Mechanisms and Search:

Aspects of Proof Theory

by

Wilfried Sieg

June 1992

Report CMU-PHIL-30

Carnegie Mellon
Philosophy
Methodology
Logic

Pittsburgh, Pennsylvania 15213-3890

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, 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

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 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.[2] 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

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

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

---

[2] 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.

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 $\wedge, \vee, \exists, \forall$. The connectives $\rightarrow, \leftrightarrow$, 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.

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**

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

functions that can be proved to exist in the theory. For example, (Π^0_∞-**CA**) denotes the theory obtained from (**BT**) by adding the comprehension principle for all formulas in Π^0_∞ and the full induction schema; (Π^0_∞-**CA**)$\upharpoonright$ or "restricted-(Π^0_∞-**CA**)" is the corresponding theory with the induction axiom. Clearly, (Π^0_∞-**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: (Π^0_∞-**CA**)$\upharpoonright$ is a conservative extension of elementary number theory (**Z**), whereas (Π^0_∞-**CA**) proves the consistency of (**Z**).

There is one very weak system we shall consider: It was introduced by Friedman and is labelled (**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:

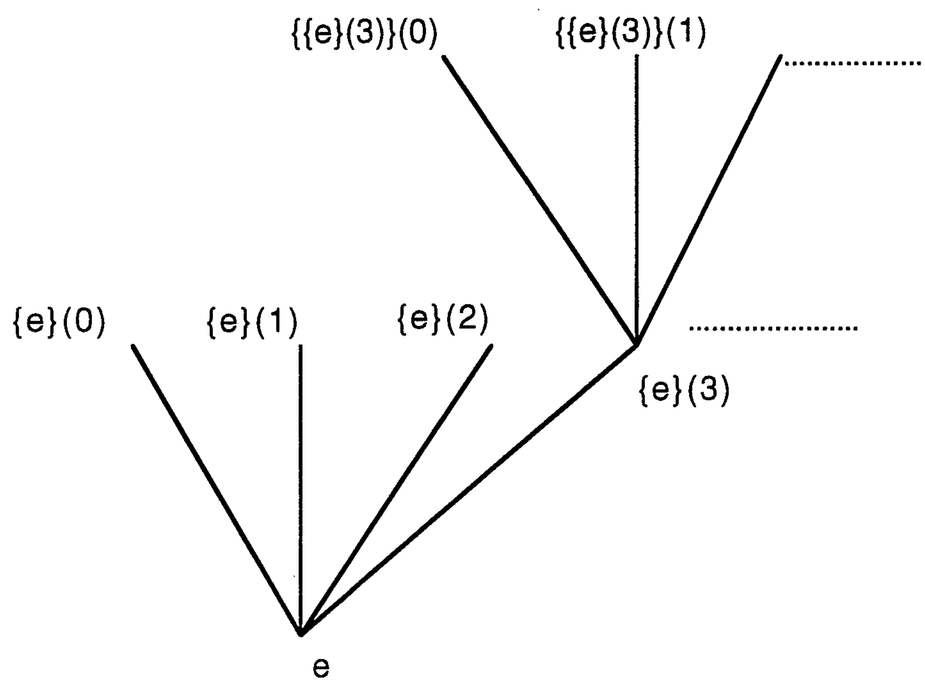
$$(\forall f)[T(f) \wedge (\forall x)(\exists y)(lh(y)=x \wedge f(y)=1) \rightarrow (\exists g)(\forall x)\, f(\bar{g}(x))=1];$$

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

$$(\forall x)(\forall y)\, [(f(x{*}y)=1 \rightarrow f(x)=1) \wedge (f(x{*}\langle y\rangle)=1 \rightarrow y\leq 1)]$$

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



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_1^0 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.

analysis beyond (Δ^1_2-**CA**). There are results for (Δ^1_2-**CA+BI**): Jäger and Pohlers determined the proof theoretic ordinal of the theory, and Jäger reduced it to 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.

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.[1]

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."[2] 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

---

[1] [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.

[2] [Gödel 1933], p. 1.

