-induction by means of quantifier-free induction and primitive recursion can be carried out when the purely arithmetic theory is expanded to the second-order theory (**BT**), allowing function parameters in the induction rule and in the defining equations for primitive recursive functions. The resulting theory is then expanded by two set-theoretical principles to Friedman's (**F**). In terms of the reductive program sketched in A.1 we want to reduce (**F**), considered as **P***, to the foundational **F***, here (**PRA**), i.e., to eliminate WKL and the Σ^0_1-axiom of choice – indeed, only the quantifier-free form of the axiom of choice, since these two forms are equivalent over (**BT**). I will use just the term-extraction lemma for the elimination of QF-AC; for the elimination of WKL one uses in addition a fact concerning primitive recursive functionals, namely, that they are "hereditarily majorizable". Leivant and Ignjatovic have obtained interesting characterizations of complexity classes, in particular of $\mathcal{P}$, via second order theories, and the techniques presented here are again most useful; see [Ignjatovic].

For the formulation of two crucial lemmata I assume that Δ consists only of existential formulas; $\Delta[\neg\text{QF-AC}_0]$ denotes the sequent obtained from Δ by adding negated instances (and instantiations) of the quantifier-free axiom of choice; $\Delta[\neg\text{WKL}]$ is defined similarly. We are working within (**BT**); explicit definition or λ-abstraction is given by: $(\forall x)\lambda y.t[y](x)=t[x]$, i.e., QF-$\lambda$A. All of the axioms are presented by a Post-system $\mathcal{K}$.

QF-AC$_0$-elimination. If D is an I-normal $\mathcal{K}$-derivation of $\Delta[\neg\text{QF-AC}_0]$, then there is an I-normal $\mathcal{K}$-derivation of Δ.

In this special situation we can eliminate QF-AC in favor of just QF-λA, i.e., quantifier-free comprehension. The same holds for Weak König's Lemma:

WKL-elimination. If D is an I-normal $\mathcal{K}$-derivation of $\Delta[\neg\text{WKL}]$, then there is an I-normal $\mathcal{K}$-derivation of Δ.

Assuming these two lemmata and the eliminability of Σ^0_1-induction, I give the proof of the conservation theorem I mentioned.

Theorem. $(\mathbf{BT}+\Sigma^0_1\text{-IA}+\Sigma^0_1\text{-AC}_0+\text{WKL})$ is conservative over (**BT**) for Π^0_2-sentences.

Proof. Notice that a derivation in $(\mathbf{BT}+\Sigma^0_1\text{-IA}+\Sigma^0_1\text{-AC}_0+\text{WKL})$ of the Π^0_2-statement $(\forall x)(\exists y)\psi xy$ can be transformed into an $\text{I}(\Sigma^0_1)$-normal $\mathcal{K}$-derivation with an endsequent of the form $[\neg\text{QF-AC}_0,\neg\text{WKL}], (\forall x)(\exists y)\psi xy$. This sequent can be assumed (by $\underline{\forall}$-inversion) to be of the form $[\neg\text{QF-AC}_0,\neg\text{WKL}], (\exists y)\psi ay$.

The main claim is this:

(*) Let Δ consist only of existential formulas; if D is an $\text{I}(\Sigma^0_1)$-normal $\mathcal{K}$-derivation of $\Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay$, then there is an I-normal $\mathcal{K}$-derivation E of $\Delta,(\exists y)\psi ay$.

Proof of (*) (proceeds by induction on the length of $\text{I}(\Sigma^0_1)$-normal $\mathcal{K}$-derivations). The induction step is trivial in case of $\underline{\text{LA}}$, $\underline{\text{C}}$ with atomic cut-formula, or when the last rule affects an element of Δ or the formula $(\exists y)\psi ay$. So we have to consider the cases that the last rule (1) is $\underline{\text{C}}$ with Σ^0_1-cut-formula, (2) is the Σ^0_1-induction rule, (3) introduces an instance of $\neg\text{QF-AC}_0$, or (4) introduces an instance of $\neg$WKL. Let me discuss the arguments for (1) and (2); those for (3) and (4) are analogous to that for (2). In case (1) the derivation ends in an inference of the form

$$\frac{\Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay,\phi \qquad\qquad \Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay,\neg\phi}{\Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay}$$

By induction hypothesis we get from D_0 immediately an I-normal $\mathcal{K}$-derivation $\text{D}_0{}^{*}$ of

(1) $$\Delta,(\exists y)\psi ay,\phi.$$

For the treatment of D_1 we first apply $\underline{\forall}$-inversion to $\neg\phi$ (assuming that ϕ is of the form $(\exists x)\chi x$) and then use the induction hypothesis to have an I-normal $\mathcal{K}$-derivation $\text{D}_1{}^{*}$ of

(2) $$\Delta,(\exists y)\psi ay,\neg\chi c.$$

By the term extraction lemma we get a term t from $\text{D}_0{}^{*}$ and an I-normal $\mathcal{K}$-derivation of

(3) $$\Delta,(\exists y)\psi ay,\chi t;$$

replace c in $D_1{}^*$ by t to get an I-normal $\mathcal{K}$-derivation of

(4) $\Delta,(\exists y)\psi ay,\neg\chi t$.

Joining the derivations leading to (3) and (4) by a cut – with a quantifier-free cut-formula – yields finally the desired I-normal $\mathcal{K}$-derivation of $\Delta,(\exists y)\psi ay$. (One remark should be added: In this proof we did not use at all the fact that we are dealing with an I-cut; only the logical complexity of the cut-formula mattered.)

In case (2) the derivation ends in an inference of the form

$$\frac{\Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay,\phi 0 \qquad \Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay,\neg\phi b,\phi b'}{\Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay,\phi t}$$

Here ϕt is again a Σ^0_1-formula. So we can apply the induction hypothesis to get from D_0 immediately an I-normal $\mathcal{K}$-derivation $D_0{}^*$ of

(5) $\Delta,(\exists y)\psi ay,\phi 0$.

Now use the standard trick to remove the universal quantifier by $\underline{\forall}$-inversion, then apply the induction hypothesis, and finally re-introduce the universal quantifier to obtain an I-normal $\mathcal{K}$-derivation $D_b{}^*$ of

(6) $\Delta,(\exists y)\psi ay,\neg\phi b,\phi b'$.

Joining the derivations leading to (5) and (6) by the Σ^0_1-induction rule, we obtain an $I(\Sigma^0_1)$-normal $\mathcal{K}$-derivation of $\Delta,(\exists y)\psi ay,\phi t$. But this derivation can be transformed into an I-normal $\mathcal{K}$-derivation of the same sequent by the Theorem concerning the elimination of Σ^0_1-induction. The remaining two cases (3) and (4) are treated similarly using the appropriate elimination lemmata. Q.E.D.

Now let us come back to the elimination lemmata we just applied to prove the conservativeness of (**BT**+Σ^0_1-IA+Σ^0_1-AC$_0$+WKL) over (**BT**) with respect to Π^0_2-sentences. Let me first give the *proof of the QF-AC$_0$-elimination lemma.*

Proof (by induction on the length of D). I focus on the crucial case when an instance of $\neg$QF-AC$_0$ has been introduced by the last rule in D. D has then the immediate subderivations D_0 and D_1 with endsequents $\Delta[\neg\text{QF-AC}_0]$, $(\forall x)(\exists y)\psi xy$ and $\Delta[\neg\text{QF-AC}_0],\neg(\exists f)(\forall x)\psi xf(x)$. By $\underline{\forall}$-inversion one obtains I-normal $\mathcal{K}$-derivations D_i' of

(1) $\Delta[\neg\text{QF-AC}_0,], (\exists y)\psi cy$

and

(2) $\Delta[\neg\text{QF-AC}_0,], \neg(\forall x)\psi xu(x)$,

where c and u are new number, respectively function parameters. $|D_i'| \le D_i$, for $i\le 1$, and the endsequents of D_i' satisfy the conditions on the complexity of the formulas. The induction hypothesis yields I-normal $\mathcal{K}$-derivations of

(3) $\Delta, (\exists y)\psi cy$

and

(4) $\Delta, \neg(\forall x)\psi xu(x)$.

By $\underline{\exists}$-inversion applied to the derivation leading to (3) we obtain a term $t[c]$ and a quasi-normal derivation of $\Delta, \psi ct[c]$ and, abbreviating $\lambda x.t[x](c)$ by $v(c)$, indeed of $\Delta, \psi cv(c)$, and thus also of

(5) $\Delta, (\forall x)\psi xv(x)$

Replace the function parameter u throughout the derivation leading to (4) by the λ-term v and get a derivation of

(6) $\Delta, \neg(\forall x)\psi xv(x)$.

The desired I-normal $\mathcal{K}$-derivation E is obtained by cutting (5) and (6) and by subsequent I-normalizing. Q.E.D.

Now let us proceed to the *proof of the elimination lemma concerning WKL.*

Proof (by induction on the length of D). I concentrate again on the central case when the last rule in D introduces an instance of $\neg$WKL; i.e.

$$T(f) \wedge (\forall x)(\exists y)(lh(y)=x \wedge f(y)=1) \wedge \neg(\exists g)(\forall x)\, f(\overline{g}(x))=1.$$

(Recall that T(f) is a purely universal statement, expressing that f is the characteristic function of a tree of 0-1-sequences.) Then there are I-normal $\mathcal{K}$-derivations D_i, $i\le 2$ and all shorter than D, of

$\Delta[\neg\text{WKL}], T(f)$

$\Delta[\neg\text{WKL}], (\forall x)(\exists y)\,(lh(y)=x \wedge f(y)=1)$, and

$\Delta[\neg\text{WKL}], (\forall g)(\exists x)\, f(\overline{g}(x))\ne 1$.

Using $\underline{\forall}$-inversion and the induction-hypothesis we obtain E_i ,$i \leq 2$, of

$\Delta, T(f)$

$\Delta, (\exists y)\,(lh(y)=c \wedge f(y)=1)$, and

$\Delta, (\exists x)\, f(\overline{u}(x)) \neq 1$

with new parameters c and u. $\underline{\exists}$-inversion provides terms t and s and also I-normal $\mathcal{K}$-derivations $F_1$and F_2 of

$\Delta, lh(t[c])=c \wedge f(t[c])=1$ and $\Delta, f(\overline{u}(s[u])) \neq 1$, respectively.

The terms s and t may contain further parameters, but u does not occur in t. Now observe: (i) t yields sequences of arbitrary length in the tree f that do not necessarily form a branch; (ii) $f(\overline{u}(s[u])) \neq 1$ expresses the well-foundedness of f. In short, we have a binary tree (according to E_0) that contains sequences of arbitrary length and is well-founded. This conflicting situation can be exploited by means of a formalized recursion theoretic observation, namely: s can be majorized (in the sense of [Howard]) by a numerical term s* that does not contain u, since u can be taken to be majorized by **1**. Let t[s*] be the 0-1 sequence

$t_0, \ldots, t_{s^*-1}$

and define with λ-abstraction the function u* by

$u^*(n) = t_n \quad \text{if } n < s^*$

and $u^*(n)$ equals 0 otherwise. $\overline{u^*}(s^*)$ equals t[s*]. According to E_0 f is provably a tree, and s* is a bound for s. Thus we have from F_2 a derivation of Δ, $f(\overline{u}(s^*)) \neq 1$. Replacing u by u* yields a derivation of and indeed a derivation G_2 of $\Delta, f(t[s^*]) \neq 1$ when taking into account the equation $\overline{u^*}(s^*)=t[s^*]$. From F_1 one can obtain a derivation G_1 of $\Delta, f(t[s^*])=1$ by $\underline{\wedge}$-inversion and the substitution lemma, replacing c by s*. A cut of G_1 and G_2 yields the sought for derivation E of Δ. Q.E.D.

