Lectures on Proof Theory

W. W. Tait

[§3. of these lecture notes, on Derivability of induction, is defective. I hope to revise it soon.]

Chapter 1

History

Proof theory was created early in the 20th century by David Hilbert to prove the consistency of the ordinary methods of reasoning used in mathematics—in arithmetic (number theory), analysis and set theory. Already in his famous "Mathematical problems" of 1900 [Hilbert, 1900] he raised, as the second problem, that of proving the consistency of the arithmetic of the real numbers. In 1904, in "On the foundations of logic and arithmetic" [Hilbert, 1905], he for the first time initiated his own program for proving consistency.

1.1 Consistency

Whence his concern for consistency? The history of the concept of consistency in mathematics has yet to be written; but there are some things we can mention.

There is of course a long tradition of skepticism. For example Descartes considered the idea we are deceived by a malicious god and that the simplest arithmetical truths might be inconsistent—but that was simply an empty skepticism. On the other hand, as we now know, in view of Gödel's incompleteness theorems, there is no relevant sense in which we can refute it.

Surely, if we are to judge by how long it took for the various successive extensions of the number system—zero and the negative numbers, and the complex numbers—to be accepted throughout the mathematical community, we have to believe that they, each in their turn, raised concerns about consistency (though at least some of the resistance to them simply resulted from the lack of immediate empirical meaning).

Computing with infinite series and with infinitesimals, for example, led to apparent inconsistencies, that called for the rigorous foundation of analysis which was achieved in the nineteenth century. Also, the problem concerning the independence of the parallel postulate in geometry, an ancient problem, was one of consistency: whether the denial of the postulate is consistent with the remaining axioms. In the eighteenth century, the investigation of this problem was intense and the attempts to prove that the denial is inconsistent finally led to the realization that it is consistent, i.e. to the development of non-Euclidean geometries.

But it seems clear that the primary impetus, if only in the psychological sense, for Hilbert's concern for the consistency problem arose from Cantor's introduction of the transfinite numbers in his *Foundations of a General Theory of Manifolds: a Mathematico-Philosophical Investigation into the Theory of the Infinite* [1883]. For the transfinite numbers are introduced essentially by the condition that, given any transitive set M of numbers, we may introduce the number $S(M)$. Define the relation $\prec$ on the transfinite numers by

$$\alpha \prec S(M) \longleftrightarrow \alpha \in M$$

Cantor regarded it as implicit in his definition of the new numbers that any sequence

$$\alpha_0 \succ \alpha_1 \succ \alpha_2 \succ \ldots$$

is finite. So, given the null set $\emptyset$, $S(\emptyset)$ is the least number 0. And when α is a number, $S(\{\alpha\})$ is the least number greater than α, i.e. $S(\alpha) = \alpha + 1$.

But there is a problem with Cantor's application of the notion of set here. Up to this point in history, the notion of a set had been clarified, primarily by Bolzano and Cantor, himself. But *this* notion of set applied to sets of elements of some *given* domain—sets of real numbers or sets of points in Euclidean space, say. And in this case, by a set of elements of a domain D was meant the extension of any property or concept defined on D. But in the case of the theory of tranfinite numbers, no domain D is given in advance from which the numbers are being picked out using the notion 'set of elements of D. Rather, a new 'domain'—call it Ω—is being *created* by stating the means for constructing its elements: namely, if M is a set of elements of Ω, then $S(M)$ is an element of Ω. Here the definition of the domain already involves the notion of a set of elements *of that very same domain.* That the old notion of set is no longer applicable in this case is a consequence of the following:

Theorem 1 *Ω is not a set.*

For, if it were, then $\alpha = S(\Omega)$ would be a number and so $\alpha \succ \alpha \succ \ldots$ would be an infinite decreasing sequence.

So, in contrast to domains such as the domain of points in Euclidean space or the real numbers, it is no longer the case that, for the domain Ω, we may take the extension of any concept defined on the domain to be a set. Indeed, we have to exclude as a set the extension of the concept '$x \in \Omega$'. This fact, which incidentally Cantor fully understood, has led to one of the dominant themes in the foundations of mathematic today: the investigation axioms of infinity or, as they are also called, large cardinal axioms. It is precisely the question of what sets of numbers exists—or, equivalently, of what numbers exist.

Failure to understand the peculiar nature of Ω led to the 'paradox of the greatest ordinal' and the 'paradox of the greatest cardinal'; although the proof of the above theorem, surely known to Cantor, yields a simpler proof of the inconsistency of assuming that Ω is a set. (And, if we assume that every totality equipollent to a set is a set, then the inconsistency of the assumption that the cardinal numbers constitute a set follows.) Frege, although he seems to have read Cantors *Foundations* and to have accepted his transfinite numbers, did fail to understand this point. He in effect took the totality of all objects as a domain and assumed that the extension of any concept defined on this domain is a set (and so in the domain). The contradiction that Russell (and Zermelo) found in Frege's system, however, did not depend upon accepting the transfinite numbers as objects; rather it depended on a later result of Cantor, namely that no set M can be mapped onto the set of subsets of M—and consequently no set can contain all of its own subsets.

Remark 1 *I believe that what further has to be understood, in order to make sense of these 'paradoxes' is that the notion of a transfinite number or, equivalently, of a set of transfinite numbers is an essentially open-ended notion: no matter what principles we introduce to construct sets of numbers, providing only that these principles are well-defined, we should be able to admit all numbers obtained by these principles as forming a set, and then proceed on to construct new numbers. So Ω cannot be regarded as a well-defined extension: we can only reason about it* intensionally*, in terms of those principles for constructing numbers that we have already admitted, leaving open in our*

reasoning the possibility—in fact, the necessity*—of always new principles for constructing numbers. When this is not understood and* Ω *is counted as a domain in the sense of a well-defined extension, then the so-called paradoxes force on us a partitioning of well-defined extensions into two categories: sets and proper classes; and the only explanation of why such an extension should be a proper class rather than a set would seem to be simply that the assumption that it is a set leads to contradiction. The paradoxes deserve the name "paradox" on this view because it offers no explanation of why there should be this dichotomy of well-defined extensions.*

It is not clear that, at the end of the day, Cantor himself entirely understood the situation. In the *Foundations* itself, realizing that his original conception of set was no longer adequate, he offers a new explanation of the notion which he clarifies in later correspondence by drawing the distinction between set and proper class simply as the distinction between consistent and inconsistent multiplicities. But this casts some doubt on whether he understood the essential open-endedness of the concept of a set of numbers. The notion of a consistent multiplicity is relative: Thus relative to the usual second-order axiom system for set theory, Morse-Kelly set theory, MKC, or its first order version, Zermelo-Fraenkel set theory, ZFC, the multiplicity of (strongly) inaccessible cardinals is consistent—i.e. if we add to MKC the axiom that there is a set M consisting of all inaccessible cardinals, then this theory is consistent if MKC is. For MKC implies that the set of sets of rank less than the least inaccessible cardinal, if there is one, form a model of MKC; and in this model $M = \emptyset$. On the other hand, a very reasonable development of the open-ended notion of number leads to the axiom that there are as many inaccessible cardinals as there are numbers. MKC together with this axiom implies that the totality of inaccessible cardinals is not a set. Ω or any other proper class fails to be a set, not because the assumption that it is a set is inconsistent, but because it is incompletely defined: in the case of Ω, anything that we admit as a set of numbers leads to bigger numbers.

The upshot of this discussion is that the paradoxes of set theory give no grounds for doubting the consistency of the theory of transfinite numbers or its embodiment in ordinary set theory. By a *well-founded* set, we mean a set s such that every sequence $s = s_0 \ni s_1 \ni s_2 \ni \ldots$ is finite. Call s a *pure* set iff every such sequence ends with a set s_n, possibly the null set. Define the *rank* of a well-founded pure set to the least number greater than the ranks of all of its elements. Let $R(\alpha)$ denote the set of all pure sets of rank $< \alpha$. Another

way of putting it is that $R(\alpha)$ is the result $P^\alpha(\emptyset)$ of iterating the PowerSet operation $s \mapsto P(s)$ α times, starting with the null set $\emptyset$. Then ordinary set theory is a theory of pure well-founded sets and its intended models are structures of the form $\langle R(\kappa), \in \rangle$, where the numbers κ will depend upon the particular axioms included in the theory. There is no appeal here to the essentially incomplete Ω or, correspondingly, the essentially incomplete totality of all pure well-founded sets.

Nevertheless, it seems that the paradoxes affected the thinking of many mathematicians at the turn of the century, including Dedekind, Poincaré and Hilbert, causing them to be concerned about methods of reasoning they had adopted which, although 'set theoretic' in one sense, have nothing to do with the source of the paradoxes. A clear statement of some of these concerns is given in a somewhat later paper by Hermann Weyl "On the new foundational crisis of mathematics" [Weyl, 1921]. Two of the main issues of concern, the Axiom of Choice and the Law of Excluded Middle, we will discuss later on. Let me comment briefly on a third issue, the PowerSet Axiom, which asserts the existence of the set $P(s)$ for any set s. We can code real and complex numbers as sets of finite ordinals, complex-valued functions of n complex variables as sets of ordered n+1-tuples of real or complex numbers, etc. So, all of analysis concerns only $P^n(\omega)$ for some small n. Now the process of passing from a domain M to the totality of all subsets of M involves only the notion of set already clarified by Bolzano and Cantor, not the open-ended notion of set involved in the theory of transfinite numbers. But, nevertheless, there is an element of open-endedness even in this notion. By a set of elements of M we mean the extension of some concept defined on M. But how may this concept be defined? It is clear that we cannot restrict ourselves to concepts defined by formulas of some particular formal language. For example, if M is infinite, then even if we admit names of elements of M in the definition, there are no more sets defined by formulas in the language than there are elements of M; but by Cantor's theorem, mentioned above, there are more subsets of M than there are elements of M. Now a very reasonable assumption is that every definable subset of M, where M is a set of sets, is defined by a formula in the language of set theory but in which names of ordinal numbers may occur. Such sets are called *ordinal definable* sets. But notice here that the totality of formulas of this kind is itself open-ended, since the totality of all ordinal numbers is. On the other hand, this only shows the open-endedness of the class of *definitions* of subsets of M, not the open-endedness of the class of its subsets. Indeed, in set theory *without* the PowerSet Axiom, we can

prove, given M, that there is an ordinal κ such that every ordinal definable subset of M is definable by means of ordinals less than κ.

My aim here is not to try to clarify this issue: I only want to present the historical context for the development of proof theory. But I do want to note that, contrary to what Weyl wrote in his paper, the so-called 'antinomies', as he called them, are not obviously a concern for analysis or, indeed, for the theory of $R(\kappa)$ for any fixed number κ.

On the other hand, both for Dedekind and Hilbert, one consequence of the so-called paradoxes of set theory was that there is no proof that there are infinite sets. Frege's assumption, in effect, that the universe of all objects is a set, and Dedekind's assumption that the objects of his thoughts form a set are both inconsistent. Now, given any infinite set, it is possible to construct a representation of the system of natural numbers; but without some given infinite set, there was no guarantee of the consistency even of elementary arithmetic. It seems to be this aspect of the paradoxes that most concerned Hilbert.

In any case, Hilbert's concern was to prove the consistency of ordinary mathematics. Prior to the late nineteenth century, the only method available for proving consistency of a theory was that interpreting the theory in another one, so that the axioms of the one theory turn out to be theorems of the other. For example, the theory of rational numbers can be interpreted in the theory of natural numbers (i.e. finite ordinals) by representing rational numbers $(k-m)/n+1$ by triples (k,m,n) of natural numbers. For example, equality of triples is defined by

$$(k,m,n) \equiv (k',m',n') \;:=\; k(n'+1)+m'(n+1) = k'(n+1)+m(n'+1)$$

and the operations of addition, subtraction, multiplication and division are defined in an obvious way which respects the equality relation. (Equals added to, subtracted from, multiplied by, or divided by equals are equal.) The ordering relation on triples likewise is defined by

$$(k,m,n) < (k',m',n') \;:=\; k(n'+1)+m'(n+1) < k'(n+1)+m(n'+1)$$

In this way, every axiom of the ordered field of rational numbers becomes a provable sentence about triples of numbers in Peano Arithmetic PA—indeed, simply about numbers, since pairs, triples, etc., of numbers can be suitably coded in this theory by numbers.

In a similar way, various non-Euclidean geometries can be proved consistent by interpreting them in Euclidean geometry. Three dimensional Euclidean geometry, in turn, can be interpreted in the theory of the ordered field of real numbers in a familiar way by interpreting points as triples of real numbers. Finally, using Dedekind's or Cantor's construction, for example, the theory of the ordered field of real numbers can be interpreted in the second order theory of the ordered field of rational numbers. So, ultimately, all of these theories can be interpreted in second order Peano Arithmetic, PA^2. Of course, this does not prove the consistency of these theories absolutely; it only proves them consistent relative to PA^2. I.e., if this latter theory is consistent, then they all are. But how do we prove PA^2 consistent? Or even just PA? Clearly the method of interpretation won't work. In order to interpret PA in a theory, the theory must imply the existence of an infinite number of elements. But what more convincing theory of this kind can we find than PA itself? By an analysis due to Dedekind [Dedekind, 1887], PA^2 can be interpreted in second order predicate logic with a unary function constant f and the nonlogical axioms

$$\exists x \forall y (f(y) \neq x)$$

$$\forall xy (f(x) = f(y) \longrightarrow x = y)$$

But this theory still implies the existence of an infinite set. A new idea is needed to prove consistency of PA.

This forms the background of Hilbert's program for proving consistency. Perhaps the best short account of it can be found in his paper "On the infinite" [Hilbert, 1926]. A more extended discussion is in [**?**; **?**].

1.2 Finitist Proof Theory

His idea was that, no matter how transfinite the concepts and methods of proof may be in a theory such as PA or even set theory, as we ordinarily understand them, our *uses* of them, our propositions and proofs, are necessarily finite. All that is needed is to provide a precisely defined language in which the concepts in question can be expressed by means of formulas and a precisely specified system of rules of inference for these formulas which adequately express what we take to be the valid inferences concerning these concepts. In other words, his idea was to represent the nonfinitist concepts and methods of proof in a formal axiomatic theory.

There are two ingredients needed for this: first, the axiomatization of the mathematical ideas in the theory, so that everything assumed in the theory about the primitive concepts and objects are explicitly stated in the axioms. This was of course an old idea, of course; but the kind of rigorous axiomatization needed to implement Hilbert's program was of much more recent origin. In fact, the most impressive example of it was his own *Foundations of Geometry* [Hilbert, 1899]. The second ingredient is the explicit statement of the principles of logic to be used in deriving theorems from the axioms, so that logic itself could be axiomatized. In this case, too, the timing was just right: the analysis of logic in Frege's *Begriffsschrift* was just what was required. (There was a remarkable meeting here of supply and demand.)

Now, having in this way completely formalized the mathematical theory, notions such as proposition, inference, proof become purely syntactical notions, referring to configurations of symbols. Propositions are represented by configurations of a certain kind called a *formulas*. A formal proof or, as we shall call it, a *deduction* is just a sequence of formulas where each formula is either in a given list of formulas, called *axioms*, or is obtained from earlier formulas by one of a list of rules for transforming formulas, called *rules of inference*. In this framework, the assertion that the original theory is consistent is the purely syntactical statement that there is no such sequence of configurations of symbols, i.e. there is no such deduction, ending with a configuration of the form $A \wedge \neg A$. The syntactical objects and concepts in question are very simple, comparable to the objects and concepts of elementary number theory. Indeed, as we now know from [Gödel, 1931], the syntactical objects can be coded by numbers in such a way that the concepts in question turn out to be very elementary arithmetic concepts.