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 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[6], 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

---

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

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 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-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 9889995496789998769 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:

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 lecture notes, i.e. without reference to computability."[11] 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 $\phi_n(x_1, x_2, ..., x_{pn})$ of a certain finite or infinite set F. Each $\phi_n(x_1, ...)$ 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 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[12] "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

---

11 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].

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

With van Heijenoort I assume that, here too, Herbrand used "intuitionistic" as synonymous with "finitist".[14] 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[15], 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 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

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

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

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 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.[1] As Tr is provably adequate we have

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

and, thus,

---

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

(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 $\Delta(s)$; $|D(a)|=|D(s)|$; $\rho(D(a))=\rho(D(s))$.

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

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 $\wedge$-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

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

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 $[(\phi \text{ is } \psi)$ or $(\psi \text{ is } \neg\xi, \xi$ is atomic and ϕ is $\xi)$ 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 $\mathbf{T}$ be a set of sequents whose elements are literals (i.e., either atoms or negated atoms); if $\mathbf{T}$ is closed under substitution[3] it is called a *Post System.* Let me describe some examples:

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

(EA$_1$) $\qquad \Gamma, t{=}t$

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

derivation and thus of $\mathfrak{T}(\mathfrak{F})\vdash D$. (Clearly, the principal formulas of axioms Γ,Δ are the elements of Δ.)

Theorem ($\mathfrak{T}$-normalization). Let D be a $\mathfrak{T}(\mathfrak{F})$-derivation of Γ; then there is a quasi-normal $\mathfrak{T}(\mathfrak{F})$-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 $\mathfrak{T}(\mathfrak{F})$-axioms are involved. If one of the axioms is a logical one then Γ_0, Γ_1 must be a $\mathfrak{T}(\mathfrak{F})$-axiom; if both are $\mathfrak{T}(\mathfrak{F})$-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 $\mathfrak{T}(\mathfrak{F})$. 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 $\mathfrak{T}(\mathfrak{F})$-derivation of Γ, then every formula in D is either a subformula of some element in Γ or of some $\mathfrak{T}(\mathfrak{F})$-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][4]: 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 $\Delta_0, \Sigma_n^0, \Pi_n^0$. The theory obtained from an extension of

[4] section 2.7.7 on pag.125

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:

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

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

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 $\mathfrak{T}(\mathfrak{F})$ of the form (QF($\mathfrak{F}$)-IA) such that $\mathfrak{T}(\mathfrak{F})$ and $\mathfrak{F}$ satisfy the following two conditions:

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

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

Theories $\mathfrak{T}(\mathfrak{F})$ 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 $\mathfrak{T}(\mathfrak{F})$ be an Herbrand theory, let Γ contain only purely existential formulas, and let ψ be quantifier-free; if D is a $\mathfrak{T}(\mathfrak{F})$-derivation of $\Gamma,(\exists x)\psi x$, then there is a term t^* and a(n I-normal) $\mathfrak{T}(\mathfrak{F})$-derivation D^* of $\Gamma,\psi t^*$.

Proof (by induction on I-normal $\mathfrak{T}(\mathfrak{F})$-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) $\Gamma,\phi 0,\psi r$

and of

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

$\mathfrak{T}(\mathfrak{F})$ 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) $\neg\phi 0,\phi t,\psi h(t)$

and

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

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

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

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_1^0-induction-principle, and that bounded iteration is exactly right for analyzing s-Σ_1^b-induction. As consequences we obtain very neat proofs of two facts: (1) the provably total functions of (Σ_1^0-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_2^1 are exactly the polynomial-time computable ones. Let me start out with the considerations for the former result.

Lemma. Let Γ contain only Σ_1^0-formulas; if D is an I-normal derivation of Γ in (Σ_1^0(**PR**)-IA), then there is an I-normal derivation of Γ in (QF(**PR**)-IA).