Clearly, the Theorem does provide computational information; that is expressed in the following corollary.

Corollary. If $(\mathbf{BT}+\Sigma^0_1\text{-IA}+\Sigma^0_1\text{-AC}_0+\text{WKL})$ proves the Π^0_2-statement $(\forall x)(\exists y)\psi xy$, then there is a primitive recursive function f and a proof of $\psi af(a)$ in (**PRA**).

(**ET**) is like (**BT**) but it has defining axioms only for the Kalmar-elementary, not for all primitive recursive function(al)s and it does not contain Σ^0_1-

induction. By the same argument one can establish a conservation result analogous to that for Friedman's (**F**); then it is possible to infer the following corollary.

Corollary. If $(\mathbf{ET}+\Sigma^0_1\text{-AC}_0+\text{WKL})$ proves the Π^0_2-statement $(\forall x)(\exists y)\psi xy$, then there is a Kalmar-elementary function f and a proof of $\psi af(a)$ in (**KEA**).

Remarks. The AC_0-elimination technique is useful also in other contexts: **(i)** It was first used to prove that $(\Sigma^1_{n+1}\text{-}\mathbf{AC})$ is conservative over $(\Pi^1_n\text{-}\mathbf{CA}_{<\varepsilon 0})$ for classes F_n of formulas; here $F_0 = \Pi^1_2$, $F_1 = \Pi^1_3$ and $F_n = \Pi^1_4$ for all $n \geq 2$. **(ii)** The fact that $(\Sigma^1_{n+1}\text{-}\mathbf{AC})\upharpoonright$ is conservative over $(\Pi^1_n\text{-}\mathbf{CA})\upharpoonright$ can be also be proved using this technique; in particular, that $(\Sigma^1_1\text{-}\mathbf{AC})\upharpoonright$ is conservative over $(\Pi^1_0\text{-}\mathbf{CA})\upharpoonright \equiv (\Pi^0_\infty\text{-}\mathbf{CA})\upharpoonright$. Since the latter is a conservative extension of (**Z**), we have a reduction of $(\Sigma^1_1\text{-}\mathbf{AC})\upharpoonright$ to (**HA**). For these results see [Feferman and Sieg 1981]. **(iii)** For fragments of number theory these techniques were refined in [Sieg 1985 and 1991].

PART C. NATURALLY NORMAL PROOFS

In the first two parts of these lectures we have seen the use of the classical sequent calculus as a technical tool for achieving two ends: For *foundational reductions* of (strong) subsystems of analysis to constructive theories and for the extraction of *computational information* from proofs, thus for the characterization of the provably total functions of theories. I mentioned a third theme of proof theoretic research that goes back to Hilbert, namely, the *cognitive psychological* one. In "Über das Unendliche" Hilbert described proof theory in such a way that it can be mistaken for cognitive psychology restricted to mathematical thinking. Let me recall his remark: "The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds." If this remark has plausibility at all, then only through the emergence of Gentzen's natural deduction calculi.[30] I am turning now to their discussion.

1. Mechanization and natural deduction proofs. The mechanization of *human* reasoning has been aimed for ever since theoretical recognition of the formal character of inference steps was complemented by practical experience with intricate mechanical devices. I remind you again of Leibniz! It is only since the end of the 19th century that we have powerful logical frameworks allowing us to formalize substantive parts of human knowledge, namely, mathematics. And it is only since the middle of our century that we have sufficiently intricate (electronic) devices providing the physical underpinnings for mechanization. Up to now, it seems to me, logical frameworks that do not reflect human reasoning have been chosen for mechanization; that applies to resolution, to sequent calculi as well as to their notational variant, tableaux.

Normal Proofs. Calculi that mirror closely the structure of ordinary argumentation have been available since the mid-thirties – Gentzen's natural deduction calculi. According to Gentzen they were to reflect "as accurately as possible the actual logical reasoning involved in mathematical proofs".[31]

[30] But one must remember that Hilbert had analyzed the role of the various connectives in such a way that his system is an axiomatic formulation of the ND-rules.

[31] Gentzen in his "Investigations into logical deduction", cf. [Szabo], p. 74.

Gentzen himself gave up using ND-calculi for his metamathematical work, since they did not seem to have the marvelous properties of sequent calculi. As far as the contemporary automated theorem proving community is concerned, Fitting's remarks in his book *First order logic and automated theorem proving* (1990) are perhaps symptomatic; Fitting writes: "Hilbert systems are inappropriate for automated theorem proving. The same applies to natural deduction, since modus ponens is a rule in both." I think, on the contrary, that if we want to make progress in automated proof search, then we have to use natural deduction calculi.

How do natural deduction calculi capture the logical structure of arguments and its dependence on the syntactic form of assumptions and conclusions? They do so by incorporating *inferences from* and *to* logically complex formulas with characteristic principal connectives. The rules for each logical connective, in the case of sentential logic $\wedge$, $\vee$, $\rightarrow$, and $\neg$, are consequently divided into "elimination" and "introduction" rules. Let me just formulate the rules for negation, because they are formulated here in a way that is not the standard (Gentzen-Prawitz) mold. The negation elimination rule $\neg$E is the distinctive rule of classical logic and it is needed to prove, for example, the law of excluded middle and Peirce's law; the introduction rule $\neg$I captures the form of indirect argumentation as used in the Pythagorean proof of the irrationality of $\sqrt{2}$:

$$\dfrac{\begin{matrix}[\neg\phi]\\ \vdots\\ \varphi\end{matrix} \qquad \begin{matrix}[\neg\phi]\\ \vdots\\ \neg\varphi\end{matrix}}{\phi} \qquad\qquad \dfrac{\begin{matrix}[\phi]\\ \vdots\\ \varphi\end{matrix} \qquad \begin{matrix}[\phi]\\ \vdots\\ \neg\varphi\end{matrix}}{\neg\phi}$$

More generally, the E-rules specify how components of assumed or already established complex formulas can be used in an argument; the I-rules provide conditions under which complex formulas can be inferred from already established components. This leads directly to the formulation of very intuitive strategies; and the calculi have, after all, the crucial metamathematical properties of sequent calculi.

To state the first of these properties recall that the premise of an elimination inference containing the characteristic connective is called *major premise* and that a derivation is called *normal*, just in case there is (roughly speaking) no formula occurrence in the derivation that is both the conclusion of an I-rule and the major premise of an E-rule. In addition, the consequence of $\neg$E should not be the major premise of an elimination rule. The first central property was established by Prawitz (1965) and can be formulated in a slightly more general way than Prawitz did:

Normalization Theorem. Any derivation of G from α in the ND-calculus can be transformed into a normal derivation leading from α to G.

Here α is the sequence of assumptions from which G is derived. Prawitz's proof specifies a particular sequence of "reduction steps" to effect the transformation.[32] The second crucial fact that holds for (normal derivations in) natural deduction calculi is a corollary of the normalization theorem and states that normal derivations D of G from α have the *subformula property* in the following sense: every formula occurring in D is (the negation of) either a subformula of G or of an element in α.

Despite the "naturalness" of natural deduction calculi, the part of proof theory that deals with them has hardly influenced developments in automated theorem proving. For that the proof theoretic tradition founded on Herbrand's work and Gentzen's work on sequent calculi have been more important. The keywords here are *resolution* and *logic programming*. From a purely logical point of view this is prima facie peculiar: It is after all the subformula property of special kinds of derivations[33] that makes resolution and related techniques possible, and normal derivations in natural deduction calculi have that very property (with the minor addition mentioned above). Why is it then that natural deduction calculi have not been exploited for **automated proof search**? The answer to this broad question lies, it seems to me, in answers to **three crucial questions:** (1) How can one specify through a calculus normal derivations? (2) How can one construct a search space that allows the formulation of strategies for finding such derivations? and (3) How can we prove the termination of strategies?

In the *case of the sequent calculus,* the analogue to the first question has a trivial answer: The calculus without the cut rule! The syntactic normalization or cut-elimination procedure is, however, not crucial for automated

[32] And holds, to be precies, only for a part of the classical calculus. The (strong) normalization theorem for the full calculus is established by Stalmark (1991).

[33] Derivations in Herbrand's calculus and derivations in the sequent calculus without cut have the *subformula property*: they contain only subformulas of their endformula, respectively endsequent.

deduction; it is the direct completeness proof for the cut-free part that is fundamental, since algorithms for finding cut-free derivations are refinements of strategies used in that proof. Such strategies realize the heuristic idea of *searching for semantic counterexamples* and are systematic procedures that yield trees σ such that *either* one of σ's branches allows the definition of a counterexample to "Δ has G as a logical consequence" *or* σ constitutes a cut-free derivation of the sequent $\neg\Delta$,G. In the *case of natural deduction calculi,* the fact that normal derivations are sufficient for obtaining all logical consequences from given assumptions is not established directly at all, but rather by combining the completeness theorem for the calculus with the normal form theorem; there is no direct characterization answering question (1) in the literature. In order to obtain an answer to the first question I introduce *intercalation calculi*; they provide natural frameworks for answering also the second question. The completeness proofs for the calculi provide the answer to the third question. I will present these considerations first for classical sentential logic; I should note that, for this logic, Richard Scheines and I implemented the first complete and heuristically guided proof search system.

Intercalation calculi (for sentential logic). The broad problem is this: How can one derive a conclusion or goal G from assumptions $\phi_1, \dots ,\phi_n$? or, to put it more vividly, how can one close – via logical rules – the gap between a conclusion G and assumptions $\phi_1, \dots ,\phi_n$? This question is at the heart of spanning the search space via intercalation calculi. The basic rules of such calculi are local reformulations of those for Gentzen's natural deduction calculi, but it is the preservation of inferential information and the restricted way in which the rules are used to close the gap (and thus to build up derivations) that is distinctive. I will discuss in this lecture only classical sentential logic; however, the theoretical considerations can be extended to predicate logic and to non-classical logics, for example, intuitionistic logic.[34] (The extension to predicate logic will be sketched in the next lecture.)