Having transformed the problem of consistency into a problem of elementary syntax or arithmetic, the program was to prove, for given mathematical theories, the syntactical or arithmetical statement that no contradiction is derivable. But, of course, the question now is: what methods are to be used in this proof of consistency? After all, the statement that the theory is consistent, though a very elementary syntactical proposition, is still a non-trivial mathematical statement, saying (in its arithmetic form) that no number has a certain elementary property. We must use mathematics to prove it. The question is: what mathematics? For example, the simplest proof of consistency of PA would be to note that the axioms are all true on the intended interpretation in the system of natural numbers and that the rules of inference preserve truth. So, since $A \wedge \neg A$ is never true, it cannot be deducible;

hence PA is consistent. But this proof is frivolous: it uses precisely the methods of proof that we are trying to prove consistent. For example, if PA were inconsistent, then we would be able to prove in it anything, including the arithmetic statement that expresses its consistency. (Of course, as we know from Gödel's second incompleteness theorem [Gödel, 1931], the converse is also true: if we could prove in PA its consistency, then it would be inconsistent. But that is getting ahead of our story.)

Hilbert's answer to this question was to require the methods used in proving consistency to be restricted to what he called *finitist* mathematics. The objects of finitist mathematics are to be finite combinations of symbols. These can be coded by natural numbers and, conversely, numbers may be regarded as combinations of symbols—for example the expressions $S(\cdots S(0)\cdots)$ with 0 or more S's, representing the result of applying the successor operation S zero or more times to 0. So we lose nothing by considering just the case of numbers. The properties and relations which are taken to be finitistically meaningful must have the property that there is an algorithm for deciding whether or not they hold for given arguments. Thus, the Law of Excluded Middle holds for finitist propositions, not as a logical assumption, but as a requirement on what we take to be a meaningful proposition. The idea of finitism, that mathematics is concerned essentially just with the natural numbers and that the only properties and relationships that should be admitted are those which come equipped with an algorithm for deciding in each case whether or not they apply, goes back to Kronecker. But there is this difference between Kronecker and Hilbert. For the former, mathematics is to be restricted entirely to finitist mathematics. Kronecker had to be completely aware of the fact that he was rejecting a large part of the mathematics of his time and in particular, a large part of analysis. Hilbert, on the other hand, was not willing to give up this mathematics. His idea was to justify it by proving, using only finitist methods, that it is consistent.

Hilbert was never completely explicit about what he meant by finitist mathematics (nor was Kronecker). The fact is that he seemed to be confident that any mathematical problem with an elementary formulation had an equally elementary solution. Thus, the statement of consistency of PA or of ZFC is elementary and so, if true, should have an elementary proof. One lesson of Gödel's incompleteness theorems is that this is false. But Hilbert, not knowing this, was probably not concerned with spelling out exactly what finitist reasoning was. Once consistency was proved, one could simply look

at what was needed in the proof.

But there is nevertheless a question of how one could spell out the idea of finitism: we are to deal only with finite things, but we have to prove general propositions, such as the statement of the consistency of some system such as PA. But this proposition, $\forall x \neg P(g, x)$, stating that no number x is a formal deduction in PA of the sentence $0 = 1$, say, (where g is the Gödel number of this sentence), has an infinite number of instances $P(g, n)$ for each number n. Such propositions are admitted in finitist mathematics only if $P(a, b)$ is a relation for which we have an algorithm for determining its truth value for all values of its free variables a and b. But that assertion itself is infinitary in the sense that it refers to an infinite system of values of the free variable. Moreover, by asserting the general proposition, we are asserting that computing out the truth value of $P(g, n)$ for each n will yield only truth, again an assertion about the infinite totality of numbers. How are we to respect the 'finite' in finitism in view of this?

1.3 Primitive Recursive Arithmetic

There is a particular analysis of the notion of finitism which seems to be in accord with Hilbert's discussion of finitist methods and in particular, of the notion of a proof of something for an 'arbitrary' number as a schematic proof for all numbers. Let the free variable b stand for an arbitrary number. So by iterating the procedure of subtracting 1, we have the sequence $b \cdots 1, 0$, where $\cdots$ refers to some arbitrary but fixed number of iterations of taking successors. Consider computing a function f for the 'arbirary' argument b. according to the recursion equations

$$f(0) = k$$

$$f(S(c)) = g(f(c))$$

where k is a given number and g a given function. The computation of $f(b)$ has the form

$$f(b) \Rightarrow \cdots \Rightarrow g(\cdots f(1) \cdots) \Rightarrow g(\cdots g(f(1)) \cdots) \Rightarrow g(\cdots g(g(k)) \cdots)$$

This computation of $f(b)$ for the 'arbitrary' b is a schematic computation involving the iteration of a certain process, but the iterations are not arbitrary now: $\cdots$ has the same meaning in all of its occurrences as it has in the

representation of b by $b, \cdots 1, 0$. Thus, the finitist does not have to accept infinite objects, such as functions; he may read $f(b)$ simply as a shorthand for the above schematic computation. Of course, the computation also involves g; and so, if this is also to be finitistically admissible, there must be such a schematic representation of $g(b)$ (which then is used to successively compute $g(k), g(g(k)), \ldots$.

In general, the idea is that we may begin with arbitrary numbers, say a and b. That is a and b represent fixed, but arbitrary, iterations. We can represent the computations of a function $f(a, b)$ or a proof of a proposition $P(a, b)$ by representing it as a schematic computation or proof involving no arbitrary iterations other than the ones given by the arbitrary numbers a and b. It would seem that this idea of generality is absolutely minimal; that anything more stringent is to deny generality entirely and so to reject non-trivial mathematics. I don't want to discuss this further, but it leads to a precise decription of finitist number theory, namely that part of arithmetic which is formalized in so-called *Primitive Recursive Arithmetic*, PRA. For further argumentation for this, see my paper "Finitism" [Tait, 1981]. What we will do now is to describe the system PRA. It was first discussed by T. Skolem in "The foundations of elementary arithmetic established by means of the recursive mode of thought, without the use of apparent variables ranging over infinite domains" [Skolem, 1923]. "Apparent variable" is an old name for bound variables. So PRA is a theory with no bound variables. We will consistently take free variables to be different symbols than bound variables. (This is common in Proof Theory, like denoting arbitrary formulas by upper case latin letters instead of by lower case greek letters. It is a badge, proclaiming that *this is a work in Proof Theory!* Of course, in the usual formalisms, one can't really speak of free or bound variables, only of free or bound *occurrences* of variables.)

Free (individual) variables will be denoted by a, b, c and d, with or without subscripts. Bound (individual) variables will be denoted by x, y, z, u and v, with or without subscripts. We assume given infinite lists of distinct free and bound variables.

The easiest way to describe the formal system PRA is to simultaneously introduce the non-logical constants and define the terms. The non-logical constants, other than the identity symbol $=$ (which we will regard in these lectures as non-logical), all are function constants—which includes individual constants as function constants of 0 arguments.

The terms and constants of PRA:

- Every free variable is a term.
- 0 is an individual constant.
- S is a function constant of 1 argument.
- If f is a function constant of n arguments, $n > 0$, and $t_1, \ldots, t_n$ are terms then $f(t_1, \ldots, t_n)$ is a term. In some cases, we write sft for $f(s,t)$ when f has 2 arguments.
- If $t(a_1, \ldots, a_n)$ is a term all of whose free variables are in the list $a_1, \ldots, a_n$ of distinct variables, then we introduce a new function constant of n arguments, which we denote by $\lambda x_1 \cdots x_n t(x_1, \ldots, x_n)$.
- . If g and h are function constants of n and n+2 arguments, respectively, then we introduce a new function constant of n+1 arguments, which we denote by $R(g,h)$.

Note that $\lambda x_1 \cdots x_n t(x_1, \ldots, x_n)$ and $R(g,h)$ denote *constants*. The constants are atomic symbols and, unlike our names for them, carry no syntactical structure. As usual, we introduce the *numerals* as abbreviations: $1 = S(0), 2 = S(1)$, etc.

The atomic formulas of PRA are just equations $s = t$ between terms. The *formulas* are built up from these equations by means of the propositional connectives $\neg, \vee, \wedge, \longrightarrow$ and $\longleftrightarrow$. We can of course (since the logic of this system is classical) take some suitable subset (e.g. $\neg$ and $\vee$) of these constants as primitive and introduce the others by contextual definition. $s \neq t$ is of course an abbreviation for $\neg s = t$. The *non-logical axioms* are

$$0 \neq S(t)$$

$$S(s) = S(t) \longrightarrow s = t$$

$$\lambda x_1 \cdots x_n t(x_1, \ldots, x_n)(s_1, \ldots, s_n) = t(s_1, \ldots, s_n)$$

$$R(g,h)(s_1, \ldots, s_n, 0) = g(s_1, \ldots, s_n)$$

$$R(g,h)(s_1, \ldots, s_n, S(t)) = h(s_1, \ldots, s_n, t, R(g,h)(s_1, \ldots, s_n, t))$$

There is just one non-logical rule of inference: the rule of Mathematical Induction

$$A(0), A(b) \longrightarrow A(S(b)) \Longrightarrow A(t)$$

Here $A(b)$ denotes some formula with 0 or more occurrences of b. $A(0), A(S(b)), A(t)$ result from $A(b)$ by replacing all occurrences of b by $0, S(b), t$, respectively. For the logical axioms, choose some complete set of axioms for propositional logic and the axioms

$$t = t$$

$$s = t \longrightarrow [A(s) \longrightarrow A(t)]$$

of identity. We can assume that the only logical rule of inference is Modus Ponens

$$A, A \longrightarrow B \Longrightarrow B$$

Note that all of the axioms that we have written down have the property that if $A(c)$ is an axiom, then so is $A(t)$ for every term t. This is also true, clearly for logical axioms. Moreover, Modus ponens is preserved by substituting t for c throughout both premises and the conclusion. However, the rule of Mathematical Induction is not preserved in general. Why?

Exercise 1 *Prove by induction on n that, if $A(c)$ has a deduction of length n in PRA, then so has $A(t)$ for any term t.* [Hint: If the last inference in the deduction is an instance

$$A(0), A(b) \longrightarrow A(S(b)) \Longrightarrow A(s)$$

of Mathematical Induction, first substitute a new variable d for b in the second premise (using the induction hypothesis), where d does not occur in t.]

It is not difficult to see that the computation of every function $f(a_1, \ldots, a_n)$ of PRA for 'arbitrary' arguments $a_1, \ldots, a_n$ using the defining axioms above and (using the above exercise) that the deduction of every formula $A(a_1, \ldots, a_n)$ with just the variables a_i using the axioms and rules of inference can be analyzed just as the case of the example $f(0) = k, f(S(b)) = g(f(b))$ above. The computation or deduction involves iterations of certain operations, but each such iteration is measured by one of those used in building up the arbitrary numbers $a_1, \ldots, a_n$ themselves from 0 using the successor operation. There is no appeal to the infinite, only to the arbitrarily large finite *as represented by the arguments* $a_1, \ldots, a_n$.

What is perhaps harder to be convinced of is that the primitive recursive functions and the methods of proof formalized in PRA represent the only functions and methods of proof satisfying this criterion. An argument for this is presented in [Tait, 1981].

Exercise 2 *Show that the axioms of* 0 *and successor, i.e.*

$$0 \neq S(t)$$

and

$$S(s) = S(t) \longrightarrow s = t$$

can be replaced by the single axiom

$$0 \neq S(0)$$

[Hint: show that there are function constants pred and sgn such that

$$pred f(0) = 0 \qquad pred(S(t)) = S(0)$$

$$sgn(0) = 0 \qquad sgn(S(t)) = S(0).$$

are theorems of PRA minus the axioms of 0 *and successor.]*

Exercise 3 *Show that there are constants* f, g *and* h *satisfying the recursion equations for addition, multiplication and exponentiation. I.e.*

$$f(s, 0) = s$$

$$f(s, S(t)) = S(f(s, t))$$

$$g(s, 0) = 0$$

$$g(s, S(t)) = f(g(s, t), s)$$

$$h(s, 0) = S(0)$$

$$h(s, S(t)) = g(h(s, t), s)$$

Exercise 4 *Prove that for every formula* A *of* PRA *there is a term* t_A *of* PRA *such that* $A \longleftrightarrow t_A = 0$ *is a theorem of* PRA. [Hint: Find a function constant f such that $f(a, b) = 0 \longleftrightarrow a = b$. Find a function constant neg such that $neg(b) = 0 \longleftrightarrow b \neq 0$ is a theorem of PRA. Find a function constant f such that $f(b, c) = 0 \longleftrightarrow b = 0 \vee c = 0$ is a theorem of PRA. (Try $f(b, c) = b \times c$).]

The functions denoted by function constants of PRA and called *primitive recusive functions*. (Surprise!)

Exercise 5 *Show that each primitive recursive function is explicitly definable in PA; i.e. each function constant f of PRA can be introduced in PA by explicit definition in such a way that all the axioms of PRA are theorems of PA (extended by explicit definitions).* [Hint: For this you need to know some of the development of the theory of syntax in PA that is needed for Gödel's incompleteness theorems. Begin the proof by stating what it is you know: PLEASE don't begin by stating that each primitive recursive function is explicitly definable in PA!]

We can understand the generalization $\forall x A(x)$, where $A(b)$ is a formula of PRA, as simply expressing $A(b)$. In this way, we can express in PRA the consistency of PA and other systems: $\forall x \neg P(g, x)$; since P(a,b) is expressed by a formula of PRA. Similarly, we can understand $\forall x \exists y A(x, y)$ as an incomplete shorthand for the formula $A(b, f(b))$, where f is some function constant. But, of course, these quantified formulas occur only as limiting cases: we cannot in general apply the propositional comnnectives to them.

Any way, the thesis I am proposing, but will not defend here, is that finitist arithmetic, as Hilbert understood it, at least up to the 1930's, is formalized in PRA.

In the sense of the last exercise, we have that $PRA \subseteq PA$. So, on our thesis, all of finitist reasoning is formalized in PA. It follows then from Gödel's second incompleteness theorem that the consistency of PA, if a fact, is not provable finitistically. For, if $P(g, b)$ is a theorem of PRA, then the consistency of PA, $\forall x P(g, x)$, is a theorem of PA.

So, if the consistency of PA or stronger systems is to be proved, stronger methods than those formalized in PRA—and so, on the above thesis, stronger methods than finitist arithmetic—are needed. But now we have opened a Pandora's box: Finitism is distinguished by the particular conception, sketched above, of generality as it occurs in the definition of functions and in proofs. This conception seems minimal, in the sense that it is difficult to see what conception of generality could admit *any* nontrivial functions and methods of proof and not admit all of PRA. So, conceptually speaking, a consistency proof in PRA is the most we could ask for. But once we have to extend beyond PRA to a system Σ in order to prove consistency of PA, we must ask what it is about Σ that makes the proof of consistency significant. At the end of the day, I think that we will see that it is not the consistency proof at all that is interesting; rather it is the details of the proof, from which we can extract a new way of understanding the concepts and methods

of proof formalized in PA.