Proof. The argument proceeds by induction on the number # of applications of theΣ_1^0-induction rule in D. Clearly, if #=0, the claim is trivial. So assume that #>0 and consider an application of the Σ_1^0-induction rule such that no other application occurs above it in D. The subderivation E determined in this way ends with the inference

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

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_1^0 or Σ_1^0; 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 bounded arithmetic, is the language of elementary arithmetic expanded by function symbols $|.|$, $\dot{\div}$, and #, where $|a|$ yields the length of the binary representation of a, $\dot{\div}$ is the shift-right-function, and a#b is $2^{|a||b|}$. The language $\mathcal{L}(\mathbf{P})$ is obtained from $\mathcal{L}(\mathbf{B})$ by adding function symbols for each element of $\mathbf{P}$. The latter class of functions is defined inductively as the smallest class of functions that contains certain initial functions (0, ', $\dot{\div}$, 2. , χ, *choice*[1]) 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[2]) iff the following holds: If τ is defined by

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

then we must have

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

and

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

x indicates a sequence of variables. -- Letting $\mathbf{F}$ stand for $\mathbf{P}$ or $\mathbf{B}$, the set of quantifier-free formulas in $\mathcal{L}(\mathbf{F})$ is denoted by QF($\mathbf{F}$). The bounded quantifiers $(\forall x \leq |t|)$ and $(\exists x \leq |t|)$, understood again as abbreviations, are called *sharply bounded*. $\Delta_0^b(\mathbf{F})$, the class of sharply bounded formulas, is built up from literals in $\mathcal{L}(\mathbf{F})$ using $\wedge$, $\vee$, and sharply bounded quantifiers; if closure under bounded existential quantification is also required, the set of formulas is called $\Sigma_1^b(\mathbf{F})$. A formula of $\mathcal{L}(\mathbf{F})$ is in s-$\Sigma_1^b(\mathbf{F})$ just in case it is of the form $(\exists x \leq t)\phi$, where ϕ is in QF($\mathbf{F}$). The theories of bounded arithmetic to be investigated contain the basic axioms for the non-logical symbols of $\mathcal{L}(\mathbf{B})$, the defining equations for the elements of $\mathbf{P}$ in case the theory is formulated in $\mathcal{L}(\mathbf{P})$, 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

---

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

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

We only have to establish that the theory (s-$\Sigma_1^b(\mathbf{P})$-LIND) is conservative over (QF($\mathbf{P}$)-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-$\Sigma_1^b(\mathbf{P})$-LIND), then there is an I-normal derivation of Γ in (QF($\mathbf{P}$)-LIND).

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

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

$\psi\mathbf{a}a$ is of the form $(\exists x)(x\leq t[\mathbf{a},a] \wedge \psi^*x\mathbf{a}a)$, where ψ^* is in QF($\mathbf{P}$) 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[3] to extract from $\mathbf{E}_0$ a term $\sigma[\mathbf{a}]$ and a derivation in (QF($\mathbf{P}$)-LIND) of

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

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

Now define:

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

ρ can be shown to be in $\mathbf{P}$. 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

---

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

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 $\mathfrak{K}$.

QF-AC$_0$-elimination. If D is an I-normal $\mathfrak{K}$-derivation of $\Delta[\neg\text{QF-AC}_0]$, then there is an I-normal $\mathfrak{K}$-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 $\mathfrak{K}$-derivation of $\Delta[\neg\text{WKL}]$, then there is an I-normal $\mathfrak{K}$-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 $I(\Sigma^0_1)$-normal $\mathfrak{K}$-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 $I(\Sigma^0_1)$-normal $\mathfrak{K}$-derivation of $\Delta[\neg\text{QF-AC}_0,\neg\text{WKL}],(\exists y)\psi ay$, then there is an I-normal $\mathfrak{K}$-derivation E of $\Delta,(\exists y)\psi ay$.

Proof of (*) (proceeds by induction on the length of $I(\Sigma^0_1)$-normal $\mathfrak{K}$-derivations). The induction step is trivial in case of <u>LA</u>, <u>C</u> 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 <u>C</u> 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

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_1^0-induction rule, we obtain an $I(\Sigma_1^0)$-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_1^0-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 ($\mathbf{BT}$+Σ_1^0-IA+Σ_1^0-AC_0+WKL) over ($\mathbf{BT}$) with respect to Π_2^0-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. $|D_i'| \leq D_i$, for $i \leq 1$, 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) $\Delta, \neg(\forall x)\psi xu(x)$.

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 Δ, $f(t[s^*])\neq 1$ when taking into account the equation $\overline{u^*}(s^*)=t[s^*]$. From F_1 one can obtain a derivation G_1 of Δ, $f(t[s^*])=1$ by $\triangle$-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**).