The intercalation rules operate on triples of the form $\alpha;\beta?G$. α is a sequence of formulas, the *available assumptions;* G is the current *goal;* β is a sequence of formulas obtained by $\wedge$-elimination and $\rightarrow$-elimination from

[34] That was done by Saverio Cittadini in his M.S. thesis written in May 1991; see [Cittadini 1992].

elements in α. To facilitate the description of rules and parts of search trees let us agree on some conventions. I let lower case Greek letters α, β, γ, δ, ... range over finite sequences of formulas; as syntactic variables over formulas we use ϕ, ψ, χ, ...; ρ, σ, τ (with indices) will range over trees. At first I consider only formulas in the language of sentential logic using the connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$; I also use $\perp$ (falsum) as an auxiliary symbol. $\phi \in \alpha$ expresses that ϕ is an element of the sequence α; α,β is short for the concatenation $\alpha * \beta$ of the sequences α and β; α,ϕ stands for the sequence $\alpha * <\phi>$, where $<\phi>$ is the sequence with ϕ as its only element. Finally, I write $\alpha \equiv \beta$ iff the sets of formulas in the sequences α and β are identical. There are three kinds of intercalation rules: those corresponding to E- rules for $\wedge$, $\vee$, $\rightarrow$; those corresponding to I-rules for $\wedge$, $\vee$, $\rightarrow$; and finally rules for negation. Let me first list the rules of the first kind, i.e., the $\downarrow$-rules:

$\downarrow\wedge_i$: $\alpha;\beta?G,\ \phi_1\wedge\phi_2 \in \alpha\beta,\ \phi_i \notin \alpha\beta \ \Rightarrow\ \alpha;\beta,\phi_i?G$ for i=1 or 2

$\downarrow\vee$: $\alpha;\beta?G,\ \phi_1\vee\phi_2 \in \alpha\beta,\ \phi_1 \notin \alpha\beta,\ \phi_2 \notin \alpha\beta \ \Rightarrow\ \alpha,\phi_1;\beta?G$ AND $\alpha,\phi_2;\beta?G$

$\downarrow\rightarrow$: $\alpha;\beta?G,\ \phi_1\rightarrow\phi_2 \in \alpha\beta,\ \phi_1 \in \alpha\beta,\ \phi_2 \notin \alpha\beta \ \Rightarrow\ \alpha;\beta,\phi_2?G$

The side conditions of these rules avoid repeating the "same questions"; $\alpha;\beta?G$ is the *same question as* $\alpha^*;\beta^*?G$ just in case the sets of formulas in the sequences α,β and α^*,β^* are identical. Now I formulate the rules that correspond to inverted introduction rules, i.e., $\uparrow$-rules.

$\uparrow\wedge$: $\alpha;\beta?\phi_1\wedge\phi_2 \ \Rightarrow\ \alpha;\beta?\phi_1$ AND $\alpha;\beta?\phi_2$

$\uparrow\vee$: $\alpha;\beta?\phi_1\vee\phi_2 \ \Rightarrow\ \alpha;\beta?\phi_1$ OR $\alpha;\beta?\phi_2$

$\uparrow\rightarrow$: $\alpha;\beta?\phi_1\rightarrow\phi_2 \ \Rightarrow\ \alpha,\phi_1;\beta?\phi_2$

The rules for negation are split into three, where $\perp$ is considered as a placeholder for (the conjunction of) a pair of contradictory formulas:

$\perp_c$: $\alpha;\beta?\phi,\ \phi\neq\perp \ \Rightarrow\ \alpha,\neg\phi;\beta?\perp$

$\perp_i$: $\alpha;\beta?\neg\phi \ \Rightarrow\ \alpha,\phi;\beta?\perp$

$\perp_{\mathfrak{F}}$: $\alpha;\beta?\perp,\ \varphi\in\mathfrak{F}(\alpha) \ \Rightarrow\ \alpha;\beta?\varphi$ AND $\alpha;\beta?\neg\varphi$.

In the last rule $\mathfrak{F}(\alpha)$ is the finite class of formulas consisting of all PROPER subformulas of elements in α. Clearly, $\perp_{\mathfrak{F}}$ is inapplicable in case $\mathfrak{F}(\alpha)$ is empty. $\mathfrak{F}(\alpha)$ is always finite; and that is crucial for the finiteness of the search space. Operations leading to smaller and yet sufficient classes can be specified; here I simply remark that double negations can be discounted in the following sense: If $\neg\phi$ is in $\mathfrak{F}(\alpha)$, then we consider only the pair $\neg\phi$ and ϕ in the first two negation rules and not also $\neg\phi$ and $\neg\neg\phi$. The various calculi we are considering are distinguished through the operation $\mathfrak{F}$, and I denote a particular calculus by IC($\mathfrak{F}$).

The problem space. The intercalation calculus provides the computational underpinnings for specifying informal approaches to proof search: its rules are used to construct a search space that contains all possible ways of closing the gap between α and G via the rules of the intercalation calculus. And as will be seen later, the space "codes" all possibilities of building up normal derivations leading from α to G in the natural deduction calculus. Within this space we search for a gap-closing subtree such that it determines uniquely a natural deduction derivation from α to G; if the search fails, the search space will contain enough information to yield a semantic counterexample. From this sketch of a completeness proof for the intercalation calculus you see that there is a family resemblance to completeness proofs for the sequent calculus without cut. The difference can be put sharply as follows: *In the case of the sequent calculus, one tries to find a semantic counterexample and, if that search fails, one actually has found a proof; in the case of the intercalation calculus, one tries to find a proof and, if that search fails, one has a counterexample.*[35]

As an example of how the intercalation rules are used to build up the search space for a question $\alpha;?G$, let me show the search tree for the question $?P\vee\neg P$. It is *partially* presented in Diagram 1 (of the Appendix to this lecture on p. 72). We start out by applying three intercalation rules to obtain three new questions, namely, ?P OR $?\neg P$ OR, proceeding indirectly, $\neg(P\vee\neg P);?\perp$. That the branching in the tree is disjunctive is indicated by □. Let us pursue the leftmost branch in the tree: To answer ?P we have to use $\perp_c$ and, because of the restriction on the choice of contradictory pairs, we have only to ask $\neg P;?P$ AND $\neg P;?\neg P$. ⊞ indicates that the branching is conjunctive here. In the first case only $\perp_c$ can be applied and leads to the same question we just analyzed: Using $\neg P$ as an assumption, $\perp$ has to be proved. Thus we *close this*

[35] The sequent calculus provides a direct framework for motivated search for a derivation; indeed, the search tree is a derivation, if the sequent is provable. But a sequent proof is far from reflecting the structure of ordinary arguments. In the case of resolution based procedures, one also has the non-trivial problem of finding an associated natural deduction derivation. Cf. Andrews, Mints, and Pfenning.

branch with a circled F, linking it to the same earlier question on the branch. In the second case the gap between assumptions and goal is obviously closed, so we top this branch with a circled T. The other parts of the tree are constructed in a similar manner. But the tree is not quite full: At the nodes that are distinguished by arrows the additional contradictory pair consisting of P and ¬P has to be considered. At nodes 2 and 3 the resulting branches do not help in closing the gap; at node 1, in contrast, the resulting subtree is of interest and will be discussed below.

The darkened subtrees (in Diagram 1) contain enough information for the extraction of derivations in a variety of styles of natural deduction. For our calculus we can easily obtain the corresponding derivations; namely:

$$\dfrac{\dfrac{\dfrac{\dfrac{\cancel{\neg P}}{P \vee \neg P} \qquad \cancel{\neg(P \vee \neg P)}}{P}}{P \vee \neg P} \qquad \cancel{\neg(P \vee \neg P)}}{P \vee \neg P}$$

The second derivation is analogous to this one, except that the roles of P and ¬P are interchanged; finally, the derivation that emerges from the undrawn part at node 1 is this:

$$\dfrac{\dfrac{\dfrac{\cancel{\neg P}}{P \vee \neg P} \qquad \cancel{\neg(P \vee \neg P)}}{P} \qquad \dfrac{\dfrac{\cancel{P}}{P \vee \neg P} \qquad \cancel{\neg(P \vee \neg P)}}{\neg P}}{P \vee \neg P}$$

The (full) search or intercalation tree is specified inductively by applying the intercalation rules to the initial question or to the "non-terminal" leaves of an already obtained partial search tree. In either case one addresses questions of the form $\alpha^*;\beta^*?G^*$. We distinguish two cases:

1. G^* is different from $\perp$: apply intercalation rules in all possible ways, e.g., in the order $\downarrow\wedge_1$, $\downarrow\wedge_2$, $\downarrow\rightarrow$, $\downarrow \mathbf{v}$, $\uparrow\wedge$, $\uparrow\rightarrow$, $\uparrow\mathbf{v}$, and finally either $\perp_i$ or $\perp_c$, **unless** $G^* \in \alpha^*,\beta^*$; in that case close the branch with **T**.
2. G^* is $\perp$: apply $\perp_{\mathfrak{F}}$ with $\varphi \in \mathfrak{F}(\alpha^*)$, **unless** $\mathfrak{F}(\alpha^*)$ is empty or there is a question $\alpha_1;\beta_1?\perp$ on the branch determined by $\alpha^*;\beta^*?\perp$ with α_1 identical to α^*; in the latter cases close the branch with F.

The *intercalation tree* is constructed in this way for any question $\alpha;?G$.. A branch in this tree constitutes a sequence of *subquestions* for $\alpha;?G$ of the form $<\alpha_i;\beta_i?G_i>_{i\in I}$; I is a subset of **N**. The sequence satisfies the obvious conditions: (1) $\alpha_0;\beta_0?G_0$ is $\alpha;?G$, and (2) for any $i>0$ the element $\alpha_i;\beta_i?G_i$ is obtained from the immediately preceding subquestion as (one of) the conclusion(s) of an intercalating rule. Due to the finiteness of $\mathfrak{F}$ and the complexity reducing character of the $\downarrow$ - and $\uparrow$ -rules the sequences of subquestions are all finite; as the intercalation tree is finitely branching we have the first part of the following proposition:

Proposition. The intercalation tree for the question $\alpha;?G$ is finite, and each branch is closed with either **T** or **F**.

Proof. Because of the above observation, only the second part of the proposition has to be established. So assume that a particular leaf with question $\alpha^*;\beta^*?G^*$ is not closed with **T**. Then we must have, first of all, that neither a $\downarrow$ -rule, nor a $\uparrow$ -rule is applicable; $\perp_i$ and $\perp_c$ are also not applicable, so G^* must be $\perp$! But then the construction is terminated, because $\perp_{\mathfrak{F}}$ is not applicable either. Thus, the branch is closed with **F**. Q.E.D.

Every branch in a search tree is finite and is topped by either a circled **T** or **F**. This assignment to the leaves can be easily (and uniquely) extended to the whole tree and thus determines the value of the original question. One can show two facts: (1) If **T** is assigned to the root of the intercalation tree, then there is a normal derivation leading from the assumptions to the goal of the question; (2) If **F** is assigned to the root of the intercalation tree, then there is not only no normal derivation, but no derivation at all: The intercalation tree contains enough information to show that the inference from α to G is semantically invalid. Let me address just (2); the first fact is established by a rather straightforward inductive argument.

Extracting Counterexamples. By the evaluation of intercalation trees we know that a question $\alpha;?G$ obtains the value **T** or **F**. In case the value is **T** we can determine an associated normal derivation. In case the question has value **F**, we have as an immediate consequence "The search failed!" But that only means the particular possibilities of building up derivations – as reflected in the construction of the intercalation tree – do not lead to a derivation that establishes G from assumptions in α. We can do better: a *special* branch in the intercalation tree can be selected and be used to define a semantic counterexample to the inference from α to G. Clearly, if the question $\alpha;?G$ evaluates as **F**, then so does $\alpha,G'';?\bot$, where G'' is $\neg G$ if G is not a negation and is its unnegated part otherwise. We establish the following lemma:

Counterexample extraction lemma. For any α and G: If the intercalation tree σ for $\alpha;?G$ evaluates as **F**, then it contains a canonical refutation branch ρ that determines a valuation v with $v'(\phi)=0$ for all $\phi\in\alpha$ and $v'(G)=1$. (That is, v is a counterexample to the inference from α to G.)