Chapter 2

Ordinal Numbers, the Veblen Hierarchy, and the Functions of Proof Theory

The first proof-theoretical consistency proof for PA was Gentzen's [Gentzen, 1936] and required that PRA be extended by the principle of transfinite induction up to a certain ordinal ϵ_0. So, since $PRA \subseteq PA$, transfinite induction up to ϵ_0 is not deducible in PA. Gentzen also proved that transfinite induction up to each ordinal $< \epsilon_0$ is deducible in PA. So ϵ_0 is called the *ordinal* of PA. Of course, this description is rough: PA does not contain variables over transfinite ordinals and so we have to understand propositions about some segment of the ordinals in PA as being about some particular well-ordering of the natural numbers representing that segment. (That means of course that the segment has to be countable.)

If someone should ask you why you want to study proof theory, you could answer: Because its there! But, if you want to be more original, you could point out that it is one of the few fields that actually motivates the study of transfinite ordinal arithmetic.

2.1 Preliminaries

By an *ordinal number* we will mean a von Neuman ordinal. I.e., each ordinal is the set of its predecessors; so an ordinal is a set of ordinals, and $\alpha < \beta$ iff $\alpha \in \beta$. Moreover, every transitive set of ordinals is an ordinal. (M is

transitive iff $x \in M$ always implies $x \subset M$.) α, β, γ and δ with or without subscripts denote ordinals. $\emptyset$ is an ordinal and, indeed, the least one: $\emptyset = 0$; and $\alpha \cup \{\alpha\}$ is an ordinal and, in fact, the least ordinal greater than α : $\alpha \cup \{\alpha\} = \alpha + 1$, called the *successor* of α. An ordinal which is neither 0 nor a *successor* is called a *limit* ordinal. ω denotes the least limit ordinal. If M is a set of ordinals, then $\bigcup M$ is an ordinal and is, in fact, the least upper bound of M. When M is null, $\bigcup M = 0$; when M has a greatest element α, then $\bigcup M = \alpha$. If M is nonempty and has no greatest element, then $\bigcup M$ is a limit. An ordinal-valued function f with domain of definition a set M of ordinals is called *order preserving* iff for all α and β in M, $\alpha < \beta$ implies $f(\alpha) < f(\beta)$ If f is an order preserving ordinal-valued function defined on some limit ordinal γ, then we define

$$Lim_{\alpha<\gamma} f(\alpha) = \bigcup \{f(\alpha) \mid \alpha < \gamma\}$$

$Lim_{\alpha<\gamma} f(\alpha)$ is necessarily a limit ordinal.

The totality of ordinals Ω is well-ordered by $<$, i.e. by $\in$. This means that $<$ is a linear ordering of Ω and that every non-empty subset of Ω has a least element. Call a class C of ordinals *hereditary* iff $\forall \alpha < \beta(\alpha \in C)$ always implies that $\beta \in C$. Then we have the principle of Transfinite Induction: Ω *is the only hereditary class of ordinals.* I.e.

$$[\forall \alpha < \beta(\alpha \in C) \longrightarrow \beta \in C] \longrightarrow \forall \alpha(\alpha \in C)$$

By a *cardinal number*, sometimes called an *initial ordinal*, we mean an ordinal which is not equal in power to one of its elements. A set is said to be of power *less than* {*greater than, equal to*} the cardinal κ iff it is in one-to-one correspondence with a cardinal which is less than {greater than, equal to} the cardinal κ. $\lambda, \kappa, \mu, \nu$ with or without subscripts will denote cardinals. A cardinal κ is called *regular* iff whenever $M \subset \kappa$ and M is of power $< \kappa$, then $\bigcup M < \kappa$. ω denotes the least infinite ordinal. If κ is a cardinal, let M be the set of all ordinals of power $\leq$ that of κ. Then $\bigcup M$ is a cardinal which is the least cardinal $> \kappa$. (If $\beta = \bigcup M$ were of power $< \kappa$, then so would $\beta + 1$ be; and so $\beta + 1 \in M$, i.e., $\beta + 1 \subseteq \bigcup M$, i.e., $\beta \in \bigcup M = \beta$ —a contradiction.). We denote $\bigcup M$ by κ^+.

Exercise 6 *a) Show that each finite ordinal, i.e. each element of ω, is a regular cardinal.*

b) Show that ω is a regular cardinal.

c) Show that each infinite cardinal (i.e. $\geq \omega$) is a limit ordinal.

d) Show that, for each κ, κ^+ is regular.

2.2 Normal Functions and Classes

If f is a function, $D(f)$ denotes its domain of definition and $R(f)$ its range of values. If M and N are sets, then $f : M \longrightarrow N$ means that f is a function with $D(f) = M$ and $R(f) \subseteq N$.

Lemma 1 *Let $f : \gamma \longrightarrow \delta$ be order preserving. Then for all $\alpha < \gamma$, $\alpha \leq f(\alpha)$.*

Suppose otherwise. Then there is a least $\alpha < \gamma$ such that $f(\alpha) < \alpha$. But since f is order preserving, it would then follow that $f(f(\alpha)) < f(\alpha)$, which contradicts the choice of α.

Lemma 2 *If $f : \gamma \longrightarrow M$ and $g : \delta \longrightarrow M$ are order preserving functions onto the set M, i.e. $R(f) = R(g) = M$, then $f = g$ and so $\gamma = \delta$.*

For g^{-1} is an order preserving function $M \longrightarrow \delta$ and so $g^{-1} \circ f : \gamma \longrightarrow \delta$ is order preserving. Hence $\alpha \leq g^{-1} \circ f(\alpha)$ for all $\alpha < \gamma$, i.e. $g(\alpha) \leq f(\alpha)$ for all $\alpha < \gamma$. By symmetry, $f(\alpha) \leq g(\alpha)$ for all $\alpha < \delta$. So $f(\alpha) = g(\alpha)$ for all $\alpha \in \gamma \cap \delta$, the least of γ and δ. If $\gamma < \delta$, then $g(\gamma)$ is in M and so must be $= f(\alpha) = g(\alpha)$ for some $\alpha < \gamma$, which is impossible since g is order preserving. For the same reason, $\delta < \gamma$ is impossible. So $\gamma = \delta$ and $f = g$.

Proposition 1 *For every subset M of an ordinal ϵ, there is a unique order preserving function $f : \delta \longrightarrow M$ for some ordinal $\delta \leq \epsilon$ with $R(f) = M$. f is called the* enumeration *of M and is denoted by $R^{-1}(M)$. δ is called the* ordinal *of M and is denoted by $o(M)$.*

We define f by recursion: Assume that $f(\alpha)$ is defined for all $\alpha < \beta$. If $\{f(\alpha) \mid \alpha < \beta\} = M$, then we are done: set $\delta = \beta$. Otherwise, there is a least element $\gamma \in M - \{f(\alpha) \mid \alpha < \beta\}$. Set $f(\beta) = \gamma$. Since f is order preserving and $M \subseteq \epsilon$, $\epsilon > f(\alpha) \geq \alpha$ for all α for which $f(\alpha)$ is defined. So, for some $\beta \leq \epsilon$, the first case must apply.

f is necessarily unique by Lemma 2.

From now on in this chapter, unless otherwise specified, we will be speaking only of ordinals $\in \kappa$, where κ is some fixed regular cardinal $> \omega$.

$M \subseteq \beta$ is called *unbounded* in β iff $\bigcup M = \beta$ iff for no $\alpha < \beta$ is $M \subseteq \alpha$. When we say simply that $M \subseteq \kappa$ is unbounded, it will always be understood that we mean "unbounded in κ". Let a subset M of β be unbounded in β. Then β must be 0 or a limit ordinal. For if $\beta = \alpha + 1$, then $\bigcup M \leq \alpha$. If β is a limit, then $\beta = Lim_{\alpha<\delta} f(\alpha)$, where $f : \delta \longrightarrow M$ is the enumeration of M.

Lemma 3 *$M \subseteq \kappa$ is unbounded iff its ordinal is κ.*

Let $f : \delta \longrightarrow M$ be the enumeration of M. $\delta \leq \kappa$. If $\delta < \kappa$, then, since M has the same power as δ, which is less than κ, and κ is regular, $\bigcup M < \kappa$—a contradiction.

So the unbounded sets are precisely the ranges of the order preserving functions $\kappa \longrightarrow \kappa$. Note that, if M and N are unbounded with enumerations f and g, respectively, and if $M \subseteq N$, then $g(\alpha) \leq f(\alpha)$ for all $\alpha < \kappa$. For $g^{-1} \circ f : \kappa \longrightarrow \kappa$ is order preserving and so $\alpha \leq g^{-1} \circ f(\alpha)$ for all $\alpha < \kappa$. I.e. $g(\alpha) \leq f(\alpha)$.

If $M \subseteq \kappa$, β is a limit ordinal $< \kappa$ and if $M \cap \beta$ is unbounded in β, then β is called a *limit point* of M (relative to κ). M is called *closed* iff it contains all of its limit points. Closed unbounded subclasses of κ are called *club* (relative to κ). Let $f : \kappa \longrightarrow \kappa$. f is called *continuous* iff for every limit ordinal $\gamma < \kappa, f(\gamma) = Lim_{\alpha<\gamma} f(\alpha)$. f is called *normal* iff it is order preserving and continuous.

Proposition 2 *Let $M \subseteq \kappa$. M is club iff $R^{-1}(M)$ is normal.*

Let $f = R^{-1}(M)$. We have proved that M is unbounded iff $D(f) = \kappa$. Let M be club and $\gamma < \kappa$ be a limit ordinal. Set $\beta = Lim_{\alpha<\gamma} f(\alpha)$. Then $M \cap \beta$ is unbounded in β and so $\beta \in M$ and moreover, β is the least element of $M - \{f(\alpha) \mid \alpha < \gamma\}$. Hence, $\beta = f(\gamma)$.

Conversely, let f be normal and let $\beta \cap M$ be unbounded in the limit ordinal β. Let γ be the least upper bound of the α such that $f(\alpha) < \beta$. Since $\beta \cap M$ be unbounded in M, γ is a limit and so $f(\gamma) = Lim_{\alpha<\gamma} f(\alpha) = \beta$. So $\beta \in M$.

In set theory, the notion of a club set plays an important role; but in proof theory, because we are interested in computing ordinals, we will see that it is the correlative notion of normal function that is important.

Exercise 7 *Prove that, if f and g are normal functions, then $f \circ g$ is normal.*

Lemma 4 *a. Let $f : \kappa \longrightarrow \kappa$. It is normal iff for every $\alpha < \kappa$*

$$f(\alpha) < f(\alpha + 1)$$

and if, for every limit $\beta < \kappa$,

$$f(\beta) = \bigcup_{\alpha < \beta} f(\alpha)$$

b. Let $\delta < \kappa$, $g : \kappa \longrightarrow \kappa$ and suppose that $\alpha < g(\alpha)$ for all $\alpha < \kappa$. Then there is a unique normal function f such that

$$f(0) = \delta$$

$$f(\alpha + 1) = g(f(\alpha))$$

a. We prove by induction on β that $\alpha < \beta \longrightarrow f(\alpha) < f(\beta)$. So assume $\alpha < \beta$. If $\beta = \delta + 1$, then $\alpha \leq \delta$ and so by the induction hypothesis, $f(\alpha) \leq f(\delta) < f(\beta)$. If β is a limit, then there is a $\delta < \beta$ with $\alpha < \delta$. So $f(\alpha) < f(\delta) \leq \bigcup_{\epsilon < \beta} f(\epsilon) = f(\beta)$. We have proved that f is order preserving and so, when β is a limit, $f(\beta) = \bigcup_{\alpha < \beta} f(\alpha) = Lim_{\alpha < \beta} f(\alpha)$.

b. For limits $\gamma < \kappa$, define $f(\gamma) = \bigcup_{\alpha < \gamma} f(\alpha)$. Since κ is regular, $\gamma < \kappa$, and $f(\alpha) < \kappa$ for all $\alpha < \kappa$, it follows that $f(\gamma) < \kappa$.

As an example, for any $\alpha < \kappa$ define $\alpha + \beta$ as a function $\alpha+ : \kappa \longrightarrow \kappa$ of β by

$$\alpha + 0 = \alpha$$

$$\alpha + (\beta + 1) = (\alpha + \beta) + 1$$

$$\alpha + \gamma = \bigcup_{\alpha < \gamma} \alpha + \beta$$

for γ a limit. (Note that the ambiguous use of '+1' is harmless, since the two meanings of $\alpha + (1)$ coincide.) Since $\alpha + \beta < \alpha + (\beta + 1)$, $\alpha+$ is a normal function. One easily shows that $\alpha+$ enumerates the set $\kappa - \alpha$. As a second example, define the function $\alpha \times (\beta) = \alpha \times \beta$ by

$$\alpha \times 0 = 0$$

$$\alpha \times \beta + 1 = \alpha \times \beta + \alpha$$

$$\alpha \times \gamma = \bigcup_{\beta<\gamma} \alpha \times \beta$$

for γ a limit. If $\alpha > 0$, $\beta + \alpha < \beta$ and so $\alpha\times$ is a normal function. By induction on β, $0 \times \beta = 0$ is easily proved.

We will shortly use the following :

Proposition 3 *Let γ and δ be limit ordinals. Then every $\alpha < \gamma \times \delta$ is uniquely of the form $\alpha = \gamma \times \zeta + \epsilon$ where $\zeta < \delta$ and $\epsilon < \gamma$.*

Choose the least θ such that $\alpha < \gamma \times \theta$. One exists, since $\gamma\times$ is a normal function. For the same reason, θ must be a successor, $\theta = \zeta + 1$. So $\gamma \times \zeta \leq \alpha < \gamma \times \zeta + \gamma$. Since $(\gamma \times \zeta)+$ enumerates $\kappa - \gamma \times \zeta$, there is a unique ϵ with $\alpha = \gamma \times \zeta + \epsilon$. Since $\alpha < \gamma \times \zeta + \gamma$, we must have $\epsilon < \gamma$. It is clear from the construction that ζ and ϵ are uniquely determined.

Let

$$M(f)$$

denote the set of fixed points $\alpha = f(\alpha)$ of the function $f : \kappa \longrightarrow \kappa$.

Proposition 4 *If f is normal, then $M(f)$ is club.*

Assume that f is normal. In order to prove that $M(f)$ is unbounded, we need to show that for each $\alpha < \kappa$ there is a fixed point of $f \geq \alpha$. If $\alpha = f(\alpha)$, we are done. Otherwise, define $\alpha_0 = \alpha$ and $\alpha_{n+1} = f(\alpha)$. Since $\alpha_0 < \alpha_1$ and f is order preserving, $\alpha_n < \alpha_{n+1}$ for all $n < \omega$. Set $\gamma = Lim_{n<\omega}\alpha_n$. Since κ is regular and $> \omega$, $\gamma < \kappa$. $f(\gamma) = Lim_{n<\omega} f(\alpha_n) = Lim_{0<n<\omega}\alpha_n = \gamma$ and so γ is a fixed point $< \alpha$. So $M(f)$ is unbounded.

Let β be a limit ordinal $< \kappa$ and let $M(f) \cap \beta$ be unbounded in β. I.e. $\beta = Lim_{\alpha<\delta} g(\alpha)$, where $g : \delta \longrightarrow M(f) \cap \beta$ is the enumeration of $M(f) \cap \beta$. Then $f(\beta) = Lim_{\alpha<\delta} f(g(\alpha)) = Lim_{\alpha<\delta} g(\alpha) = \beta$, since the values of g are fixed points of f. So $\beta \in M(f)$.