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.[1] 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

[1] 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.

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.[3] 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[4] 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 <u>through a calculus</u> normal derivations? (2) How can one construct a search space that allows the

[3] 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).

[4] 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.

predicate logic and to non-classical logics, for example, intuitionistic logic.[5] (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 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** $\alpha;\beta?\phi_2$

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

---

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

As an example of how the intercalation rules are used to build up the search space for a question α;?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 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 $\neg$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 $\neg$P are interchanged; finally, the derivation that emerges from the undrawn part at node 1 is this:

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.

so G^* must be $\bot$! But then the construction is terminated, because $\bot_{\mathcal{F}}$ 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 α;?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 α;?G evaluates as **F**, then so does α,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 α;?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 $\mathcal{F}(\alpha*\langle G''\rangle)$ 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 applicable. Application of that rule with any formula in $\mathfrak{F}(\alpha*<G''>)$, in particular with H_0, leads to the canonical closing indicated in the diagram.

Let $\Gamma := \{\phi \mid \phi \in \alpha_{v+1}\}$; 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 $\alpha*<G''>$. We establish first the closure properties.

Closure lemma. For all subformulas ϕ_1, ϕ_2 of $\alpha*<G''>$ 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 :

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 $\alpha;?G$ 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.

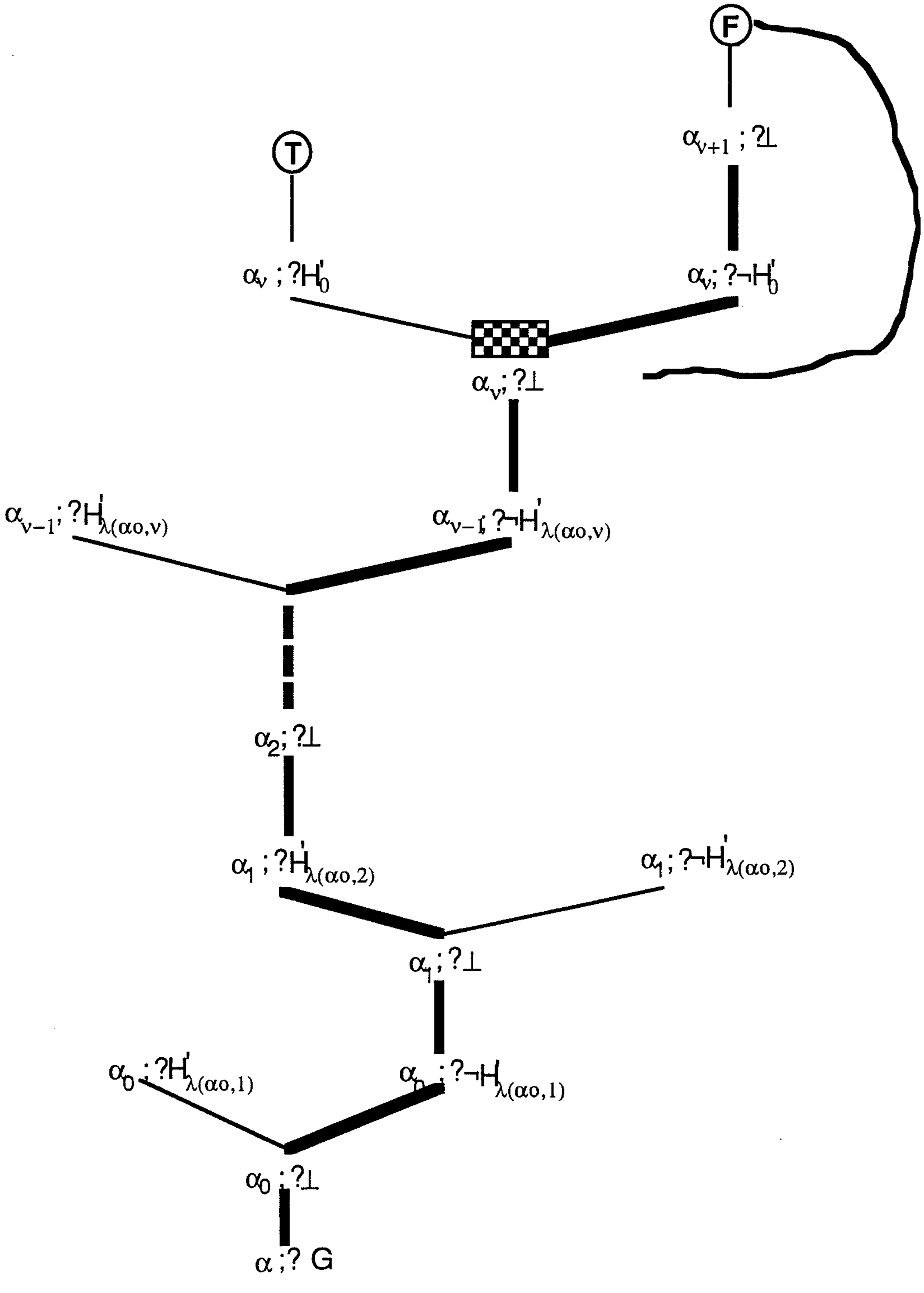

T
F
α_{v+1} ; ?⊥
α_v ; ?H'_0
α_v; ?¬H'_0
α_v; ?⊥
α_{v−1}; ?H'_{λ(αo,v)}
α_{v−1}; ?¬H'_{λ(αo,v)}
α_2; ?⊥
α_1 ; ?H'_{λ(αo,2)}
α_1 ; ?¬H'_{λ(αo,2)}
α_1; ?⊥
α_0 ; ?H'_{λ(αo,1)}
α_0 ; ?¬H_{λ(αo,1)}
α_0 ; ?⊥
α ; ? G

Diagram 2

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

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