The intercalation tree σ is evaluated as **F** and thus it will be quite direct to see that the following construction leads to a branch ρ through σ, if $\mathfrak{F}(\alpha*\langle G''\rangle)$ is non-empty. If this set is empty, $\alpha*\langle G''\rangle$ consists only of sentential letters and the valuation v, defined by $v(P)=0$ iff $P\in\alpha*\langle G''\rangle$, is a counterexample. If the set of proper subformulas of the elements of $\alpha*\langle G''\rangle$ is non-empty, we need a more sophisticated argument and, naturally, some auxiliary definitions. The finite set $\mathfrak{F}(\alpha*\langle G''\rangle)$ for the negation rule $\bot_{\mathcal{F}}$ can be enumerated (without repetition). $\langle H'_i\rangle_{i\in I}$ is the sequence of formulas obtained from such an enumeration by letting H' be H if H is not a negation and its unnegated part otherwise; $I := \{ i \mid 1\leq i\leq n \}$; let H'_0 be G''. For $i\in I$ define

$$\kappa(\beta,i) = \begin{array}{ll} \mu k.\ (i\leq k\leq n \wedge H'_k\notin\beta \wedge \neg H'_k\notin\beta) & \text{if there is such an } H'_k \\ 0 & \text{otherwise} \end{array}$$

The sequence of nodes of ρ (and more) is determined as follows:

$$\begin{array}{lcll}
\rho^*(0) & = & \alpha;?G & \\
\alpha_0 & = & \alpha*\langle G''\rangle & \\
\lambda(\alpha_0,1) & = & \kappa(\alpha_0,1) & \\
\rho^*(1) & = & \alpha_0;?\bot & \\
\text{if } 0<m: & & & \\
\rho^*(2m) & = & \alpha_{m-1};?H'_{\lambda(\alpha0,m)} & \text{if } [\alpha_{m-1};?H'_{\lambda(\alpha0,m)}] \text{ is } \mathbf{F} \\
 & & \alpha_{m-1};?\neg H'_{\lambda(\alpha0,m)} & \text{otherwise} \\
\alpha_m & = & \alpha_{m-1}*\langle\neg H'_{\lambda(\alpha0,m)}\rangle & \text{if goal of } \rho^*(2m) \text{ is } H'_{\lambda(\alpha0,m)} \\
 & & \alpha_{m-1}*\langle H'_{\lambda(\alpha0,m)}\rangle & \text{if goal of } \rho^*(2m) \text{ is } \neg H'_{\lambda(\alpha0,m)} \\
\lambda(\alpha_0,m+1) & = & \kappa(\alpha_m,\lambda(\alpha_0,m)) & \\
\rho(2m+1) & = & \alpha_m;?\bot &
\end{array}$$

Let ν be the smallest m with $\lambda(\alpha_0,m+1)=0$; then $\alpha_\nu \equiv \alpha_{\nu+1}$, and the refutation branch ρ is the restriction of ρ^* to $\{m \mid m2\nu+1\}$. Let me illustrate (and clarify) this construction through Diagram 2: At each step in selecting the next node of the canonical branch ρ one or the other indicated possibility of proceeding must obtain (as long as the set of assumptions can be properly extended), because not both conclusions of $\bot_{\mathcal{F}}$ with the contradictory pair H'_k and $\neg H'_k$ can be evaluated as **T**. (In case both are evaluated as **F**, choose the leftmost.) So we have selected a branch ρ through the intercalation tree σ that is **F**-closed, all of whose nodes evaluate as **F**, and whose "closing node", indicated by the checkered rectangle, is such that no rule other than $\bot_{\mathcal{F}}$ is applicable. Application of that rule with any formula in $\mathfrak{F}(\alpha*\langle G''\rangle)$, in particular with H_0, leads to the canonical closing indicated in the diagram.

Let $\Gamma := \{\phi \mid \phi\in\alpha_{\nu+1}\}$; thus, Γ consists of all the formulas appearing on the l.h.s. of the question mark at ρ's top node. The set Γ has important syntactic closure properties and this can be exploited to define a valuation that will serve as a model for $\alpha*\langle G''\rangle$. We establish first the closure properties.

Closure lemma. For all subformulas ϕ_1, ϕ_2 of $\alpha*\langle G''\rangle$ we have:

(i) either ϕ_1 or $\neg\phi_1$ is in Γ, but not both;

(ii) $\neg\neg\phi_1\in\Gamma \Rightarrow \phi_1\in\Gamma$;

(iii) $(\phi_1\wedge\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ and $\phi_2\in\Gamma$;
$\neg(\phi_1\wedge\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ or $\neg\phi_2\in\Gamma$;

(iv) $(\phi_1\vee\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ or $\phi_2\in\Gamma$;
$\neg(\phi_1\vee\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ and $\neg\phi_2\in\Gamma$;

(v) $(\phi_1\rightarrow\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ or $\phi_2\in\Gamma$;
$\neg(\phi_1\rightarrow\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ and $\neg\phi_2\in\Gamma$.

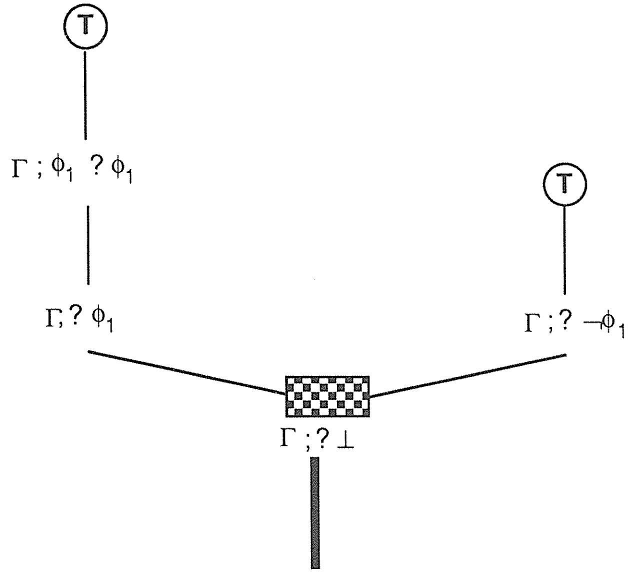
Proof. (i) is direct from the construction. (ii) is an almost immediate consequence of (i): Assume $\neg\neg\phi_1\in\Gamma$ and $\phi_1\notin\Gamma$; from the second assumption and the first part of (i) it follows that $\neg\phi_1\in\Gamma$. But that together with the first assumption contradicts the second part of (i).

Now let me establish (iii) paradigmatically to show the pattern of further argumentation. We have to show:

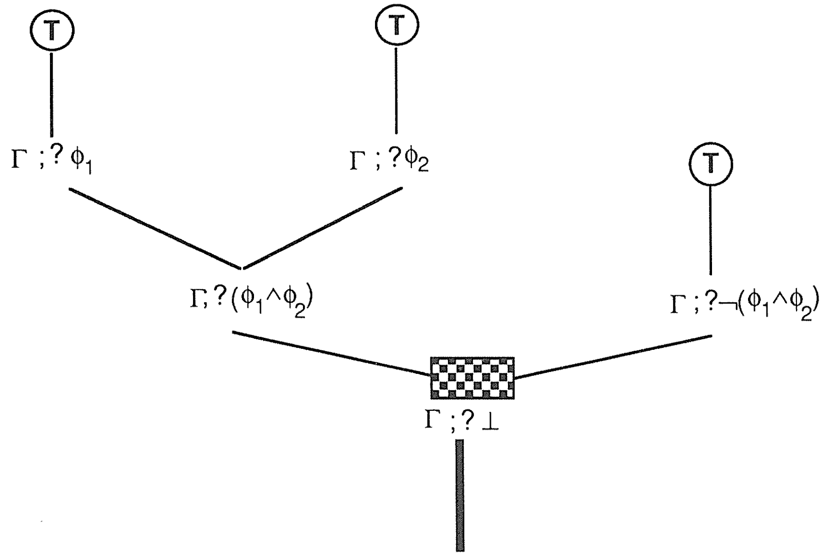
(*) $(\phi_1\wedge\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ and $\phi_2\in\Gamma$ and

(**) $\neg(\phi_1\wedge\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ or $\neg\phi_2\in\Gamma$.

For (*) assume $(\phi_1\wedge\phi_2)\in\Gamma$ and $\phi_1\notin\Gamma$ (the case $\phi_2\notin\Gamma$ is symmetric); by (i) $\neg\phi_1\in\Gamma$. Given these conditions we can close the branch as follows, applying $\downarrow\wedge_1$ to the left node above the checkered one :



This contradicts the fact that the checkered node is evaluated as **F**. (**) is established in an analogous way applying $\uparrow \wedge$ instead of $\downarrow \wedge_1$: Assume that $\neg(\phi_1 \wedge \phi_2) \in \Gamma$, $\neg\phi_1 \notin \Gamma$, and $\neg\phi_2 \notin \Gamma$; from the last two assumptions and (i) follows $\phi_1 \in \Gamma$ and $\phi_2 \in \Gamma$, and the branch can be closed as indicated in the next diagram.



Q.E.D.

Now define a valuation by $v(P) = 0$ *iff* $P \in \Gamma$. Using this valuation and the closure lemma we can prove the **Proposition** that for every $\phi \in \Gamma$: $v'(\phi)=0$. Hence v is a model for $\alpha * <G''>$; this concludes the proof of the lemma concerning the extraction of counterexamples. Putting these considerations together, we obtain a completeness theorem for classical sentential logic in the following form:

Completeness theorem. The intercalation tree for the question α;?G allows us to determine either a normal derivation G from α or a branch that provides a counterexample to the inference from α to G.

So we have a semantic argument for the normalizability of ND proofs.

Normal form theorem. If G can be proved from assumptions in α, then there is a normal proof of G from α.

This is, as far as I know, the first *semantic proof of the normal form theorem* for a natural deduction calculus. It is also extremely easy to obtain (from intercalation derivations) the interpolation theorem.