So, when f is normal, the enumeration of $M(f)$ is normal. We denote it by f' and call it the *derivative* of f.

Proposition 5 *Let $\delta < \kappa$ and for each $\alpha < \delta$, let M_α be a club subset of κ. Then $M = \bigcap_{\alpha<\delta} M_\alpha$ is club.*

Let $\beta < \kappa$ be a limit and let $M \cap \beta$ be unbounded in β. Then $M_\alpha \cap \beta$ is unbounded in β for each $\alpha < \delta$. So $\beta \in M_\alpha$ for each α, i.e. $\beta \in M$ and so M is closed.

We need to show that, for each $\alpha < \kappa$, there is an element of $M > \alpha$. Since δ and $\omega < \kappa$, $\delta \times \omega < \kappa$. We define an increasing sequence $\langle \alpha_\xi \mid \xi < \delta \times \omega \rangle$ by recursion. $\alpha_0 = \alpha + 1$. For $0 < \xi$, ξ is uniquely of the form $\delta \times n + \theta$, where $n < \omega$ and $\theta < \delta$. α_ξ is the least element of M_θ greater than $Lim_{\zeta<\xi}\alpha_\zeta$. Since κ is regular and so $Lim_{\zeta<\xi}\alpha_\zeta < \kappa$, one such element exists since M_θ is unbounded in κ. Again, by regularity, $\beta = Lim_{\xi<\delta\times\omega}\alpha_\xi < \kappa$. Then for each $\theta < \delta$, $\beta = Lim_{n<\omega}\alpha_{\delta\times n+\theta}$; i.e., β is the limit of a sequence of elements of M_θ. Hence, since M_θ is closed, $\beta \in M_\theta$ for each $\theta < \delta$. I.e., $\beta \in M$.

Exercise 8 *Call a subset $S \subseteq \kappa$* stationary *iff $S \cap C \neq \emptyset$ for every club set C.*

a) Let $\langle C_\alpha \mid \alpha < \kappa\}$ be a sequence of club sets. Prove that the diagonal *of this sequence*

$$\Delta_\alpha C_\alpha = \{\beta < \kappa \mid \forall \alpha < \beta(\beta \in C_\alpha)\}$$

is club. [Hint: Use Proposition 5 to show that we can assume $\alpha < \beta$ implies $C_\beta \subseteq C_\alpha$.]

b) Let $f : \kappa \longrightarrow \kappa$ and suppose that the set $S = \{\alpha < \kappa \mid f(\alpha) < \alpha\}$ is stationary. Prove that there is a γ such that $M_\gamma = \{\alpha \in S \mid f(\alpha) = \gamma\}$ is stationary. [Fodor, 1956] [Hint: Otherwise, for each γ there is a club set C_γ such that $M_\gamma \cap C_\gamma = \emptyset$. Consider $S \cap \Delta_\gamma C_\gamma$.]

2.3 Veblen Hierarchies

Using Propositions 2, 4 and 5, we can construct the following hierarchies of normal functions.

Definition 1 *Let f be a normal function. The* Veblen hierarchy $\langle f_\alpha \mid \alpha < \kappa \rangle$ *of normal functions based on f is defined by*

$$f_0 = f$$

and for $\beta > 0$

$$f_\beta = R^{-1}(\bigcap_{\alpha<\beta} M(f_\alpha))$$

Thus, for $\beta > 0$, f_β enumerates the common fixed points of all the f_α for $\alpha < \beta$. Let $\alpha < \beta$. Then $M(f_\beta) \subseteq R(f_\beta) \subseteq M(f_\alpha)$. So

$$\bigcap_{\alpha \leq \beta} M(f_\alpha) = M(f_\beta)$$

and, if γ is a limit

$$\bigcap_{\alpha < \gamma} M(f_\alpha) = \bigcap_{\alpha < \gamma} R(f_\alpha)$$

In view of these equations, we have the following alternative characterization of the hierarchy.

Proposition 6 *If f is a normal function, then*

$$f_0 = f$$

$$f_{\alpha+1} = (f_\alpha)'$$

and if γ is a limit

$$f_\gamma = R^{-1} \bigcap_{\alpha < \gamma} R(f_\alpha)$$

The following proposition characterizes the order of the ordinals of the form $f_\alpha(\xi)$.

Proposition 7 *$f_\alpha(\xi) < f_\beta(\zeta)$ iff*

a. $\alpha < \beta$ and $\xi < f_\beta(\zeta)$, or

b. $\alpha = \beta$ and $\xi < \zeta$, or

c. $\beta < \alpha$ and $f_\alpha(\xi) < \zeta$.

Proof: Let $\alpha < \beta$. $\xi < f_\beta(\zeta)$ iff $f_\alpha(\xi) < f_\alpha(f_\beta(\zeta)) = f_\beta(\zeta)$. Let $\alpha = \beta$. $\xi < \zeta$ iff $f_\alpha(\xi) < f_\alpha(\zeta) = f_\beta(\zeta)$. Let $\beta < \alpha$. $f_\alpha(\xi) < \zeta$ iff $f_\alpha(\xi) = f_\beta(f_\alpha(\zeta)) < f_\beta(\zeta)$.

κ is a club set and its enumeration is the identity function i_κ with $i_\kappa(\alpha) = \alpha$ for all $\alpha < \kappa$. Every $\alpha < \kappa$ is a fixed point of i_κ and so its derivative $i'_\kappa = i_\kappa$. It easily follows that the whole Veblen hierarchy in this case collapses: $(i_\kappa)_\alpha = i_\kappa$ for all α. But the following proposition shows that this is the only case in which the hierarchy collapses: For f a normal function other than i_κ, $\alpha \neq \beta \longrightarrow f_\alpha \neq f_\beta$.

Proposition 8 *Let f be a normal function, $f \neq i_\kappa$ and let $\delta = \delta_f$ be the least ordinal such that $f(\delta) \neq \delta$. Then*

a. $f_\alpha(\xi) = \xi$ for all $\alpha < \kappa$ and all $\xi < \delta$.

b. $\delta < f_\alpha(\delta)$ for all $\alpha < \kappa$.

c. $f^*(\alpha) = f_\alpha(\delta)$ defines a normal function f^*.

Proof.

a. This holds for $\alpha = 0$ by the definition of δ. Suppose that it holds for all $\alpha < \beta$. Then $\delta \subseteq M(f_\alpha)$ for all $\alpha < \beta$ and so $\delta \subseteq \bigcap_{\alpha<\beta} M(f_\alpha) = R(f_\beta)$.

b. $\delta < f(\delta)$ since $\delta \neq f(\delta)$ and $\delta \leq f(\delta)$ since f is normal. $R(f_\alpha) \subseteq R(f)$ and so $\delta < f(\delta) \leq f_\alpha(\delta)$.

c. Let $\alpha < \beta$. $\delta < f_\beta(\delta)$ by b) and so $f_\alpha(\delta) < f_\alpha(f_\beta(\delta)) = f_\beta(\delta)$. So f^* is order preserving; and so we need only prove that it is continuous. Let γ be a limit and $\beta = Lim_{\alpha<\gamma} f_\alpha(\delta)$. We must prove that $f_\gamma(\delta) = \beta$. By a), $\delta \subseteq R(f_\gamma)$ and so $f_\gamma(\delta)$ is the least element of $R(f_\gamma) - \delta$. $\beta \in \bigcap_{\alpha<\gamma} R(f_\alpha) = R(f_\gamma)$, since for each $\alpha < \gamma, \beta = Lim_{\alpha<\xi<\gamma} f_\xi(\delta) \in R(f_\alpha)$ (since $R(f_\alpha)$ is closed). So by b), $\beta \in R(f_\gamma) - \delta$; and hence $f_\gamma(\delta) \leq \beta$. By b), $\delta < f_\gamma(\delta)$ and so $f_\alpha(\delta) < f_\alpha(f_\gamma(\delta)) = f_\gamma(\delta)$ for each $\alpha < \gamma$. Hence, $\beta = Lim_{\alpha<\gamma} f_\alpha(\delta) \leq f_\gamma(\delta)$. I.e. $\beta = f_\gamma(\delta)$.

Let $f \neq i_\kappa$ be normal. The normal function f_* is defined by

$$f_* = (f^*)' = R^{-1}(M(f^*)$$

Proposition 9 *Let $f \neq i_\kappa$ be normal.*

a. $f_*(0)$ is a limit.

b. If $f_*(\gamma)$ is a limit, then

$$\alpha, \beta < f_*(\gamma) \Rightarrow f_*(\beta) < f_\beta(\alpha) < f_*(\gamma)$$

c. If $\delta_f < \theta < f_*(0)$, then there are $\alpha, \beta < \theta$ with $f_\beta(\alpha) \geq \theta$. In fact, we may take $\alpha = \delta$.

Proof.

a. $0 \leq \delta < f_{f_*(0)}(\delta) = f_*(0)$. Let $\alpha + 1 \in M(f^*)$. Then $\alpha \leq f_\alpha(\delta) < f_{\alpha+1}(\delta) = \alpha + 1$. So $\alpha \in M(f^*)$. So the least element $f_*(0)$ of $M(f^*)$ is a limit.

b. Let $\epsilon = Max\{\alpha, \beta\} + 1$. Since $f_*(\gamma)$ is a limit, $\epsilon < f_*(\gamma)$. So

$$f_\beta(\alpha) \leq f_\beta(f_\alpha(\delta)) < f_\beta(f_\epsilon(\delta)) = f_\epsilon(\delta) < f_{f_*(\gamma)}(\delta) = f_*(\gamma)$$

c. $f_\theta = f^*(\theta) > \theta$, since $f_*(0)$ is the least fixed point of f^*. So $\forall \beta < \theta (f_\beta(\delta) < \theta$ contradicts the continuity of f^*.

$f_*(\gamma)$ need not be a limit for $\gamma > 0$. For example, given a normal function $f \neq i_\kappa$, define g by $g(\alpha) = f(\alpha)$ for $\alpha < f_*(0)$ and $g(\alpha) = \alpha$ for $\alpha \geq f_*(0)$. g is normal and $\neq i_\kappa$ and $g_*(\alpha) = f_*(0) + \alpha$ for all $\alpha < \kappa$.

Proposition 10 *Let $f \neq i_\kappa$ be normal and $\delta = \delta(f) \leq \alpha \in R(f)$. Then there is a greatest ordinal $\beta = \beta_\alpha$ such that $\alpha \in R(f_\beta)$. $\beta \leq \alpha$ and for $\alpha = f_\beta(\gamma), \gamma < \alpha$. If $\alpha < f_*(0)$, then $\beta < \alpha$.*

Proof. $\alpha \leq f_\alpha(\delta) < f_{\alpha+1}(\delta) \leq f_{\alpha+1}(\gamma)$ for all $\gamma \geq \delta$. If $\gamma < \delta$, then $f_{\alpha+1}(\gamma) = \gamma < \alpha$. So $\alpha \notin R(f_{\alpha+1})$. Let γ be the least ordinal with $\alpha \notin R(f_\gamma)$. $\gamma > 0$, since $\alpha \in R(f)$. γ is not a limit since, for a limit $\epsilon, R(f_\epsilon) = \bigcap_{\xi<\epsilon} R(f_\xi)$. So $\gamma = \beta+1$ for some $\beta = \beta_\alpha$. $\alpha \in R(f_\beta)$ and, for $\theta > \beta, \alpha \notin R(f_\theta) \subseteq R(f_{\beta+1})$. $\beta + 1 \leq \alpha + 1$, and so $\beta \leq \alpha$. Let $\alpha = f_\beta(\xi)$. $\xi \leq \alpha$ since f_β is normal and $\alpha \neq \xi$, since $\alpha \notin R(f_{\beta+1})$. So $\xi < \alpha$. Note that $\delta \leq \xi$, since $\xi < \delta$ implies $f_\beta(\xi) = \xi < \alpha$. Now suppose that $\alpha < f_*(0)$. Then $\beta_\alpha < \alpha$, since $\beta = \alpha$ implies $\alpha < f_\alpha(\delta) \leq f_\alpha(\xi) = \alpha$—a contradiction.

2.4 Ordinal Arithmetic

We have already defined ordinal addition and multiplication. Ordinal exponentiation $\alpha - exp\beta = \alpha^\beta$ is defined by

$$\alpha^0 = 1$$
$$\alpha^{\beta+1} = \alpha^\beta \times \alpha$$

and for limits γ,

$$\alpha^\gamma = \bigcup_{\beta<\gamma} \alpha^\beta$$

$0^0 = 0$ and $0^\beta = 0$ for all $\beta > 0$; and $1^\beta = 1$ for all β. But, if $\alpha > 1$, then $\alpha - exp$ is normal by Lemma 4, since $\beta \times \alpha > \beta$.

Remark 2 *If we want to prove an equation $s(\alpha) = t(\alpha)$ where the terms $s(\alpha)$ and $t(\alpha)$ express normal functions of α (e.g. when they are built up by means of a composition of normal functions), then it suffices to prove $s(0) = t(0)$ and that $s(\alpha) = t(\alpha)$ implies $s(\alpha + 1) = t(\alpha = 1)$. The equation then follows by induction since, if $s(\alpha) = t(\alpha)$ holds for all $\alpha < \gamma$, where γ is a limit, then $s(\gamma) = t(\gamma)$ holds by continuity.*

Exercise 9 *a) Prove that $+$ and $\times$ are associative:*

$$(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$$

$$(\alpha \times \beta) \times \gamma = \alpha \times (\beta \times \gamma$$

b) Show that neither $+$ nor $\times$ is commutative. In particular, show that

$$1 + \omega = \omega < \omega + 1$$

$$2 \times \omega = \omega < \omega \times 2$$

c) Prove the left distribution law

$$\alpha \times (\beta + \gamma) = \alpha \times \beta + \alpha \times \gamma$$

d) Show that the right distribution law

$$(\alpha + \beta) \times \gamma = \alpha \times \gamma + \beta \times \gamma$$

fails. [Hint: Prove that $1 \times \alpha = \alpha$. Now, using b), consider $2 \times \omega = (1 + 1) \times \omega$.]

e) Prove

$$\alpha^{\beta+\gamma} = \alpha^\beta \times \alpha^\gamma$$

$$\alpha^{\beta\times\gamma} = (\alpha^\beta)^\gamma$$

The main result we need here is a generalization of a theorem of arithmetic of finite ordinals.

Definition 2 *The* ordering by first differences *of finite sequences of ordinals is defined by*

$$\langle \alpha_1, \ldots, \alpha_m \rangle \prec \langle \beta_1, \ldots, \beta_n \rangle$$

iff there is a $j \leq n$ *such that* $\alpha_i = \beta_i$ *for* $i < j$ *and either* $j = m + 1$ *or else* $j \leq m$ *and* $\alpha_j < \beta_j$.

Exercise 10 *Prove that* $\prec$ *does not well-order the set of finite sequences of ordinals* $< \kappa$*—in fact, that it does not even well-order the set of finite sequences of ordinals* < 2.