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

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 $\bot_{\mathcal{F}}$ with $\mathcal{F}$ containing all proper subformulas of α^* (where subformulas of quantified formulas are taken only with terms in $\mathbf{T}(\alpha^*,\bot)$). 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:

$[N]_{\sigma^*}$	=	v(N)	if N is a leaf of σ^*
$[N]_{\sigma^*}$	=	$[M]_{\sigma^*}$	if M is the unique predecessor of N

in case N is at a conjunctive branching,

$[N]_{\sigma^*}$	=	**T**	if for all immediate predecessors M of N: $[M]_{\sigma^*}$=**T**
		F	if for some immediate predecesor M of N: $[M]_{\sigma^*}$=**F**
		O	otherwise

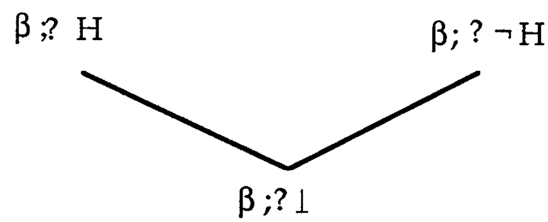
in case N is at a disjunctive branching,

$[N]_{\sigma^*}$	=	**F**	if for all immediate predecessors M of N: $[M]_{\sigma^*}$=**F**
		T	if for some immediate predecesor M of N: $[M]_{\sigma^*}$=**T**
		O	otherwise

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 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 $\alpha*\langle G''\rangle$ (where immediate subformulas of quantified formulas are taken only with terms in $\mathbf{T}(\alpha*\langle G''\rangle,\bot)$:

$$\tau(0) = \alpha;?G$$
$$\alpha_0 = \alpha*\langle G''\rangle$$
$$\lambda(\alpha_0,1) = \kappa(\alpha_0,1)$$
$$\tau(1) = \alpha_0;?\bot$$

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



if both questions $\beta;?H$ and $\beta;?\neg H$ 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 $\beta;?\bot$ 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

$(\exists x)\phi x \in \Gamma \Rightarrow \phi t\in\Gamma$ for some term t

and

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

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 definition of a structure $\mathcal{M}$ from Γ_p 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. [1] 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.

---

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

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.[3]

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.

---

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

> 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.[6]

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.*[7]

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

6 [Gödel 1951], pp. 5-6.

7 [Gödel 1951], p. 7.

"intended model" constituted by *inductively generated* elements.[10] 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

---

[10] 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

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

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