APPENDIX to Lecture C.1

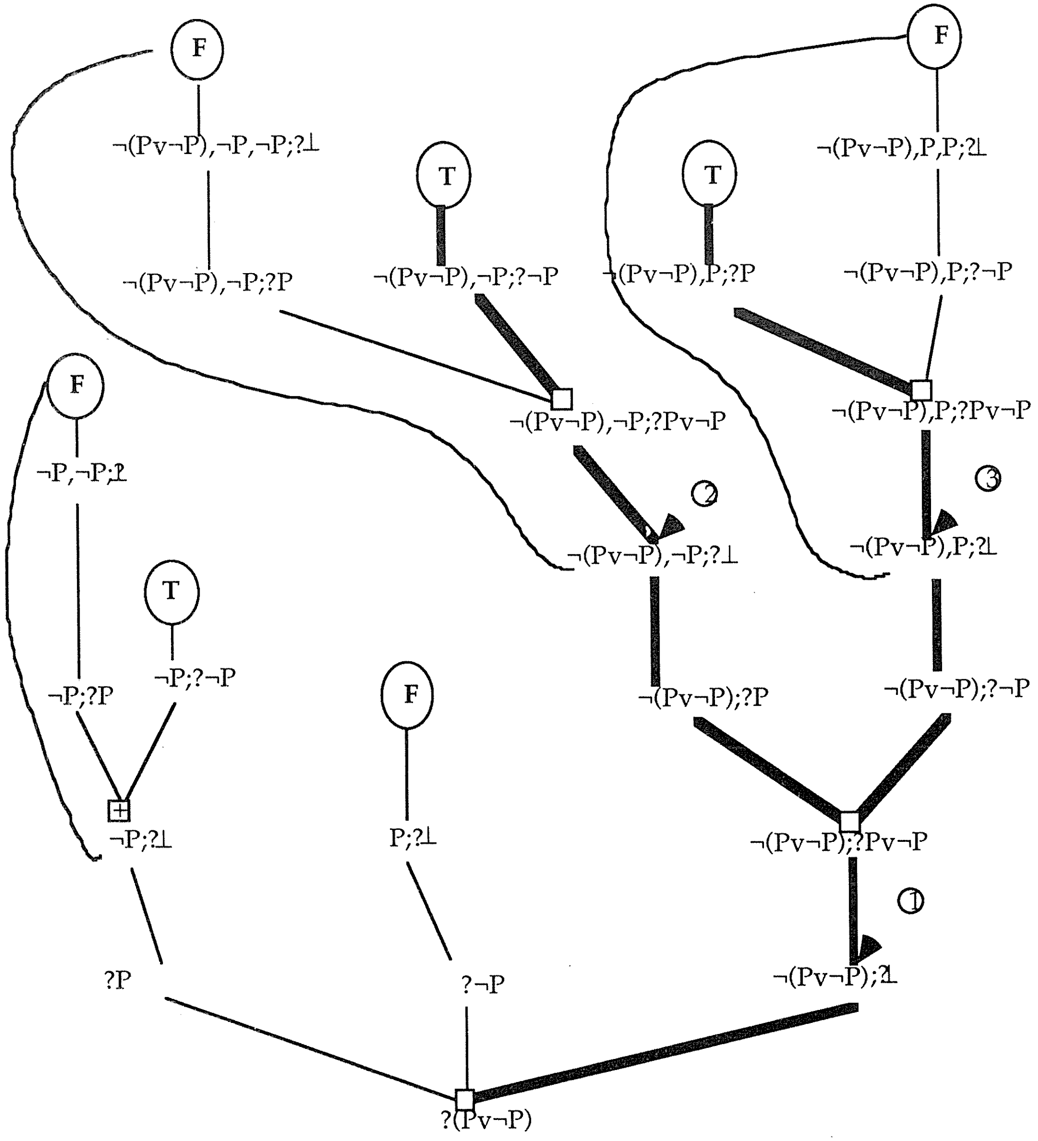


Diagram 1

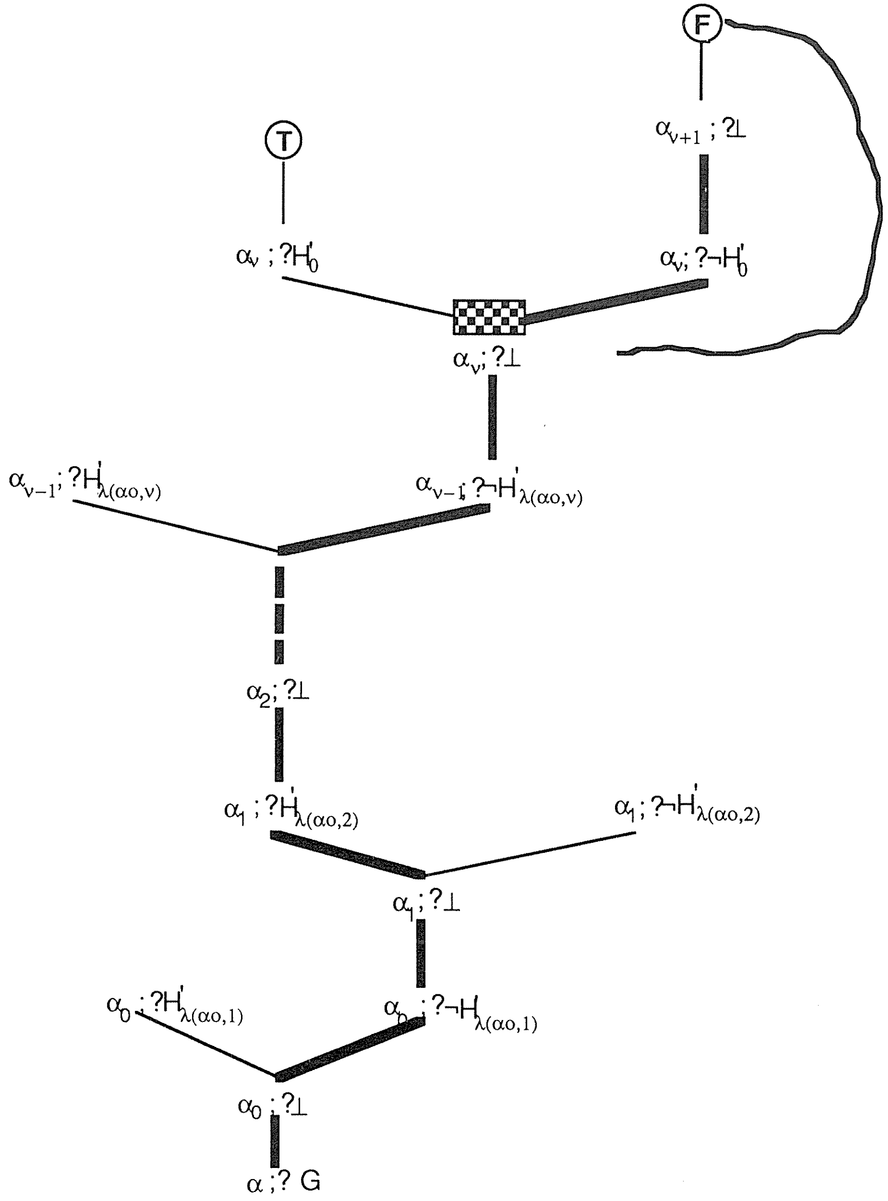

F
T
$\alpha_{\nu+1}$; ?⊥
α_ν ; ?H'_0
α_ν; ?¬H'_0
α_ν; ?⊥
$\alpha_{\nu-1}$; ?$H'_{\lambda(\alpha o,\nu)}$
$\alpha_{\nu-1}$; ?¬$H'_{\lambda(\alpha o,\nu)}$
α_2; ?⊥
α_1 ; ?$H'_{\lambda(\alpha o,2)}$
α_1; ?¬$H'_{\lambda(\alpha o,2)}$
α_1 ; ?⊥
α_0 ; ?$H'_{\lambda(\alpha o,1)}$
α_0 ; ?¬$H_{\lambda(\alpha o,1)}$
α_0 ; ?⊥
α ; ? G

Diagram 2

2. Proof search. This last lecture will consist of two parts: a systematic, logical part extending the intercalating considerations to predicate logic; a speculative, partly historical part extending my earlier analytic report on computablity. The latter will be concerned mainly with Turing's and Gödel's views on whether and how the limits of mechanical procedures can be overcome in mathematics. They agree that such a phenomenon would show up in the search for proofs – *including* the finding of axioms and the introduction of concepts.

Problem space for predicate logic. The metamathematical considerations of the last lecture can be extended to *classical* predicate logic. To that end we consider the following formulation of the elimination and introduction rules for the quantifiers. For $\forall$:

$\forall$E
$$\frac{(\forall x)\phi x}{\phi t}$$

$\forall$I
$$\frac{\phi a}{(\forall x)\phi x}$$

The I-rule must satisfy the restriction that a does not occur in any assumption on which the derivation of ϕa depends. – For $\exists$ we have the rules:

$\exists$E
$$\frac{(\exists x)\phi x \qquad \begin{matrix}[\phi a]\\ \vdots \\ \eta\end{matrix}}{\eta}$$

$\exists$I
$$\frac{\phi t}{(\exists x)\phi x}$$

with the usual restriction on the elimination rule, namely, a must not occur in η or $(\exists x)\phi x$ nor in any assumption (other than ϕa) on which the proof of (the upper occurrence of) η depends.

In building up the intercalation tree one also applies quantifier rules "to close the gap between assumptions and conclusions". In the formulation of the rules $T(\gamma,G)$ denotes the set of terms occurring in the formulas of γ,G.

$\downarrow\forall$: $\alpha;\beta?G,\ (\forall x)\phi x\in\alpha\beta,\ t\in\mathfrak{T}(\alpha\beta,G),\ \phi t\notin\alpha\beta \ \Rightarrow\ \alpha;\beta,\phi t?G$

$\downarrow\exists$: $\alpha;\beta?G,\ (\exists x)\phi x\in\alpha\beta$, a is new for $\alpha,(\exists x)\phi x,G$, and there is no $t\in\mathfrak{T}(\alpha\beta,G)$ with $\phi t\in\alpha\beta \ \Rightarrow\ \alpha,\phi a;\beta?G$

$\uparrow\forall$: $\alpha;\beta?(\forall x)\phi x$, a is new for $\alpha,(\forall x)\phi x \ \Rightarrow\ \alpha;\beta?\phi a$

$\uparrow\exists$: $\alpha;\beta?(\exists x)\phi x,\ t\in\mathfrak{T}(\alpha\beta,G) \ \Rightarrow\ \alpha;\beta?\phi t$

Intercalation trees are now inductively specified as in the case of sentential logic: if $\alpha^*;\beta^*?G^*$ is an open question, all possibilities of intercalating formulas are considered. In case G^* is different from $\bot$ one proceeds, e.g., in the order $\downarrow\forall$, $\downarrow\&_1$, $\downarrow\&_2$, $\downarrow\rightarrow$, $\downarrow\exists$, $\downarrow\mathbf{v}$, $\uparrow\forall$, $\uparrow\&$, $\uparrow\rightarrow$, $\uparrow\exists$, $\uparrow\mathbf{v}$, and finally either $\bot_i$ or $\bot_c$; in case G^* is $\bot$ we apply $\bot_{\mathfrak{F}}$ with $\mathfrak{F}$ containing all proper subformulas of α^* (where subformulas of quantified formulas are taken only with terms in $\mathfrak{T}(\alpha^*,\bot)$). Branches are closed with **T** and **F** under the same conditions as before. However, intercalation trees will in general not be finite; that means at every stage there will be a branch without a definite value, and to evaluate partial trees σ^* we assign a third value **O** to the leaves of such branches. Given the valuation v_{σ^*}, the value of the question at σ^*'s root is determined by recursion on σ^* following Kleene's scheme [IM, p. 334] for three-valued logic:

$$\begin{array}{lcll} [N]_{\sigma^*} & = & v(N) & \text{if N is a leaf of } \sigma^* \\ [N]_{\sigma^*} & = & [M]_{\sigma^*} & \text{if M is the unique predecessor of N} \end{array}$$

in case N is at a conjunctive branching,

$$\begin{array}{lcll} [N]_{\sigma^*} & = & \mathbf{T} & \text{if for all immediate predecessors M of N: } [M]_{\sigma^*}=\mathbf{T} \\ & & \mathbf{F} & \text{if for some immediate predecesor M of N: } [M]_{\sigma^*}=\mathbf{F} \\ & & \mathbf{O} & \text{otherwise} \end{array}$$

in case N is at a disjunctive branching,

$$\begin{array}{lcll} [N]_{\sigma^*} & = & \mathbf{F} & \text{if for all immediate predecessors M of N: } [M]_{\sigma^*}=\mathbf{F} \\ & & \mathbf{T} & \text{if for some immediate predecesor M of N: } [M]_{\sigma^*}=\mathbf{T} \\ & & \mathbf{O} & \text{otherwise} \end{array}$$