Proposition 11 *Let* $\alpha > 1$.

a. Every $\beta > 0$ *is uniquely of the form*

$$\beta = \alpha^{\beta_1} \times \zeta_1 + \cdots + \alpha^{\beta_m} \times \zeta_m$$

where $m \geq 0$, $0 < \zeta_i < \alpha$ *and* $\beta_1 > \cdots > \beta_m$. *We call this the* normal form *of* β *to the* base α. *We include the case of* $\beta = 0$ *by allowing m=0.*

b. If

$$\gamma = \alpha^{\gamma_1} \times \eta_1 + \cdots + \alpha^{\gamma_n} \times \eta_n$$

is also in normal form, then $\beta < \gamma$ *iff*

$$\langle \alpha^{\beta_1} \times \zeta_1, \ldots, \alpha^{\beta_m} \times \zeta_m \rangle \prec \langle \alpha^{\gamma_1} \times \eta_1, \ldots, \alpha^{\gamma_n} \times \eta_n \rangle$$

Proof. a. We prove this by induction on β. Let γ be the least ordinal such that $\beta < \alpha^\gamma$. Since αexp is a normal function (when $\alpha > 1$), γ cannot be a limit. If $\gamma = 0$, then $\beta = 0$ and m=0 gives the required form. Otherwise, $\gamma = \beta_1 + 1$ for some β_1 and so $\alpha^{\beta_1} \leq \beta < \alpha^{\beta_1} \times \alpha$. By Proposition 3, β is uniquely of the form $\alpha^{\beta_1} \times \zeta_1 + \theta$ for some ζ_1 with $0 < \zeta_1 < \alpha$ and $\theta < \alpha^{\beta_1}$. The last inequality implies $\theta < \beta$. So by the induction hypothesis, θ has the required form, which we can write as $\theta = \alpha^{\beta_2} + \cdots + \alpha^{\beta_m} \times \zeta_m$. Note that Since $\alpha - exp$ is normal and $\theta < \alpha^{\beta_1}, \beta_1 > \beta_2$.

b. It clearly suffices to show that, if $\beta = \alpha^{\beta_1} \times \zeta_1 + \cdots + \alpha^{\beta_m} \times \zeta_m$ is the normal form of β to the base α and $\gamma > \beta_1$, then $\alpha^\gamma > \beta$. We prove this by induction on m. If m=0, $\beta = 0 < \alpha^\gamma$. Let m = k+1. By the inductive hypothesis, $\alpha^{\beta_1} > \alpha^{\beta_2} \times \zeta_2 + \cdots + \alpha^{\beta_m} \times \zeta_m$. So $\alpha^\gamma \geq \alpha^{\beta_1 + 1} = \alpha^{\beta_1} \times \alpha \geq \alpha^{\beta_1} \times (\zeta_1 + 1) = \alpha^{\beta_1} \times \zeta_{1'} + \alpha^{\beta_1} > \beta$.

When $\alpha = 2$, the ζ_i must all be =1; and so the normal form to base 2 is

$$\beta = 2^{\beta_1} + \cdots + 2^{\beta_m}$$

where $\beta_1 > \cdots > \beta_m$. When $\alpha = \omega$, then the ζ_i are all $< \omega$. So $\omega^{\beta_i} \times \zeta_i = \omega^{\beta_i} + \cdots + \omega^{\beta_i}$ (ζ_i summands) and so the normal form to base ω can also be written as

$$\beta = \omega^{\beta_1} + \cdots + \omega^{\beta_m}$$

where $\beta_1 \leq \cdots \leq \beta_m$. In both cases, normal form to base 2 and normal form to base ω, $\beta < \gamma$ iff $\langle \beta_1, \ldots, \beta_m \rangle \prec \langle \gamma_1, \ldots, \gamma_n \rangle$, where the β_i and γ_j are the exponents in the corresponding normal forms.

O. Veblen [1908] introduced the hierarchy $\langle k_\alpha \mid \alpha < \omega^+ \rangle$ based on the normal function $k = \omega - exp$. His purpose was to introduce a unique notation for each ordinal $< k_*(0)$. We shall instead consider the hierarchy $\langle h_\alpha \mid \alpha < \omega^+ \rangle$, where $h = 2 - exp$.

Exercise 11 *a) Show that the fixed points of $n+$ for $0 < n \leq \omega$ are precisely the infinite ordinals ($\alpha \geq \omega$).*

b) Show that the fixed points of $\omega+$ are precisely the ordinals of the form $\omega \times \alpha$ for α infinite.

c) Show that every ordinal is a fixed point of $1 \times$.

d) Show that the fixed points of $\alpha\times$ are precisely the ordinals ω^δ greater than α. [Hint: Show by induction on δ that $\alpha, \beta < \omega^\delta$ implies $\alpha + \beta < \omega^\delta$.]

e) Let $\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_m}$, $m > 0$, and $\beta = \omega^{\beta_1} + \cdots + \omega^{\beta_n}$, where $\alpha_1 \geq \cdots \geq \alpha_m$ and $\beta_1 \geq \cdots \geq \beta_n$. Show that

$$\alpha + \beta = \omega^{\alpha_1} + \cdots + \omega^{\alpha_i} + \beta$$

for the greatest $i = 1, \ldots m$ such that $\alpha_i \geq \beta_1$. [Hint: use d).]

f) Let $\alpha = \omega^{\alpha_1} \times k_1 + \cdots + \omega^{\alpha_m} \times k_m$ be the normal form of α to the base ω. Show that $\alpha \times n = \omega^{\alpha_1} \times k_1 \times n + \omega^{\alpha_2} \times k_2 + \cdots + \omega^{\alpha_m} \times k_m$. [Hint: Use e).]

g) Let α be as in f) and let $\gamma > 0$. Show that $\alpha \times \omega^\gamma = \omega^{\alpha_1 + \gamma}$.

h) Now let α and β be as in e). Show that

$$\alpha \times \beta = \omega^{\alpha_1+\beta_1} + \cdots + \omega^{\alpha_1+\beta_n}$$

i) Show that $2^\alpha = \omega^\alpha$ iff α is a fixed point of $\omega\times$.

j) Show that the fixed points of $\omega - exp$ are $> \omega$.

k) Show that the fixed points of $2 - exp$ are ω and the fixed points of $\omega - exp$, so that

$$2 - exp_1(0) = \omega$$

$$2 - exp_1(1+\alpha) = \omega - exp_1(\alpha)$$

l) Show that, for $\beta > 1, 2 - exp_\beta = \omega - exp_\beta$.

Going back to Cantor, the values of the derived function $\omega - exp'$ of $k = \omega - exp$ are denoted by $\omega - exp'(\alpha) = \epsilon_\alpha$ and called the *ϵ-numbers*. So these are the fixed points $\omega^\epsilon = \epsilon$. So we have $2-exp(0) = \omega$ and $2-exp(1+\alpha) = \epsilon_\alpha$.

2.5 The Functions of Proof Theory

In proof theory it is not the function h or k that we want, but the function ψ:

Definition 3

- *$\psi : \kappa \longrightarrow \kappa$ is defined by*

$$\psi(0) = 0$$

$$\psi(1+\alpha) = 2^\alpha = 2 - exp(\alpha)$$

- *Let $\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n}$, where $\alpha_1 \geq \cdots \geq \alpha_n$. This form is unique and so we can define $\chi_\alpha : \kappa \longrightarrow \kappa$ by*

$$\chi_\alpha(\beta) = \psi_{\alpha_1}(\cdots(\psi_{\alpha_n}(\beta))\cdots)$$

ψ is clearly normal. $\delta_\psi = 3$, since $\psi(\alpha) = \alpha$ for $\alpha < 3$ and $\psi(3) = 4$. From now on

$$\langle \psi_\alpha \mid \alpha < \kappa \rangle$$

will refer to the hierarchy based on this particular ψ. $\langle \chi_\alpha \mid \alpha < \kappa \rangle$ is not a Veblen hierarchy, of course, since χ_0 is the identity function i_κ on κ and the

hierarchy based on this function collapses to just one function. But since the composition of normal functions is normal, each χ_α is normal.

Every ordinal α is uniquely of the form $\gamma + n$, where γ is 0 or a limit and $n < \omega$. So

$$2^\alpha = 2^\gamma \times 2^n = 2^\gamma + \cdots + 2^\gamma$$

with 2^n summands. It follows then from Proposition 11 that every ordinal α is uniquely of the form

$$\alpha = 2^{\alpha_1} + \cdots + 2^{\alpha_n} \tag{2.1}$$

where each α_i is a limit or 0 and $\alpha_1 \geq \cdots \geq \alpha_n$. We will call this the *dyadic form* of α. Let

$$\beta = 2^{\beta_1} + \cdots + 2^{\beta_m} \tag{2.2}$$

also be in dyadic form. It follows easily from Proposition 11 that $\alpha < \beta$ iff $\langle \alpha_1, \ldots, \alpha_n \rangle \prec \langle \beta_1, \ldots, \beta_m \rangle$. Now, let $\zeta_1 \geq \cdots \geq \zeta_{m+n}$ be the α_i's and β_j's, counting multiplicities. The *commutative sum* of α and β is defined by

$$\alpha \oplus \beta = 2^{\zeta_1} + \cdots + 2^{\zeta_{m+n}} \tag{2.3}$$

Note that this is also in dyadic form.

Proposition 12 *$\langle \psi, \oplus \rangle$ is the least pair of functions such that*

a. $\oplus$ is commutative: $\alpha \oplus \beta = \beta \oplus \alpha$

b. $\alpha\oplus$ is order preserving: $\beta < \gamma \longrightarrow \alpha \oplus \beta < \alpha \oplus \gamma$

c. ψ is order preserving

d. $\psi(\alpha) \oplus \psi(\alpha) \leq \psi(\alpha + 1)$.

In other words, if $\langle g, \otimes \rangle$ also satisfies these conditions, then for all α and β

$$\alpha \oplus \beta \leq \alpha \otimes \beta$$

$$\psi(\alpha) \leq g(\alpha)$$

Proof. First we must show that ψ and $\oplus$ satisfy these conditions.
a. is immediate.

b. Let (2.1) and (2.2) and $\gamma = 2^{\gamma_1} + \cdots + 2^{\gamma_k}$ be dyadic normal forms. $\alpha \oplus \beta$ is given by 2.3. Let $\alpha \oplus \gamma = 2^{\eta_1} + \cdots + 2^{\eta_{n+k}}$. $\beta < \gamma$ implies $\langle \beta_1, \ldots, \beta_m \rangle \prec \langle \gamma_1, \ldots, \gamma_k \rangle$, which implies $\langle \zeta_1, \ldots, \zeta_{m+n} \rangle \prec \langle \eta_1, , \ldots, \eta_{n+k} \rangle$; and this implies $\alpha \oplus \beta < \alpha \oplus \gamma$.

c. is immediate since ψ is normal.

d. This is immediate for $\alpha = 0$. Let $\alpha > 0$ and $\psi(\alpha)$ have the dyadic form $2^\gamma + \cdots 2^\gamma$ with 2^k summands ($1 + \alpha = \gamma + k$). So $\psi(\alpha) \oplus \psi(\alpha) = 2^\beta + 2^\beta = 2^{\beta+1} = \psi(\alpha + 1)$.

In fact, ψ is the least order preserving function satisfying $\psi(\alpha) + \psi(\alpha) \leq \psi(\alpha + 1)$. If g is another such function, then $\psi(0) = 0 \leq g(0)$. Now assume $\psi(\alpha) \leq g(\alpha)$. $\psi(\alpha+1) = \psi(\alpha) \times 2 \leq g(\alpha) \times 2 \leq g(\alpha+1)$. If γ is a limit and we assume $\psi(\alpha) \leq g(\alpha)$ for all $\alpha < \gamma$, then $\psi(\gamma) = \bigcup_{\alpha<\gamma} \psi(\alpha) \leq \bigcup_{\alpha<\gamma} g(\alpha) \leq g(\gamma)$.

Now we assume that $\langle g, \otimes \rangle$ satisfies a.- d. $\alpha+$ is the least order preserving function $j : \kappa \longrightarrow \kappa$ with $j(0) = \alpha$. So by b., $\alpha + \beta \leq \alpha \otimes \beta$. Hence

$$g(\alpha) + g(\alpha) \leq g(\alpha) \otimes g(\alpha) \leq g(\alpha + 1)$$

But ψ is the least function satisfying $\psi(\alpha) \times 2 \leq \psi(\alpha+1)$, and so $\psi(\alpha) \leq g(\alpha)$ for all α.

Now we prove $\alpha \oplus \beta \leq \alpha \otimes \beta$ by induction on $\alpha \oplus \beta$. Let the dyadic forms of α, β, and $\alpha \oplus \beta$ be (2.1), (2.2) and 2.3), respectively. Suppose

$$\alpha \otimes \beta < \alpha \oplus \beta = 2^{\zeta_1} + \cdots + 2^{\zeta_{m+n}}$$

By the commutativity of $\oplus$ and $\otimes$, we can assume that $\zeta_{m+n} = \beta_m$. m cannot be 0, since otherwise $\beta = 0$ and we have $\alpha \otimes 0 < \alpha \oplus 0 = \alpha$, contradicting the fact that $0 \otimes \alpha = \alpha \otimes 0$ is an order preserving function of α. If $\beta_m = 0$, set $\gamma = 2^{\beta_1} + \cdots + 2^{\beta_{m-1}}$. If β_m is a limit, then there are $\eta_k \leq \cdots \leq \eta_1 \leq \zeta_{m+n-1}$ such that

$$\alpha \otimes \beta < 2^{\zeta_1} + \cdots + 2^{\zeta_{m+n-1}} + 2^{\eta_1} + \cdots + 2^{\eta_k}$$

Set $\gamma = 2^{\beta_1} + \cdots + 2^{\beta_{m-1}} + 2^{\eta_1} + \cdots + 2^{\eta_k}$. In either case, we have $\gamma < \beta$ with $\alpha \otimes \gamma < \alpha \oplus \gamma$. Hence $\alpha \otimes \gamma < \alpha \otimes \beta \leq \alpha \oplus \gamma < \alpha \oplus \beta$, contradicting the induction hypothesis that $\zeta \oplus \eta \leq \zeta \otimes \eta$ for all $\zeta \oplus \eta < \alpha \oplus \beta$. QED

We should compare the hierarchies $\langle \psi_\alpha \mid \alpha < \kappa \rangle$ and $\langle 2 - exp_\alpha \mid \alpha < \kappa \rangle$.

If k and m are normal functions, then so is their composition $k \circ m$ and, since $\alpha = k(m(\alpha))$ iff $\alpha = m(\alpha) = k(\alpha)$,

$$M(k \circ m) = M(k) \cap M(m).$$

$2 - exp(\alpha) = \psi(1 + \alpha)$, and so $M(2 - exp) = M(\psi) - \omega$. The finite fixed points of ψ are the $\alpha < 3 = \delta_\psi$ and so for $\beta = 1$,

$$2 - exp_\beta(\alpha) = f_\beta(3 + \alpha) \tag{2.4}$$

Assume that this holds for some β. $\delta(\psi_\beta)$ is again 3, and so $M(2 - exp_\beta) = M(\psi_\beta) - \omega$. Hence (2.4) holds also for $\beta + 1$. By continuity, then, it holds for all $\beta > 0$. $\delta_{2-exp} = 0$ and so $e2 - exp^*(\alpha) = 2 - exp_\alpha(0) = \psi_\alpha(3) = \psi^*(\alpha)$. I.e. $\psi^* = 2 - exp^*$ and so $\psi_* = 2 - exp_*$. So $2 - exp_*(0)$ supports the hierarchy $\langle \psi_\alpha \mid \alpha < \kappa \rangle$, i.e.

$$\alpha, \beta < 2 - exp_*(0) \longrightarrow \psi_\alpha(\beta) < 2 - exp_*(0)$$

If $\alpha, \beta < 2 - exp_*(0)$, then $\alpha \oplus \beta \leq \psi(\alpha) \oplus \psi(\beta) \leq \psi(\alpha \cup \beta) + 1) < 2 - exp_*(0)$. Thus $2 - exp_*(0)$ also supports the function $\oplus$.

It follows then that, as far as the functions ψ_α and $\oplus$ of proof theory are concerned, we may as well take $\kappa = 2 - exp_*(0)$ rather than a regular uncountable cardinal. The function $2 - exp_*$ has a name in proof theory: namely, $2 - exp_(*) = \Gamma$. But we will have another use for the symbol Γ.

Chapter 3

The Cut Elimination Theorem

3.1 Deductive Systems

Definition 4 *A* formula system *consists of*

1. *A set of objects, called* formulas*. The formulas are partitioned into three disjoint kinds:* prime *formulas,* disjunctive *or* $\bigvee$ *formulas, and* conjunctive *or* $\bigwedge$ *formulas. With each disjunctive or conjunctive formula is associated a class of formulas called its* components*. Prime formulas have no components. A disjunctive formula with components* $\{A_i \mid i \in I\}$*, is often denoted by*

$$\bigvee_{i \in I} A_i$$

A conjunctive formula with these components is often denoted by

$$\bigwedge_{i \in I} A_i$$

But CAUTION: This notation is ambiguous, since it is not excluded that two disjunctive, respectively, conjunctive formulas have the same components. So $\bigvee_{i \in I} A_i$ *denotes* some *disjunctive formula with these components. Similarly in the conjunctive case.*

2. *An operation* $A \mapsto \overline{A}$ *associating with each formula* A *a formula* $\overline{A}$*, called its* complement*, such that the complement of a prime formula is prime,*

$$\overline{\bigvee_{i \in I} A_i} = \bigwedge_{i \in I} \overline{A_i} \quad \overline{\bigwedge_{i \in I} A_i} = \bigvee_{i \in I} \overline{A_i}$$

and

$$\overline{\overline{A}} = A$$

So complements come in pairs, where each member of the pair is the complement of the other.

3. *An operation $A \mapsto |A|$, where $|A|$ is an ordinal, called the* rank *of A. The rank operation is to satisfiy two conditions:*

   (a) $|\overline{A}| = |A|$

   (b) $|A| < |B|$ *for every component A of B.*

We give three examples of formulas systems.

Example 1 *The Formula System of Countable Propositional Logic.*

- Formulas. *Assume given a countably infinite list of distinct* propositional constants*: $P_0, P_1, \ldots$. The set of formulas is defined by*

  - *Every propositional constant is a formula.*
  - *If M is a set of formulas, then $\bigvee M$ and $\bigwedge M$ are formulas.*

  The propositional constants are the prime formulas, $\bigvee M$ is disjunctive and $\bigwedge M$ is conjunctive.

- Components *The components of $\bigvee M$ and $\bigwedge M$ are the formulas in M. (We are not excluding the case of $M = \emptyset$.)*

- Complements.

$$\overline{P_{2n}} = P_{2n+1} \quad \overline{P_{2n+1}} = P_{2n}$$

$$\overline{\bigvee M} = \bigwedge \overline{M}$$
$$\overline{\bigwedge M} = \bigvee \overline{M}$$

  where $\overline{M}$ is the set of $\overline{A}$ for $A \in M$.

- Ranks.

$$|P_n| = 0$$

$$|\bigvee M| = |\bigwedge M| = \bigcup_{A \in M} |A|$$

Exercise 12 *Show that, for each countable ordinal α, there is a formula A of Countable Propositional Logic with $|A| = \alpha$.*

Example 2 *The Formula System of Classical First Order Predicate Logic, CPL.*

- Formulas. *The terms are built up in the usual way from free variables, individual constants and function constants. The prime formulas are of the form*

$$P(t_1, \ldots, t_n) \qquad \bar{P}(t_1, \ldots, t_n)$$

where P is a predicate constant of n arguments and the t_i are terms. We allow the case of n=0, in which case P is a propositional constant and we write P and $\bar{P}$ rather than $P()$ and $\bar{P}()$. The formulas are built up from the prime formulas and the propositional constants $\top$ (for trivial truth) and $\bot$ (for absurdity or trivial falsety) by means of disjunction $A \vee B$, conjunction $A \wedge B$, existential quantification $\exists x A(x)$, and universal quantification $\forall x A(x)$, where $A(x)$ is obtained from a formula $A(b)$ by replacing each occurrence of the free variable b by the bound variable x which does not occur in $A(b)$. Thus, besides the prime formulas, we have the formulas

$\bot$	$\top$
$A \vee B$	$A \wedge B$
$\exists x A(x)$	$\forall x A(x)$

- $\bot$, $A \vee B$, *and* $\exists x A(x)$ *are disjunctive formulas.* $\top$, $A \wedge B$, *and* $\forall x A(x)$ *are conjunctive.*

- Components. $\bot$ *and* $\top$ *have no components,* $A \vee B$ *and* $A \wedge B$ *have the components* A *and* B, *and* $\exists x A(x)$ *and* $\forall x A(x)$ *have the components* $A(t)$, *for every term* t.

- Complements. *The complement operation is defined by*

$\overline{P(t_1, \ldots, t_n)} = \bar{P}(t_1, \ldots, t_n)$	$\overline{\bar{P}(t_1, \ldots, t_n)} = P(t_1, \ldots, t_n)$
$\overline{\bot} = \top$	$\overline{\top} = \bot$
$\overline{A \vee B} = \bar{A} \wedge \bar{B}$	$\overline{A \wedge B} = \bar{A} \vee \bar{B}$
$\overline{\exists x A(x)} = \forall x \overline{A(x)}$	$\overline{\forall x A(x)} = \exists x \overline{A(x)}$

- Rank.
  - *Prime formulas have rank 0.*
  - $|\bot| = |\top| = 0$.
  - $|A \vee B| = |A \wedge B| = Max\{|A|, |B|\} + 1$.
  - $|\exists x A(x)| = |\forall x A(x)| = |A(b)| + 1$.

  In the case of the quantifiers, note that $|A(t)| = |A(b)|$ *for every term* t. *So we do have* $|\exists x A(x)| = |\bigvee_{t \in Tm} A(t)| > |A(t)|$ *for all components* $A(t)$*; and similarly for* $\forall x A(x)$.

Note that, unlike the case of Countable Propositional Logic, the ranks of the formulas of CPL are all finite. This will be true of the following example, too. However, we will later consider natural formula systems containing formulas of infinite rank.

Example 3 *The Formula Systems of Classical First Order Arithmetic,* CPA *and* CPA^*. *These formula systems are defined exactly as in the case of* CPL, *with these exceptions:*

- *In both* CPA *and* CPA^*, *the only individual constant is* $\bar{0}$ *and the only function constants are* S *(for successor) and the function constants of* PRA. *In* CPA *the only predicate constants are* $=$ *and a single unary predicate constant* U.
- *In* CPA^*, *the* formulas *are the* closed *formulas, or* sentences, *of* CPA—*i.e. containing no free variables.*
- *We identify the closed term* t *with* $\bar{n}$, *where* n *is the value of* t. *The* $\bar{n}$*'s are the* numerals*:*
$$\bar{1} = S(\bar{0}), \bar{2} = S(\bar{1}), \ldots$$
- *In* CPA^*, *the components of* $\exists x A(x)$ *and* $\forall x A(x)$ *are just the formulas* $A(\bar{n})$.

Remark 3 *First order arithmetic is usually formalized using only the constants for the functions* $+$ *and* $\times$, *rather than for all primitive recursive functions. The other primitive recursive functions can be introduced by explicit definition so that the defining equations associated with them become*

theorems. This shows that our formalization of first order arithmetic is conservative over the usual formalism, in the sense that any sentence of the latter that is provable in our system is already provable in latter system.

In all of these examples of *classical systems*, we might (but will not) introduce the negation $\neg A$ of a formula A by

$$\neg A \ := \ \overline{A}$$

Thus, with this definition, the classical inference from $\neg\neg A$ to A becomes the trivial inference from A to itself. This shows that, for intuitionistic logic, our notion of a formula system is not sufficiently rich.

Definition 5 *A* signed *formula system is a formula system in which each formula is classified as* positive*, +, or* negative*, -, subject to the condition that A is positive iff $\overline{A}$ is negative.*

Example 4 *The Formula Systems of* Intuitionistic *First Order Predicate Logic, IPL, and of* Intuitionistic *First Order Arithmetic, IPA and IPA^*, are defined exactly as in the corresponding classical systems except that we assign signs to the formulas as follows:*

- *$P(t_1, \ldots, t_n)$ is positive and $\overline{P}(t_1, \ldots, t_n)$ is negative.*
- *$\bot$ is positive and $\top$ is negative.*
- *$\exists x A(x)$ and $\forall x A(x)$ are positive iff $A(x)$ is positive; otherwise they are negative.*
- *The signs for $A \vee B$ and $A \wedge B$ are given by the table*

A	B	$A \vee B$	$A \wedge B$
+	+	+	+
+	-	+	-
-	+	+	-
-	-	-	-

From now on, we will take all formula systems to be signed: we have made our point that, in the classical case, the sign plays no role; but it also does no harm and can be assigned arbitrarily (consistent with A and $\overline{A}$ having opposite signs). In particular, the formula systems of CPL, and IPL are the

same, the formula systems of CPA and IPA are the same and the formula systems of CPA^* and IPA^* are the same.

Implication and negation are defined by

$$A \longrightarrow B \ := \ \overline{A} \vee B$$

$$\neg A \ := \ A \longrightarrow \bot$$

So $A \longrightarrow B$ is positive except when A is positive and B is negative; hence, $\neg A$ is always positive. Bi-implication is defined by

$$A \longleftrightarrow B \ := \ (A \longrightarrow B) \wedge (B \longrightarrow A)$$

We would like to define the notion of a *deduction*, not for single formulas, but for finite sets of formulas.

$$\Gamma, \ \Delta, \ \Theta, \ \Gamma_\alpha, \ldots$$

will denote finite sets of formulas.

$$\Gamma, \Delta$$

will denote the union $\Gamma \cup \Delta$ of Γ and Δ. It is not excluded in this notation that Γ and Δ contain some common formulas.

$$\Gamma, A_1, \ldots, A_n$$

will denote the set $\Gamma \cup \{A_1, \ldots, A_n\}$. An *intuitionistic set* or *i-set* of formulas is a finite set of formulas from a signed formula system which contains at most one + formula.

In classical logic, we may think of a set Θ of formulas occurring as a premise or conclusion of an inference as expressing the disjunction $\bigvee \Theta$ of the formulas in Θ.

In intuitionistic logic, negative formulas have no meaning as assertions; rather they stand for hypotheses. To interpet Θ intuitionistically, we should divide it into two parts, $\overline{\Gamma}$ (where $\overline{\Gamma}$ is the set of complements of formulas in Γ) and Δ, consisting of the negative formulas and the positive formulas in Θ, respectively. Then Θ expresses that the conjunction of the (positive) formulas in Γ implies the disjunction of the (positive) formulas in Δ, i.e. $\bigwedge \Gamma \longrightarrow \bigvee \Delta$.

Notice that, in classical logic, $\bigvee \Theta$ and $\bigwedge \Gamma \longrightarrow \bigvee \Delta$ are logically equivalent, since the complement of a formula expresses its negation and $A \longrightarrow B$ means $\neg A \vee B$; but in intuitionistic logic these are not equivalent: for example, $A \longrightarrow A$, but $\neg A \vee A$ is not a valid principle of intuitionistic logic.

In the original formalism of the *Sequence Calculi* of [Gentzen, 1935], θ is written as the *sequent*

$$\Gamma \Rightarrow \Delta$$

expressing $\bigwedge \Gamma \longrightarrow \bigvee \Delta$. So in this notation, all formulas occuring in inferences are positive. It follows that an i-set of formulas in the sequent notation is either

$$\Gamma \Rightarrow \emptyset$$

which is usually written

$$\Gamma \Rightarrow$$

or else of the form

$$\Gamma \Rightarrow A$$

As a matter of fact, Gentzen took $\overline{\Gamma}$ and Δ to be, not sets of formulas, but sequences. It simplifies the rules of inference, however, to replace sequences by sets and sequents by sets of positive and negative formulas.

Definition 6

- *A* Classical Deductive System Σ *consists of a formula system* Σ_F *together with a set* Σ_A *of* axioms sets. *The axiom sets are finite sets of prime formula of the system. The only conditions on* Σ_A *are*
  - *For every prime formula* P*, some subset of* $P, \overline{P}$ *is in* Σ_A. [Completeness Condition]
  - *If* $\Gamma, A \in \Sigma_A$ *and* $\Delta, \overline{A} \in \Sigma_A$*, then some subset of* Γ, Δ *is in* Σ_A. [Cut Condition]
- *An Intuitionistic Deductive System is a classical deductive system except that the axiom sets are i-sets.*

The rules of inference for a deductive system Σ are:

AX $\qquad \Gamma, \underline{\Delta} \qquad (\Delta \in \Sigma_A)$

$$\bigvee \qquad \frac{\Delta, \underline{A_j}}{\Gamma, \underline{\bigvee_{i\in I} A_i}} \qquad (\text{some } j \in I, \Delta \subseteq \Gamma)$$

$$\bigwedge \qquad \frac{\Delta_j, \underline{A_j}}{\Gamma, \underline{\bigwedge_{i\in I} A_i}} \qquad (\text{all } j \in I, \Delta_j \subseteq \Gamma)$$

$$\textbf{CUT} \qquad \frac{\Delta, \underline{A} \qquad \Theta, \underline{\overline{A}}}{\Gamma} \qquad (\Delta, \Theta \subseteq \Gamma)$$

Notice that, when Δ, A and $\Theta\overline{A}$ are i-sets, then Δ, Θ is an i-set; so it is always possible to cut A with the conclusion Δ, Θ. The rules of inference other than CUT are called *normal* rules of inferences and their instances are called normal. The sets of formulas above the line are called *premises* and the set below the line is called the *conclusion*. Thus **AX** has no premises, $\bigvee$ has just one premise, $\bigwedge$ has a premise corresponding to each component A_j of $\bigwedge_{i\in I} A_i$. **CUT** has two premises. Note that in each premise of each inference we have underlined one formula. This is called the *minor formulas* or *mf* of the premise, and together they are called the mf of the inference. The mf of a cut are called its *cut-formulas* or *cf*. In each normal inference, there are underlined formulas in the conclusion, called the *principal formulas* or *pf* of the inference. In **AX**, all the formulas in Δ are pf. In the other normal inferences, there are just one pf. **CUT** has no pf. Normal inferences have *cut-degree* $= 0$. The cut-degree of a cut, however, is the rank $|A| = |\overline{A}|$ of its cut-formulas.

We shall adopt the following convention regarding the notion of an inference: An inference is given, not only by its premises and conclusion, but also by its mf and pf. Thus, there may be two distinct inferences with the same premises and conclusion. For example

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

Intuitionistic Restriction In intuitionistic deductive systems, all inferences are restricted to the case that every premise and conclusion is an i-set.