The intercalation tree σ for $\alpha;?G$ is thus defined in stages as follows: σ_0 is $\alpha;?G$; σ_{n+1} is σ_n if $[\alpha;?G]_{\sigma n}$ is either **T** or **F**, otherwise σ_{n+1} is obtained from σ_n by expanding each open branch by all applicable rules. Three possibilities can arise: (1) for some $n\in N$ $[\alpha;?G]_{\sigma n}=\mathbf{T}$, (2) for some $n\in N$ $[\alpha;?G]_{\sigma n}=\mathbf{F}$, and (3) for all $n\in N$ $[\alpha;?G]_{\sigma n}=\mathbf{O}$. In the first case a normal derivation can be associated with a subtree of σ_n; the second case provides a finite counterexample; the third case requires additional considerations. For case (1) one selects an appropriate subtree and proves (by induction) that it determines uniquely a normal derivation of G from elements in α. The associated derivations have the subformula property; i.e., every formula that occurs in them is either a subformula of an element of α or of G, except possibly for assumptions that are cancelled by $\neg$E. "Subformula" is again taken in the sense appropriate for predicate logic: any instance ϕt is a subformula of the quantified formula $(Qx)\phi x$.

Completeness and normal form. For case (2) we can construct a finite canonical refutation branch as in sentential logic and define from it a counterexample. Case (3) requires a little more circumspection: Instead of directly constructing a refutation branch, we determine first a particular infinite subtree of the intercalation tree; König's Lemma is then applied to this *canonical refutation tree* and yields an infinite branch from which a counterexample can be defined.

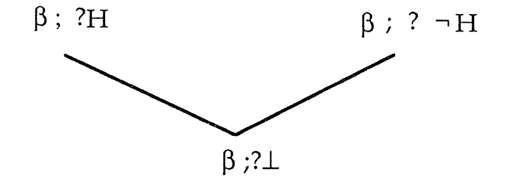
Counterexample extraction. For any α and G: if the intercalation tree σ for α;?G is such that for each natural number n $[\alpha;?G]_{\sigma n}=\mathbf{O}$, then σ contains an infinite refutation branch ρ that determines a structure $\mathfrak{M}$ with $\mathfrak{M} \models \phi$, for all ϕ in α, and $\mathfrak{M} \models \neg G$. Thus, $\mathfrak{M}$ is a counterexample to the inference from α to G.

The reason for having to cut down the intercalation tree σ to the canonical refutation tree τ is this: Refutation branches have to satisfy suitable syntactic closure conditions, and it is trivial to construct infinite branches of σ that don't. So we define τ in such a way that all of its infinite branches satisfy the closure conditions. The pertinent considerations extend those for sentential logic with variations on Henkin and tableaux constructions, and thus I emphasize only the crucial points.

The construction of τ (as a subtree of the intercalation tree σ) for the question α;?G proceeds in two waves: The first aims for "sub-maximization" with respect to a given finite set of formulas, whereas the second introduces new subformulas by witnessing – through instances with new variables – existential and negated universal formulas that occur on the l.h.s. of ?. We start out the construction of the binary tree τ (using conventions and definitions from the last lecture) with the first wave for the enumeration of the proper subformulas of formulas in $\alpha*<G">$ (where immediate subformulas of quantified formulas are taken only with terms in $\mathfrak{T}(\alpha*<G">,\bot)$:

$$\begin{array}{lll}
\tau(0) & = & \alpha;?G \\
\alpha_0 & = & \alpha*<G"> \\
\lambda(\alpha_0,1) & = & \kappa(\alpha_0,1) \\
\tau(1) & = & \alpha_0;?\bot
\end{array}$$

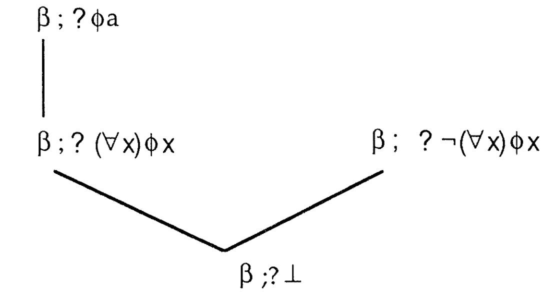
Now let 0<m; at level 2m we extend each open branch with a question of the form β;?$\bot$ at its leaf by



if both questions β;?H and β;?$\neg$H evaluate as **O**; if only one of them evaluates as **O**, then the branch is extended at just that question. And one of these cases must hold, because the question β;?$\bot$ evaluates as **O**. (Clearly, as before, H is the first element in the given enumeration that extends β properly.) At the next level 2m+1, every open branch is extended by applying the appropriate negation rule. After finitely many steps this construction cannot be continued. However, at least one branch in the tree constructed so far has to be open (for extensions by rules other than $\bot_{\mathcal{F}}$), as for all n$\in$N $[\alpha;?G]_{\sigma n}=\mathbf{O}$. In sentential logic, as we saw, that cannot happen; the resulting set of formulas Γ is deductively closed in the sense of the earlier Closure Lemma. Here, some of the Γ's associated with the top nodes cannot satisfy the closure conditions

$$\begin{array}{ll}
 & (\exists x)\phi x \in \Gamma \;\Rightarrow\; \phi t \in \Gamma \text{ for some term t} \\
\text{and} & \neg(\forall x)\phi x \in \Gamma \;\Rightarrow\; \neg\phi t \in \Gamma \text{ for some term t.}
\end{array}$$

In the first case the rule $\downarrow\exists$ is applicable (with a canonically chosen new variable); in the second case we are able to extend the branch in the following way (also with a canonically chosen new variable):



The right extension closes with **T**, whereas the left one remains open. This then brings us to the second wave: We apply $\downarrow\exists$ in all possible cases and then perform the above analysis for all $\neg(\forall x)\phi x$ for which no negated instance is available. Now the first wave can be repeated for an extended set of formulas, and so on, obviously! We obtain in this way an infinite, binary subtree τ of the intercalation tree; König's Lemma applied to this canonical refutation tree yields an infinite branch ρ. Define $\Gamma_\rho = \{\psi \mid \psi$ occurs on the l.h.s. of ? in some question on $\rho\}$; this set has all the appropriate closure properties needed to serve as the basis for the counterexample definition.

Closure Lemma. For all subformulas ϕ_1, ϕ_2 of $\alpha*\langle G''\rangle$ we have:

(i) either ϕ_1 or $\neg\phi_1$ is in Γ, but not both;

(ii) $\neg\neg\phi_1\in\Gamma \Rightarrow \phi_1\in\Gamma$;

(iii) $(\phi_1\wedge\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ and $\phi_2\in\Gamma$;
$\neg(\phi_1\wedge\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ or $\neg\phi_2\in\Gamma$;

(iv) $(\phi_1\vee\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ or $\phi_2\in\Gamma$;
$\neg(\phi_1\vee\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ and $\neg\phi_2\in\Gamma$;

(v) $(\phi_1\rightarrow\phi_2)\in\Gamma \Rightarrow \neg\phi_1\in\Gamma$ or $\phi_2\in\Gamma$;
$\neg(\phi_1\rightarrow\phi_2)\in\Gamma \Rightarrow \phi_1\in\Gamma$ and $\neg\phi_2\in\Gamma$;

(vi) $(\exists x)\phi x\in\Gamma \Rightarrow \phi t\in\Gamma$ for some term t;
$\neg(\exists x)\phi x\in\Gamma \Rightarrow \neg\phi t\in\Gamma$ for all terms t;

(vii) $(\forall x)\phi x\in\Gamma \Rightarrow \phi t\in\Gamma$ for all terms t;
$\neg(\forall x)\phi x\in\Gamma \Rightarrow \neg\phi t\in\Gamma$ for some term t.

The definition of a structure $\mathfrak{M}$ from Γ_ρ is now utterly standard, and we obtain a completeness theorem for classical predicate logic in the following form:

Completeness Theorem. The intercalation tree for the question α;?G allows us to determine either a normal derivation G from α or a branch that provides a counterexample to the inference from α to G.

So we have a semantic argument for the normalizability of ND proofs.

Corollary (Normal Form Theorem): If G can be proved from assumptions in α, then there is a normal proof of G from α.

Remark. As in the case of sentential logic the Interpolation Theorem with its standard consequences (Beth Definability, Robinson Joint Consistency) can be obtained easily and constructively.

Let me address the question of finding proofs in mathematics – with logical *and* mathematical understanding. If one looks, as one naturally would, at Georg Polya's writings on mathematical reasoning and heuristics, one realizes very quickly that his most general strategies for argumentation

are logical ones. Quite sophisticated strategies are involved in a program, the Carnegie Mellon Proof Tutor, that searches automatically and efficiently for natural deduction proofs in sentential logic; that program was developed by Richard Scheines and myself with assistance from Jonathan Pressler and Chris Walton. [36] Presently we are extending the program to predicate logic. Though it is undoubtedly not logical formality per se that facilitates the finding of proofs, logic does help to bridge the abyss between assumptions and conclusions. It does so by suggesting *very rough structures* for arguments, that is, *logical* structures that depend solely on the syntactic form of assumptions and conclusions. This role of logic may seem modest, but it seems to be critical for penetrating to essential subject-specific considerations supporting a conclusion. It is our very ambitious goal (that will take some years of sustained work) to do automated proof search in elementary set theory, say, up to the Schröder-Bernstein Theorem; and in combinatorics, say up to van der Waerden's Theorem and other Ramsey type theorems.

So what? Let us assume *for speculation's sake* that we have written a program that develops – in a completely automatic mode and guided by intelligible heuristics (in Polya's sense) – the parts of mathematics indicated at the end of the last section. *What may we have learned?* Perhaps only that computational power is the crucial ingredient for the success of the program; but perhaps more, namely, how to make partially explicit the collective wisdom contained in the structure of mathematics.

Proofs provide explanations of what they prove by putting their conclusions in a context that shows them to be correct. The deductive organization of parts of mathematics is *the* classical methodology for specifying such contexts. This methodology has two crucial aspects: the formulation of appropriate principles and the reasoning from such principles. For foundational purposes one formulates quasi-constructive principles as *axioms* – principles that underly the "construction" of objects in the intended model – e.g., of natural numbers, sets in the cumulative hierarchy, or elements of inductively defined classes; for mathematical practice one formulates principles for concepts that characterize general structures without canonically generated elements – in order to make analogies between

[36] For details, in particular concerning heuristics, see [Sieg and Scheines 1992].

different parts of mathematics precise and to achieve generality of arguments in that way. *Reasoning* from principles is mediated through logical inferences and subject-specific lemmata. These two aspects correspond schematically to *intuition* and *ingenuity*, the two faculties Turing thought are involved in mathematical reasoning.