Exercise 13 *Rewrite the rules of inference in the notation of sequents involving just positive formulas.*

Definition 7 (Deductions) *We define the notion of a* deduction $\mathcal{D}$ *of* Γ *of* rank $\leq \alpha$ *and* cut-degree $\leq \delta$

$$\mathcal{D} \vdash \Gamma \ [\alpha, \delta]$$

by induction on α*. Namely, if*

$$(*) \qquad \frac{\cdots \Gamma_i \cdots}{\Gamma} \qquad (i \in I)$$

is an inference with cut-degree $< \gamma$ *and for each* $i \in I$*,* $\mathcal{D}_i \vdash \Gamma_i \ [\alpha_i, \gamma]$*, then* $\mathcal{D} = \langle (*), \langle \mathcal{D}_i \mid i \in I \rangle, \Gamma \rangle$ *is a deduction of* Γ *of rank* $\leq \alpha$ *and cut-degree* $\leq \delta$*.* $(*)$ *is called the* inference *of* $\mathcal{D}$ *and the* $\mathcal{D}_i$ *are called its* subdeductions*.*

We write

$$\vdash \Gamma \ [\alpha, \delta]$$

to mean that there is a deduction $\mathcal{D}$ of Γ of rank $\leq \alpha$ and cut-degree $\leq \delta$ and

$$\vdash \Gamma \ [< \alpha, \delta]$$

to mean that there is such a deduction whose rank is $< \alpha$.

Note that the *rank* and *cut-degree* of a deduction have not actually been defined; but we may define them as the least α such that the deduction has rank $\leq \alpha$ and the least δ such that the deduction has cut-degree $\leq \delta$, respectively. It is natural to think of the deduction as a well-founded tree (well-founded meaning that every path upward is finite), where each node is the conclusion Γ of an inference and the nodes immediately above Γ, if any, are the premises of the inference. Then the rank of the deduction is the height of the tree. It cut degree is the least ordinal greater than the cut-degree of each cut in the deduction. So a deduction contains no cuts just in case its cut degree is 0: in this case, we say that the deduction is *normal*.

Exercise 14 *Assume given a deductive system, either classical or intuitionistic.*

a) Prove that $\vdash A, \overline{A}\ [2 \times |A|, 0]$*, for every* A*.*

b) Prove that $\vdash \overline{A} \vee A\ [2 \times |A| + 2, 0]$ *for every* A *(providing a formula* $\overline{A} \vee A$ *exists in the deductive system).* [Hint: Be careful in the intuitionistic case. Only i-sets are admitted.]

c) Prove that $\vdash \top\ [0, 0]$*, providing a conjunctive formula* $\top$ *with no components exist in the system.*

d) Prove that, if the disjunctive formula $\bot$ *with no components exists in the system, then* $\vdash \Gamma, \bot\ [\alpha, \delta]$ *implies* $\vdash \Gamma\ [\alpha, \delta]$*.*

Lemma 5 (Weakening Lemma) *If* $\vdash \Gamma\ [\alpha, \delta]$*, then* $\vdash \Gamma, \Delta\ [\alpha, \delta]$*, providing in the intuitionistic case that* Γ, Δ *is an i-set.*

Proof. Simply add Δ to the conclusion of the inference of the given deduction of Γ.

There are two properties of classical systems which we shall want to use later on. By a *contraction*, we mean an inference in which the conclusion is included in each premise.

Lemma 6 (Contraction Lemma) *In classical deductive systems, if* $\vdash \Gamma\ [\alpha, \delta]$*, then there is a deduction of* Γ *of rank* $\leq \alpha$ *and cut-degree* $\leq \delta$ *in which every inference is a contraction.*

Proof by induction on α. Let

$$\frac{\cdots \Gamma_i \cdots}{\Gamma}$$

be the inference of the given deduction. $\Gamma_i \subseteq \Gamma$ and so, by Weakening, we may suppose that each $\Gamma_i = \Gamma$. Now apply the induction hypotheis to the deductions of the premises.

Lemma 7 (Reduction Lemma) *In a classical system assume*

$$\vdash \Gamma\ [\alpha, \delta]$$

a) For each $j \in I$

$$\vdash \Gamma - \{\bigwedge_{i\in I} A_i\}, A_j \ [\alpha, \delta]$$

b) If I is finite, then

$$\vdash \Gamma - \{\bigvee_{i\in I} A_i\}, \{A_j \mid j \in I\} \ [\alpha, \delta]$$

c) If $\{A\}$ is an axiom set, then

$$\vdash \Gamma - \{\bar{A}\} \ [\alpha, \delta]$$

Proof of part a) by induction on α. The other parts are equally straightforward. We can assume the the given deduction contains only contractions. If the pf of its inference is other than $\bigwedge_{i\in I} A_i$, then the inference is preserved by replacing Γ by $\Gamma - \{\bigwedge_{i\in I} A_i\}, A_j$ in all the premises and in the conclusion. Now apply the induction hypothesis to the deductions of the premises. If $\bigwedge_{i\in I} A_i$ is the pf of the final inference, then one of the premises is Γ, A_j. Apply the induction hypotheis to this premise.

3.2 The Elimination Theorem

Lemma 8 (Elimination Lemma) *Let $\mathcal{D} \vdash \Gamma, A \ [\alpha, \delta]$ and $\mathcal{D}' \vdash \Gamma', \overline{A} \ [\beta, \delta]$, where $|A| \leq \delta$. Then $\vdash \Gamma, \Gamma' \ [\alpha \oplus \beta, \delta]$.*

Of course, we obtain $\vdash \Gamma, \Gamma' \ [\alpha \cup \beta, \delta + 1]$ using a cut with cut formula A. In fact, if $|A| < \delta$, there is no increase in cut-degree. The real content is when $|A| = \delta$.

Proof by induction on $\theta = \alpha \oplus \beta$. In order to prove the lemma in the intuitionistic case, we need only observe that when $\mathcal{D}$ and $\mathcal{D}'$ are intuitionistic deductions, then so is the deduction we construct of Γ, Γ'. This is routine in each case and is left to the reader.

CASE 1. A is not a pf of $\mathcal{D}$ or else $\overline{A}$ is not a pf of $\mathcal{D}'$. Since $\oplus$ is commutative, $|A| = |\overline{A}|$ and $A = \overline{\overline{A}}$, the lemma is symmetric in A and $\overline{A}$. Hence, we can assume that A is not a pf of $\mathcal{D}$. So the inference of $\mathcal{D}$ is of the form

$$(I) \qquad \frac{\cdots \Gamma_i, \underline{B_i} \cdots \Gamma_j, \underline{B_j}, A, \cdots}{\Gamma, A} \qquad (i \in I, j \in J)$$

indicating the possibility of some premises containing A as something other than a mf and others not.

$$\vdash \Gamma_i, B_i \ [\alpha_i, \delta]$$

and

$$\vdash \Gamma_j, B_j, A \ [\alpha_j, \delta]$$

where the α_i and α_j are all $< \alpha$. By the induction hypothesis, since $\alpha_j \oplus \beta < \theta$,

$$\vdash \Gamma_j, B_j, \Gamma' \ [\alpha_i \oplus \beta, \delta]$$

$$(II) \qquad \frac{\cdots \Gamma_i, \underline{B_i} \cdots \Gamma_j, \underline{B_j}, \Gamma' \cdots}{\Gamma, \Gamma'} \qquad (i \in I, j \in J)$$

is an inference, and so since $\alpha_i, \alpha_j \oplus \beta < \theta$, $\vdash \Gamma, \Gamma' \ [\theta, \delta]$.

CASE 2. A is pf of $\mathcal{D}_0$ and $\overline{A}$ is pf of $\mathcal{D}_1$.

CASE 2a. A is a prime formula. Then $\Gamma, A = \Delta, A, \Theta$, where Δ, A is an axiom set, and $\Gamma', \overline{A} = \Delta', \overline{A}, \Theta'$, where $\Delta', \overline{A}$ is an axiom set. By the Cut Condition on axiom sets, this means that $\Delta, \Delta' \subseteq \Gamma, \Gamma'$ includes an axiom set. So $\vdash \Gamma, \Gamma' \ [0, 0]$.

CASE 2b. A is non-prime. Note that in this case, α and β must be > 0. By symmetry we may suppose that $A = \bigvee_{i \in I} A_i$ and so $\overline{A} = \bigwedge_{i \in I} \overline{A_i}$. So the inference of $\mathcal{D}$ has a single premise Δ, A_j with mf A_j, where

$$\vdash \Delta, A_j \ [\alpha', \delta]$$

and $\alpha' < \alpha$. The inference of $\mathcal{D}'$ has the premise $\Delta', \overline{A_j}$ with mf $\overline{A_j}$, where

$$\vdash \Delta', \overline{A_j} \ [\beta', \delta]$$

and $\beta' < \beta$. Since α and β are > 0, we have $\alpha = \alpha \oplus 0 < \theta$ and, similarly, $\beta < \theta$. Also, $\alpha' \oplus \beta, \alpha \oplus \beta' < \theta$. Here we are using the fact that $\oplus$ is order preserving in both arguments.

If $\Delta = \Delta - \{A\}$, then

$$\vdash \Delta - \{A\}, A_j \ [\alpha', \delta] \tag{3.1}$$

If $A \in \Delta$, then by the induction hypotheis applied to $\alpha' \oplus \beta$

$$\vdash \Gamma', \Delta - \{A\}, A_j \ [\alpha' \oplus \beta, \delta] \tag{3.2}$$

If $\Delta' = \Delta' - \{\overline{A}\}$, then

$$\vdash \Delta' - \{\overline{A}\}, \overline{A_j}[\beta', \delta] \tag{3.3}$$

If $\overline{A} \in \Delta'$, then by the induction hypothesis applied to $\alpha \oplus \beta'$,

$$\vdash \Gamma, \Delta' - \{\overline{A}\}, \overline{A_j} \ [\alpha \oplus \beta', \delta] \tag{3.4}$$

Since $|A_j| < |A| < \delta$, we may apply cut to one of the four pairs (3.1,3.3), (3.1, 3.4), (3.2, 3.3) or (3.2, 3.4) of premises to obtain

$$\vdash \Gamma, \Gamma' \ [\theta, \delta]$$

Recall that, when $\gamma = \omega^{\gamma_1} + \cdots + \omega^{\gamma_n}$ is the normal form to base ω of γ, i.e. where $\gamma_1 \geq \cdots \geq \gamma_n$, then

$$\chi_\gamma(\beta) = \psi_{\gamma_1}(\cdots \chi_{\gamma_n}(\beta) \cdots)$$

We now state the main theorem of this chapter.

Theorem 2 (Cut Elimination Theorem) *In every deductive system*

a. $\mathcal{D} \vdash \Gamma \ [\alpha, \delta + \omega^\gamma]$ *implies* $\vdash \Gamma \ [\psi_\gamma(\alpha), \delta]$.

b. $\vdash \Gamma \ [\alpha, \delta + \gamma]$ *implies* $\vdash \Gamma \ [\chi_\gamma(\alpha), \delta]$.

In particular, taking $\delta = 0$, a deduction of Γ of rank α and cut-degree γ can be transformed into a normal deduction of rank $\leq \chi_\gamma(\alpha)$.

b) is obtained by setting $\gamma = \omega^{\gamma_1} + \cdots + \omega^{\gamma_n}$ in normal form and iterating a) n times.

a) The proof is by induction on γ and, within that, by induction on α. In other words, we assume that a) holds for all α' and γ' such that (i) $\gamma' < \gamma$ or (ii) $\gamma' = \gamma$ and $\alpha' < \alpha$. Under these hypotheses, we prove it for α and γ.

First assume that the last inference of $\mathcal{D}$ is not a cut. Then it is of the form

$$(**) \qquad \frac{\cdots \Gamma_j \cdots}{\Gamma} \qquad (j \in J)$$

where for each $j \in J$ there is an $\alpha_j < \alpha$ with $\vdash \Gamma_j\ [\alpha_j, \delta + \omega^\gamma]$. By the induction hypothesis (ii), $\vdash \Gamma_j\ [\psi_\gamma(\alpha_j), \delta]$ for each j. So, since $\psi_\gamma(\alpha_j) < \psi_\gamma(\alpha)$ the result follows using (**).

So we can assume that the last inference of $\mathcal{D}$ is a cut

$$\frac{\Delta, A \qquad \Theta, \overline{A}}{\Gamma}$$

where $\Delta, \Theta \subseteq \Gamma$. It follows that $|A| < \delta + \omega^\gamma$ and that there are $\alpha', \alpha'' < \alpha$ with $\vdash \Delta, A\ [\alpha', \delta + \omega^\gamma]$ and $\vdash \Theta, \overline{A}\ [\alpha'', \delta + \omega^\gamma]$. Let $\beta = \alpha' \cup \alpha''$. Then $\beta < \alpha$. By the induction hypothesis (ii)

$$\vdash \Delta, A\ [\psi_\gamma(\beta), \delta]$$
$$\vdash \Theta, \overline{A}\ [\psi_\gamma(\beta), \delta]$$

CASE 1. $\gamma = 0$. Then $|A| \leq \delta$, and so by the Elimination Lemma, $\vdash \Gamma\ [f(\beta) \oplus f(\beta), \delta]$. But $f(\beta) \oplus f(\beta) \leq f(\beta + 1) \leq f(\alpha) = \psi_0(\alpha)$.

CASE 2. $\gamma > 0$. Since $|A| < \delta + \omega^\gamma$, there is a $\theta < \gamma$ and $m < \omega$ with $|A| < \delta + \omega^\theta \times k$. So applying the above cut,

$$\vdash \Gamma\ [\psi_\gamma(\alpha), \delta + \omega^\theta \times k]$$

Iterating the induction hypothesis (i) k times and noticing that the values of ψ_γ are fixed points of ψ_θ

$$\vdash \Gamma\ [\psi_\theta(\cdots \psi_\theta(\psi_\gamma(\alpha)) \cdots), \delta]$$

I.e.

$$\vdash \Gamma\ [\psi_\gamma(\alpha), \delta]$$

$f(0) = 0 < \omega^0$ and when $0 < \alpha < \omega$, $f(\alpha+1) = 2^\alpha < 2^{\alpha+1}$. When $\alpha \geq \omega$, $f(\alpha) = 2^\alpha$. So in any case, $f(\alpha) \leq 2^\alpha$. For all α, n and β, define β^α_n

$$\beta^\alpha_0 = \alpha \qquad \beta^\alpha_{n+1} = \beta^{\beta^\alpha_n}$$

Then $f^n(\alpha) \leq 2^\alpha_n$. So

Corollary 1 *If $k < \omega$, then*

$$\vdash \Gamma\ [\alpha, \delta + k]$$

implies

$$\vdash \Gamma\ [2^\alpha_k, \delta]$$

Definition 8 • *The set $SubF(A)$ of* subformulas *of a formula A in a formula system is defined by induction on the rank of A:*

$$SubF(A) = \{A\} \cup \bigcup \{SubF(B) \mid B \text{ is a component of } A\}$$

- *$SubF(\Gamma)$ is the set of all formulas occuring as subformulas of some formula in Γ.*
- *Let $\mathcal{D}$ be a deduction. We define the notion of a* formula *in $\mathcal{D}$ by induction on its rank. Namely, A is a formula in $\mathcal{D}$ iff it is in the conclusion of the inference of $\mathcal{D}$ or it is in some subdeduction of $\mathcal{D}$.*

Corollary 2 (Subformula Property) *If $\mathcal{D} \vdash \Gamma\ [\alpha, 0]$, then every formula in $\mathcal{D}$ is a subformula of some formula in Γ.*

Proof. Every formula in a premise of a normal inference is a subformula of a formula in the conclusion.