> The activity of the intuition consists in making spontaneous judgments which are not the result of conscious trains of reasoning. These judgments are often but by no means invariably correct ... Often it is possible to find some other way of verifying the correctness of an intuitive judgment. We may, for instance, judge that all positive integers are uniquely factorizable into primes; a detailed mathematical argument leads to the same result. This argument will also involve intuitive judgments, but they will be less open to criticism than the original judgment about factorization.[37]

Ingenuity is to aid intuition by "suitable arrangements of propositions, and perhaps geometrical figures and drawings". If the latter are arranged suitably, then "the validity of the intuitive steps which are required cannot seriously be doubted". Clearly, the role played by these faculties differs from mathematician to mathematician, from subfield to subfield. This arbitrariness, Turing believed, is removed by the introduction of a "formal logic"; formal rules (that correspond to intuitively valid inferences) reduce greatly the necessity for appealing to intuition, and the idea of ingenuity takes on a more definite shape, when we work in a formal logic:

> In general a formal logic will be framed so as to admit a considerable variety of possible steps in any stage in a proof. Ingenuity will then determine which steps are the more profitable for the purpose of proving a particular proposition.

These broad considerations are connected directly to the discussion of actual or projected computing devices in his *Lecture to the London Mathematical Society* and *Intelligent Machinery*, where Turing calls for both "intellectual searches" (i.e., heuristically guided searches) and "initiative" (that includes, in the context of mathematics, proposing intuitive steps). So Turing faces both problems: formulating heuristics with respect to a fixed search space, that is, derivations of a particular formal system, but also finding new principles. The latter problem has to be addressed since, in Turing's own phrase, the necessity for intuition cannot be entirely eliminated because of Gödel's theorems.

Indeed, in his investigation of ordinal logics, Turing was not about to formulate "ingenious" ways of finding proofs; on the contrary, ingenuity was replaced by "patience" based on the fact that the theorems of a formal logic can always be effectively enumerated and on the assumption that "all proofs take the form of a search through this enumeration for the theorem for which a proof is desired". And he focused on ways of transcending the limitations imposed by the Incompleteness Theorems. In 1947, when he was more concerned with the actual construction of computing machines, he nevertheless emphasized the shift of the theoretical issues:

> As regards mathematical philosophy, since the machines will be doing more and more mathematics themselves, the centre of gravity of the human interest will be driven further and further into philosophical questions of what can in principle be done etc.[38]

If the interpretation of the Incompleteness Theorems (seen as formulating particular answers to the question of what in principle can be done) is to be informative, the relation of Turing computability to effective calculability and the informal understanding of the latter notion must come to the fore.

Gödel's disjunct. Post emphasized in his 1936 paper that the Incompleteness and Undecidability Theorems exemplify "a fundamental discovery in the limitations of the mathematizing power of Homo Sapiens". In his 1944-paper he remarked with respect to these results:

> Like the classical unsolvability proofs, these proofs are of unsolvability by means of given instruments. What is new is that in the present case these instruments, in effect, seem to be the only instruments at man's disposal.[39]

For Gödel – in contrast to Post – the Incompleteness Theorems do *not* establish "any bounds for the powers of human reason, but rather for the potentialities of pure formalism in mathematics".[40] Turing's work provides, according to Gödel, "a precise and unquestionably adequate definition of the general concept of formal system"; consequently, the Incompleteness Theorems hold for *arbitrary formal* systems (satisfying the usual conditions). Curiously enough, in [Gödel 1972a] there is a discussion of a "philosophical error in Turing's work" that can be regarded as a footnote to the word "mathematics" in the first quotation. Gödel claims that Turing, on page 136 of [Davis 1965], gives an argument to show that "mental procedures cannot go beyond mechanical procedures". What *is* given on that page is a very brief argument showing that "the number of states of mind that need be taken into

[37] [Turing 1939], pp. 208-209.

[38] [Turing 1947], p. 122.

[39] [Post 1944], p. 310 in [Davis 1965].

[40] [Gödel 1964] , pp. 72-73.

account is finite". The context makes crystal-clear that mechanical procedures are being analyzed, and thus I cannot see a philosophical error in Turing's work; rather, I believe the error is in Gödel's interpretation.

However, the interest of Gödel's remarks in this note is quite independent of his error; they summarize points for which he had argued more extensively in his Gibbs Lecture (1951). If mathematics, Gödel stated there, is viewed as a body of propositions that "hold in an absolute sense", then the Incompleteness Theorems express the fact that mathematics is not exhaustible by a mechanical enumeration of its theorems. After all, the First Theorem yields, for any consistent formal system S containing a modicum of number theory, a simple arithmetic sentence that is independent of S. But Gödel emphasized that it is the Second Theorem that makes this phenomenon of inexhaustibility particularly evident.

> For it makes it impossible that someone should set up a certain well-defined system of axioms and rules and consistently make the following assertion about it: All of these axioms and rules I perceive (with mathematical certitude) to be correct, and moreover I believe that they contain all of mathematics.[41]

If someone claims this, he contradicts himself: Recognizing the correctness of all axioms and rules means recognizing the consistency of the system. Thus, a mathematical insight that does not follow from the axioms has been gained. To explain carefully the meaning of this situation, Gödel distinguished between "objective" and "subjective" mathematics: *Objective mathematics* consists of all true mathematical propositions; *subjective mathematics* contains all humanly provable mathematical propositions. Clearly, there cannot be a complete formal system for objective mathematics; but it is not excluded that, for mathematics in the subjective sense, there might be a finite procedure yielding all of its evident axioms (though we could never be certain that all of these axioms are correct). But if there were such a procedure, then – at least as far as mathematics is concerned – the human mind would be equivalent to a Turing machine. Furthermore, there would be simple arithmetic problems that could not be decided by any mathematical proof intelligible to the human mind. If we call such a problem *absolutely undecidable* we have established with full mathematical rigor that ***either*** *mathematics is inexhaustible in the sense that its evident axioms cannot be generated by a finite procedure* or *there are absolutely undecidable arithmetic problems*.[42]

Aspects of mathematical experience. This theorem appears to Gödel to be of "great philosophical interest". That is not surprising, since he explicates the first alternative in the following way: "... that is to say, the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine". However, if one takes seriously this reformulation, then one certainly should try to see in what ways the human mind "transcends" the limits of mechanical computors. Gödel suggested in (1972a) that there may be (humanly) effective, but non-mechanical procedures. Yet even the most specific of his proposals, Gödel admitted, "would require a substantial advance in our understanding of the basic concepts of mathematics". That proposal concerned the extension of systems of axiomatic set theory by axioms of infinity, i.e., extending segments of the cumulative hierarchy. The problem of extending what I call *accessible domains* is not special to the case of set theory; rather, there are completely analogous issues for the theory of primitive recursive functionals and for the theory of constructive ordinals in the second number class. This is the first of the two aspects of mathematical experience I want to describe briefly; as a matter of fact, both aspects are related to features of "mental procedures" Gödel discussed.

Accessible domains, constituted by inductively generated elements, are most familiar from mathematics and logic. In proof theory, for example, inductively defined higher constructive number classes have been used in consistency proofs for impredicative subsystems of analysis. These and other classes provide special cases in which generating procedures allow us to grasp the intrinsic build-up of mathematical objects. And such an understanding is a fundamental source of our knowledge of mathematical principles for the domains constituted by them; for it is the case, I suppose, that the definition and proof principles for such domains follow directly from the comprehended build-up.[43] If we understand, for example, the set-theoretic

[41] [Gödel 1951], pp. 5-6.

[42] [Gödel 1951], p. 7.

[43] A broad framework for the "inductive or rule governed generation" of mathematical objects is described in [Aczel 1977]; it is indeed so general that it encompasses not only finitary i.d. classes, higher number classes, and models of a variety of constructive theories, but also segments of the cumulative

generation procedure for a segment of the cumulative hierarchy, then it is indeed the case that the axioms of ZF^- (i.e., ZF without the postulate for the existence of, the first infinite ordinal), together with a suitable axiom of infinity, "force themselves upon us as being true" in Gödel's famous phrase; they simply formulate the principles underlying the "construction" of the objects in this segment.[44]

The sketch of this *quasi-constructive* aspect of mathematical experience is extremely schematic and yet, I think, helpful for further orientation. For Dedekind, consistency proofs were to ensure that axiomatically characterized notions (like that of a complete ordered field) were free from "internal contradictions". Here we are dealing with *abstract* notions *without* an "intended model" constituted by *inductively generated* elements.[45] And these notions are distilled from mathematical practice for the purpose of comprehending complex connections, of making analogies precise, and of obtaining a more profound understanding. It is in this way that the axiomatic method teaches us, as Bourbaki (1950) expressed it in Dedekind's spirit,

> to look for the deep-lying reasons for such a discovery [that two, or several, quite distinct theories lend each other "unexpected support"], to find the common ideas of these theories, ... to bring these ideas forward and to put them in their proper light.

Notions like group, field, topological space, and differentiable manifold are abstract in this sense and are properly investigated, i.e., in full generality, in category theory. Another example of such a notion is that of Turing's mechanical computor! Though Gödel (1972 a) uses "abstract" in a more inclusive way than I do here, it seems that the notion of computability exemplifies his broad claim "that we understand abstract terms more and more precisely as we go on using them, and that more and more abstract terms enter the sphere of our understanding". This *conceptional aspect* of mathematical experience and its profound function in mathematics have been entirely neglected in the logico-philosophical literature on the foundations of mathematics - except in the writings of Paul Bernays.

Final remarks. I argued that the sharpening of axiomatic theories to formal ones was motivated by epistemological concerns. A central point was the requirement that the checking of proofs ought to be done in a radically intersubjective way; it should involve only operations similar to those used by a computor when carrying out an arithmetic calculation. Turing analyzed the processes underlying such operations and formulated a notion of computability by means of his machines; that was in 1936. In a paper written about ten years later and entitled *Intelligent Machinery*, Turing stated what really is the central problem of cognitive psychology:

> If the untrained infant's mind is to become an intelligent one, it must acquire both discipline and initiative. So far we have been considering only discipline [via the universal machine, W.S.]. ... But discipline is certainly not enough in itself to produce intelligence. That which is required in addition we call initiative. This statement will have to serve as a definition. Our task is to discover the nature of this residue as it occurs in man, and to try and copy it in machines.

The task of copying is difficult, some would argue impossible in the case of mathematical thinking. But before we can start copying, we have to discover – at least partially – "the nature of the residue". Thus we are led back to the questions: What are essential aspects of mathematical experience? Are they mechanizable? I have tried to give a very tentative and partial answer to the first question. As far as the second question is concerned, I don't have even a conjecture on how it will be answered. Whatever the right answers may be, mathematical experience represents an extremely important component of Turing's problem, and we should investigate crucial aspects vigorously – by historical case studies, theoretical analyses, psychological experimentation and, *quite in Turing's open spirit*, by machine simulation.

hierarchy. It provides a uniform framework in which the difficulties (in our understanding) of generating procedures can be compared and explicated.

[44] There is a rich literature dealing with the "iterative conception of set" including papers by Parsons and Wang; that cannot be discussed here. For references to this literature, see the second edition of *Philosophy of Mathematics*, edited by Benacerraf and Putnam, Cambridge, 1983.

[45] The categoricity of the second-order theory of complete ordered fields does not argue against this point; as another example of a theory exhibiting similar features consider the theory of dense linear orderings without endpoints.

BIBLIOGRAPHY

This is a highly selective bibliography in which I listed sources only, when they are actually referred to in the lectures.

P. Aczel An introduction to inductive definitions; in: *Handbook of Mathematical Logic*, J. Barwise (ed.), North-Holland Publishing Company, 1977, 739-782

P. Andrews Transforming matings into natural deduction proofs; 5th Conference on automated deduction , W. Bibel and R. Kowalski (eds.), Springer-Verlag, 1980, 281-292

P. Bernays Über Hilberts Gedanken zur Grundlegung der Arithmetik; Jahresbericht DMV 31 (1922), 10-19

Hilbert, David; in: *Encyclopedia of Philosophy*, vol. 3, 1967, 496-504

Die schematische Korrespondenz und die idealisierten Strukturen; Dialectica 24 (1970), 53-66; reprinted in: P. Bernays, *Abhandlungen zur Philosophie der Mathematik*, Darmstadt 1976, 176-188

W. Bledsoe Non-resolution theorem proving; Artificial Intelligence 9 (1977), 1-36

N. Bourbaki The architecture of mathematics; Math. Monthly 57 (1950), 221-32

W. Buchholz A new system of proof-theoretical ordinal functions, Ann. Pure and Applied Logic 32 (1986), 195-207

W. Buchholz, S. Feferman, W. Pohlers, W. Sieg

Iterated Inductive Definitions and Subsystems of Analysis; Lecture Notes in Mathematics, vol. 897, Springer-Verlag, 1981

W. Buchholz and K. Schütte

Proof theory of impredicative subsystems of analysis; Bibliopolis, 1988

W. Buchholz and W. Sieg

A note on polynomial time computable arithmetic; Contemp. Math. 106 (1990), 51-56

S. Buss *Bounded Arithmetic*; Bibliopolis, 1986

A. Church An unsolvable problem of elementary number theory; Amer. J. Math. 58 (1936), 345-63

Review of [Turing 1936], J. Symbolic Logic 2 (1937), 42-3

S. Cittadini Intercalation calculus for intuitionistic propositional logic; Report 29 in Philosophy, Methodology, Logic Series, Carnegie Mellon University, 1992

M. Davis (ed.) *The Undecidable*; New York, 1965

R. Dedekind Stetigkeit und irrationale Zahlen; Braunschweig 1872

Was sind und was sollen die Zahlen; Braunschweig 1888

M. Fitting First-order logic and automated theorem proving; Springer-Verlag, 1990

S. Feferman Formal theories of transfinite iterations of generalized inductive definitions and some subsystems of analysis; in: Kino, Myhill, and Vessley (eds.), *Intuitionism and Proof Theory*, North-Holland, 1970

S. Feferman and W. Sieg

Proof theoretic equivalences between classical and constructive theories of analysis; in: [Buchholz e.a.], 78-142

F. Ferreira Polynomial time computable arithmetic; Contemp. Math. 106 (1990), 137-156

G. Frege *Grundgesetze der Arithmetik*, begirffsschriftlich abgeleitet; Jena, 1893

Nachgelassene Schriften; H. Hermes, F. Kambartel, F. Kaulbach (eds.), Hamburg, 1969

Collected papers on mathematics, logic, and philosophy; B. McGuinness (ed.), Oxford University Press, 1984

H. Friedman Iterated inductive definitions and Σ_1^2 -AC; in: Kino, Myhill, and Vessley (eds.), *Intuitionism and Proof Theory*, North-Holland, 1970, 435-42

Some systems of second order arithmetic and their use; Proc. Int. Cong. Mathematicians, vol. 1, 1975, 235-242

H. Friedman, S. Simpson, R. Smith

Countable algebra and set existence axioms; Ann. Pure and Applied Logic 25 (1983), 141-183

J. Gallier *Logic for Computer Science*, Harper and Row, 1986

R. Gandy The confluence of ideas in 1936; in: *The universal Turing machine - a half-century survey*, R. Herken (ed.), Oxford University Press, 1988, 55-111

G. Gentzen Untersuchungen über das Logische Schließen I, II; Math. Zeitschrift 39 (1934), 176-210, and (1935), 405-431

J.-Y. Girard Proof theory and logical complexity; Bibliopolis, 1987

J.-Y. Girard, Y. Lafont, P. Taylor

Proofs and types; Cambridge University Press, 1989

K. Gödel Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I, Monatshefte für Mathematik und Physik 38 (1931), 173-198

The present situation in the foundations of mathematics; Lecture delivered at the meeting of the Amer. Math. Society, Dec. 29-30, 1933

On undecidable propositions of formal mathematical systems; Lecture Notes, Princeton, 1934, reprinted in [Davis], 39-71

Über die Länge von Beweisen; Ergebnisse eines math. Kolloquiums 7 (1936), 23-24

Remarks before the Princeton bicentennial conference on problems in mathematics; 1946, reprinted in [Davis], 84-88

J. Herbrand Investigations in proof theory; 1930 , in: (1971), 44-202

Unsigned note on Herbrand's thesis, written by Herbrand himself; 1931a, in: (1971), 272-75

Note for Jacques Hadamard; 1931b, in: (1971), 277-81

On the consistency of arithmetic; 1931c, in (1971), 282-98

Logical Writings, W. Goldfarb (ed.), Cambridge, 1971

D. Hilbert Über den Zahlbegriff; Jahresberichte der Deutschen Mathematiker-Vereinigung 8 (1900), 180-194

Sur les problèmes futurs des mathématiques; Compte Rendu du Deuxième Congrès International des Mathématiciens, Paris, 1902, 59-114

Über die Grundlagen der Logik und Arithmetik; 1904, reprinted in [van Heijenoort], 129-138

Über das Unendliche; Math. Annalen 95 (1926), 161-190

D. Hilbert and P. Bernays

Die Grundlagen der Mathematik, vol. I; Springer-Verlag, 1934

Die Grundlagen der Mathematik, vol. II; Springer-Verlag, 1939

W. A. Howard

Hereditarily majorizable functionals of finite type; in: Lecture Notes in Mathematics 344; Springer-Verlag, 1973, 454-461

A. Ignjatovic Delineating classes of computational complexity via second order theories with weak set existence principles (I); to appear in J. Symbolic Logic

G. Jäger *Theories for admissible sets* - a unifying approach to proof theory; Bibliopolis, 1986

S.C. Kleene General recursive functions of natural numbers; Math. Annalen 112 (1936), 727-42

Introduction to Metamathematics; Groningen, 1952

G. Kreisel On the interpretaion of non-finitist proofs I; J. Symbolic Logic 16 (1951), 241-267

Hilbert's Programme; Dialectica 12 (1958A), 346-372

Mathematical significance of consistency proofs; J. Symbolic Logic 23 (1958B), 321-388

Survey of proof theory; J. Symbolic Logic 33 (1968), 321-388

L. Kronecker Über den Zahlbegriff; published in 1887, reprinted in *Werke*, Vol. III, Part 1, Teubner, 1899, 251-274

D. Leivant A foundational delineation of computational feasibility; manuscript, 1991

G. Mints Proof theory in the USSR: 1925-1969; J. Symbolic Logic 56 (1991), 385-424

C.D. Parsons On a number-theoretic choice schema and its relation to induction; in: Kino, Myhill, and Vessley (eds.), *Intuitionism and Proof Theory*, North-Holland, 1970, 459-473

On n-quantifier-induction; J. Symbolic Logic 36 (1972), 466-482

F. Pfenning Proof transformations in higher-order logic; Ph.D. dissertation, Carnegie Mellon University, 1987

W. Pohlers *Proof theory: an introduction*; Lecture Notes in Mathematics, vol. 1407, Springer-Verlag, 1989

E. Post Finite combinatory processes. Formulation I; J. Symbolic Logic 1 (1936), 103-5

Recursively enumerable sets of positive integers and their decision problems; Bull. Amer. Math. Soc. 50 (1944), 284-337

D. Prawitz *Natural Deduction*, A proof-theoretical study; Stockholm, 1965

M. Rathjen Proof-theoretic analysis of KPM; Arch. Math. Logic 30 (1991), 377-403

C. Reid *Hilbert*; Springer-Verlag, 1970

K. Schütte *Proof theory*; Springer-Verlag, 1977

H. Schwichtenberg

Proof-theory: some applications of cut-elimination; in: *Handbook of Mathematical Logic*, J. Barwise (ed.), North-Holland Publishing Company, 1977, 867-96

W. Sieg Inductive definitions, constructive ordinals, and normal derivations; in: [Buchholz e.a.], 143- 186

Fragments of arithmetic; Ann. Pure Appl. Logic 28 (1985), 33-71

Hilbert's program sixty years later; J. Symbolic Logic 53 (1988), 338-348

Relative consistency and accessible domains; Synthese 84 (1990), 259-297

Herbrand analyses; Arch. Math. Logic 30 (1991), 409-441

Mechanical procedures and mathematical experience; to appear in: *Mathematics and Mind* (A. George, ed.), Oxford University Press, 1992

Intercalation calculi and automated proof search; manuscript, 1992

W. Sieg and R. Scheines

Searching for proofs (in sentential logic), in: *Philosophy and the Computer*, L. Burkholder (ed.), Westview Press, 1992, 137-159

G. Stalmark Normalization theorems for full first order classical natural deduction; J. Symbolic Logic 56 (1991), 129-149

W.W. Tait Normal derivability in classical logic; in: *The syntax and semantics of infinitary languages*, J. Barwise (ed.), Lecture Notes in Mathematics 72, 1968, 204-236

G. Takeuti *Proof theory* (Second edition); North Holland, 1987

A. Turing On computable numbers, with an application to the Entscheidungsproblem; Proc. London Math. Soc. 42 (1936), 230-265; reprinted in [Davis], 116-151

Systems of logic based on ordinals; Proc. London Math. Soc. 45 (1939), 161-228; eprinted in [Davis], 155-222

Lecture to London Mathematical Society on 20 February 1947; in: A.M. Turing's ACE report of 1946 and other papers, B.E. Carpenter and R.W. Doran (eds.), Cambridge (Mass.) 1986, 106-124

Intelligent machinery; written in September 1947, submitted to the National Physical Laboratory in 1948, and reprinted in: *Machine Intelligence 5*, Edingburgh, 3-23

J. van Heijenoort

From Frege to Gödel; Cambridge (Mass.), 1967

Selected essays; Bibliopolis, 1985

J. von Neumann

Zur Hilbertschen Beweistheorie; Math. Zeitschrift 26 (1927), 1-46

Hao Wang Toward mechanical mathematics; IBM Journal for Research and Development 4 (1960); reprinted in: *A survey of mathematical logic*, Peking and Amsterdam, 1963, 224-268

A. Weiermann

Vereinfachte Kollabierungsfunktionen und ihre Anwendungen; Arch. Math. Logic 31 (1991), 85-94

A.N. Whitehead and B. Russell

Principia Mathematica, vol. 1; Cambridge University Press, 1910

———, vol. 2, 1912

———, vol. 3, 1913