Exercise 15 *a) Show that, in a classical deductive system $\vdash \Gamma, \bigwedge_{i\in I} A_i\ [\alpha, \delta]$ implies $\vdash \Gamma, A_j\ [\alpha, \delta]$ for every $j \in I$.*

b) Show that a) holds in an intuitionistic deductive system if $\bigwedge_{i\in I} A_i$ is positive or its components are all negative. Show that the restriction is necessary in the intuitionistic case.

c) Assume that $\bigvee_{i\in I} A_i$ *is positive or all of its components are negative and assume that no subformula of a formula in* Γ *is a* $\bigwedge$*-formula with more than one component. Show that, in an intuitionistic deductive system,* $\vdash \Gamma, \bigvee_{i\in I} A_i\ [\alpha, \delta]$ *implies* $\vdash \Gamma, A_j\ [\chi_\delta(\alpha), 0]$ *for some* $j \in I$. [Hint: first apply the Elimination Theorem to the given deduction.] *Show that this fails in in general in a classical deductive system.*

d) In CPL, assume that every conjunctive subformula of $\bigvee_{i\in I} A_i$ *or of a formula in* Γ *has only a finite number of components and that* $\vdash \Gamma, \bigvee_{i\in I} A_i\ [\alpha, \delta]$. *Show that there is a finite set* Δ *of components* A_j *of* $\bigvee_{i\in I} A_i$ *such that* $\vdash \Gamma, \Delta\ [\chi_\delta(\alpha), 0]$

e) Assume that there is a disjunctive formula $\perp$ *with no components and that the disjunctive formula* $A \vee B$ *with components* A *and* B *exists for each* A *and* B. *Define* $\neg A = \overline{A} \vee \perp$. *Show that* $\vdash A \vee \neg A$ *is deducible in the classical system.*

f) Make the same assumption as in e) and, moreover, assume that P *is a prime formula which occurs in no axiom set except* $\{P, \overline{P}\}$. *Show that* $P \vee \neg P$ *is not deducible intuitionistically.*

e) Call a deductive system consistent *iff there is no formula* A *such that both* $\vdash A$ *and* $\vdash \overline{A}$. *Show that the system is consistent iff the null set of formulas is not derivable.*

We mention some extensions of the Cut Elimination Theorem.

I. We can add to the rules of inference

AX′ $\qquad\qquad \Gamma, \underline{A}, \underline{\overline{A}}$

for arbitrary non-prime formulas A; where, in the intuitionistic case, Γ must consist entirely of negative formulas. (In the case of prime formulas A, each inference of this form is an instance of **AX**, by the Completeness Property.) Of course, by Exercise 14 a), there is a normal deduction of each instance of this rules of rank $2 \times |A|$. By adding the new rule, though, we are able to reduce ranks of deductions in some cases. The proof of the Elimination Lemma goes through as before, with the addition of one new possibility in Case 2b: namely, when Γ, A or $\Gamma', \overline{A}$ is an instance of **AX′**. By symmetry

we can assume the former. But then $\overline{A} \in \Gamma$ and so $\vdash \Gamma, \Gamma'[\beta, \delta]$ follows by applying weakening to $\mathcal{D}'$.

We will refer to deductions using $\mathbf{AX}'$ as *deductions in the extended sense.*

II. If P is a prime formula and $\langle B_i \mid i \in I \rangle$ is some family of formulas, then we could introduce the rule of inference

$$\mathbf{P} \qquad \frac{\cdots \Gamma_i, \underline{B_i} \cdots}{\Gamma, \underline{P}} \qquad (\Gamma_i \subseteq \Gamma)$$

where, in the intuitionistic case, all the sets must be i-sets. We may add $\mathbf{P}$ for any number of prime formulas P, providing that the only inferences having $\overline{P}$ as a pf are instances of $\mathbf{AX}$ with principle formulas P and $\overline{P}$. The proof of cut elimination goes through as before, with one new consideration in Case 1a) in the proof of the Elimination Lemma: namely,when Γ, P is obtained by $\mathbf{P}$. But in this case, $\Gamma', \overline{P}$ can only be an instance of $\mathbf{AX}$ with pf P and $\overline{P}$. But then $\overline{P} \in \Gamma'$ and so $\vdash \Gamma, \Gamma'[\beta, \gamma]$ by Weakening.

III. We may add to a deductive system inferences of the form $\mathbf{P}$ for one or more *arbitrary* formulas P and not just prime formulas, and without any restrictions on inferences in which $\overline{P}$ is a pf. We shall call such inferences *extra-logical inferences.* The proofs of the Elimination Lemma and Cut Elimination Theorem go through without change, except that, in defining the cut-degree of an inference, we ignore cuts

$$\frac{\Delta, P \qquad \Theta, \overline{P}}{\Gamma}$$

where Δ, P is obtained by $\mathbf{P}$. Thus, a normal deduction in this case is a deduction in which the only cuts are of this form, where Δ, P is obtained by $\mathbf{P}$. The Subformula Property in this case no longer holds: rather, we can only say that, all the formulas occuring in a normal deduction of Γ are subformulas of formulas in Γ or of P or $\overline{P}$, where P is pf of some extralogical inference.

3.3 First Order Predicate Logic

We have already described the formula systems CPL (Example 2) and IPL (Example 4). It is usual to take the axiom sets of both CPL and IPL to be the i-sets $\{A, \overline{A}\}$ for atomic formulas A. However, we will assume only that the axiom sets satisfy the Completeness and Cut Conditions. In particular, this will include First Order Predicate Logic with Identity. In this deductive system, a particular binary relation constant $=$ is singled out and we take the axiom sets to be the following i-sets:

$$A, \overline{A}$$

$$t = t$$

$$s \neq t, \overline{A(s)}, A(t)$$

for terms s and t and all prime formulas A and $A(b)$ (where $s \neq t$ is the complement of $s = t$). The augmented collection of axiom sets clearly still satisfies the Completeness and Cut Conditions. The only distinction between CPL and IPL is that deductions in IPL are restricted to i-sets, whereas deductions in CPL may involve arbitrary sets.

Notice that the only inferences that may have more than two premises are instances of $\bigwedge$ when the principal formula has an infinite number of components.

Exercise 16 *Prove that, in any deductive system, when and only when there are formulas with infinitely many components can deductions have infinite rank.*

The systems CPL and IPL, which we will denote collectively by PL, have formulas with infinitely many components and yet, as we ordinarily consider them, have only deductions of finite rank. The reason is that the rule of inference for $\forall x A(x)$ usually given is not $\bigwedge$ but

$$\forall \qquad \frac{\Delta, A(b)}{\Gamma, \forall x A(x)} \qquad (b \notin \Delta, \Delta \subseteq \Gamma)$$

$b \notin \Delta$ means that b does not occur in a formula in Δ. Of course the other conjunctive formulas of this system, viz. $\top$ and binary conjunctions $A \wedge B$

have 0 and 2 components, respectively; and so replacing $\bigwedge$ by $\forall$ does mean that all deductions are finite. Let $\vdash' \Gamma[m,n]$ mean that there is a deduction of Γ in PL of rank $\leq m$ and cut-degree $\leq n$, but with $\bigwedge$ replaced by $\forall$ in the case of pf of the form $\forall x A(x)$. Note that all deductions in the sense of $\vdash'$ in PL have finite cut-degree, since only a finite number of formulal occur in a deduction and all formuilas have finite rank. The connection between the two versions of deduction in PL is given by

Exercise 17 *a) Prove by induction on α that $\mathcal{D} \vdash \Gamma[\alpha, \delta]$ implies $\vdash' \Gamma[m,n]$, where $m < \omega$, $m \leq \alpha$ $n < \omega$ and $n \leq \delta$.*

b) Prove that, if the free variable b does not occur in any formula in Γ and $\vdash' \Gamma, A(b)[m,n]$, then for every term t, $\vdash' \Gamma, A(t)[m,n]$.

c) Prove (using b)) that $\vdash' \Gamma[m,n]$ implies $\vdash \Gamma[m,n]$.

By c) and a), we have [Gentzen, 1935]:

Corollary 3 (Gentzen's Hauptsatz) *$\vdash' \Gamma[m,n]$ implies $\vdash' \Gamma[2_n^m, 0]$.*

As a matter of fact, Gentzen did not provide the bound on the rank of the cut-free deduction; but it was his use of ordinals in connection with his consistency proof for PA that leads to this bound.

Of course, a direct proof of Corollary 3 is easily obtained from our proof above for $\vdash$. The only modification that is needed is in the proof of the Elimination Lemma in the case that $A = \exists x A(X)$ is pf of the inference of $\mathcal{D}$ and $\overline{A} = \forall x \overline{A(x)}$ is pf of $\mathcal{D}'$. So we have

$$\frac{\Delta, A(t)}{\Gamma, A}$$

and

$$\frac{\Delta', \overline{A(b)}}{\Gamma, \overline{A}}$$

where b does not occur in Δ'. But now apply Exercise 17 b) to this latter inference to obtain

$$\frac{\Delta', \overline{A(t)}}{\Gamma, \overline{A}}$$

with no increase in rank or cut-degree, which puts us back on track in the proof of the Elimination Lemma.

Exercise 18 *Prove that every theorem of classical or intuitionistic first-order predicate logic, in your favorite formalization of this system, is deducible (or, strictly speaking, its unit set is deducible) in the deductive system CPL or IPL, respectively.* [In the case of CPL, it is not hard to simply prove directly that the normal rules of inference are semantically complete: If, in every structure which interprets the non-logical constants in the formulas in Γ and for every assignment of values in the domain of the structure, some formula in Γ is satisfied, then there is a normal deduction of Γ.]

As a consequence of Exercise 15 *c*) and *d*) and Lemma 7, we have

Corollary 4 ($\bigvee$-Instantiation) *a) If no subformula of Γ is a $\wedge$-formula or a $\forall$-formula, then in IPL*

i) $\vdash \Gamma, A \vee B\ [\alpha, 0]$ implies $\vdash \Gamma, A\ [\alpha, 0]$ or $IPL \vdash \Gamma, B\ [\alpha, 0]$.

ii) $\vdash \Gamma, \exists x A(x)\ [\alpha, 0]$ implies $\vdash \Gamma, A(t)\ [\alpha, 0]$ for some term t.

b) In CPL, $\vdash \Gamma, A \vee B\ [\alpha, 0]$ implies $\vdash \Gamma, A, B\ [\alpha, 0]$.

c) If no subformula of Γ or $A(b)$ is a $\forall$-formula, then in CPL, $\vdash \Gamma, \exists x A(x)[\alpha, 0]$ implies $\vdash \Gamma, A(t_0), \ldots, A(t_n)[\alpha, 0]$ for some list $t_0, \ldots, t_n$ of terms.

3.4 First Order Arithmetic

The formula systems of CPA, IPA, CPA^* and IPA^* have already been described in Examples 3 and 4. We will refer to CPA and IPA collectively as PA and to CPA^* and IPA^* as PA^*.

The axiom system of PA consists of the unit set of each defining equation of a primitive recursive function, the unit sets

$$0 \neq S(t)$$

$$S(s) \neq S(t), s = t$$

$$t = t$$

and

$$U(t), \bar{U}(t)$$
$$s \neq t, \overline{A(s)}, A(t)$$

for each prime formula $A(b)$. To the rules of inference of PL we add the rule of Mathematical Induction

$$\textbf{MI} \qquad \frac{\Delta, A(\bar{0}) \quad \Theta, \overline{A(b)}, A(S(b))}{\Gamma, A(t)} \qquad (\Delta, \Theta \subseteq \Gamma, b \notin \Theta)$$

In the inference **MI**, the mf are $A(\bar{0}), \overline{A(b)}$ and $A(S(b))$, and the pf is $A(t)$, which all have the same rank as the *induction formula* $A(b)$.

The maximum of the numbers $n+1$, where n is the rank of an induction formula in the deduction will be called the *induction-degree* of the deduction. Note that the extension III of the Cut Elimination Theorem applies to PA to show that, if there is a deduction of Γ in PA, then there is one whose cut-degree is $\leq$ its induction degree.

Notice that every deduction in PRA is a deduction in PA.

The axiom system of PA^* consists of

$$\bar{n} = \bar{n}$$

$$\bar{m} \neq \bar{n}$$

when $m \neq n$,

$$U(\bar{n}), \bar{U}(\bar{n})$$

Recall that in PA^* we identify closed terms with the same value. So every axiom set of sentences in PA is an instance of **AX** in PA^*. Given a deduction in PA, it will contain only a finite number of instances of **MI** and only a finite number of cuts. Let Γ be a set of formulas of PA. A *numerical instance* of Γ is a set of sentences that results from substituting a numeral for each free variable in the formulas of Γ (the same numeral for each occurence of the variable in each formula in Γ).

Lemma 9 *If there is a deduction in PA of Γ of induction-degree k, then each numerical instance of Γ has a deduction in PA^* of rank $< \omega \times \omega$ and cut-degree k.*

Proof. First, we can assume that the cut-degree of the given deduction $\mathcal{D}$ is $\leq k$. We show by induction on n that, if $\mathcal{D}$ is of rank n, then there is a deduction of each numerical instance of Γ of rank $\leq \omega \times n$ and cut-degree $\leq k$. Let Γ' be the numerical instance. If Γ is an instance of **AX** in PA, then Γ' clearly includes an axiom of PA^*. If the last inference in $\mathcal{D}$ is other than an instance of **AX** or **MI**, then the result follows immediately by the induction hypothesis. So assume that the last inference is

$$\frac{\Delta, A(\bar{0}) \quad \Theta, \overline{A(b)}, A(S(b))}{\Lambda, A(\bar{p})}$$

so that $\Gamma' = \Lambda', A(\bar{p})'$. By the induction hypothesis, there are deductions of $\Delta', A(\bar{0})'$ and $\Theta', \overline{A(\bar{m})}', A(\overline{m+1})'$ all of some rank $\alpha < \omega \times n$ and cut-degree $\leq k$, for each m. We obtain a deduction of Γ' of rank $\leq \alpha + p < \omega \times n$ and cut-degree $\leq k$ by induction on p. If $p = 0$, then we obtain Γ‘ from $\Delta', A(\bar{0})'$ by Weakening. Let $p = r + 1$. We have a deduction of $\Lambda', A(\bar{r})'$ of rank $\leq \alpha + r$ and so, by a cut with cut-formula $A(\bar{r})'$, we have a deduction of Γ' of rank $\leq \alpha + p$. QED

So every deduction in PA of a set Γ of sentences can be transformed into a deduction in PA^* of rank $< \omega \times \omega$ and some finite cut degree n and so, by the Cut Elimination Theorem, into a normal deduction of Γ of rank

$$2_n^{\omega \times \omega} < \epsilon_0$$

where $\epsilon_0 = \psi_1(5) = 2 - exp_1(2) = \omega - exp_1(1)$ is the least fixed point of ψ greater than ω.

Theorem 3 (Schütte, 1951) *If Γ is a set of sentences deducible in PA, then it has a normal deduction in PA^* of rank $< \epsilon_0$.*

Using the primitive recursive functions *pred* and *sgn*, defined by

$$predf(0) = 0 \qquad pred(S(t)) = S(0)$$

$$sgn(0) = 0 \qquad sgn(S(t)) = S(0).$$

we can construct for each formula A of PRA a term $[A]$ with the same free variables, such that $A \longleftrightarrow [A] = 0$. First, define

$$s - 0 = s \qquad s - S(t) = pred(s - t)$$

$$|s - t| = s - t + t - s.$$

Now define

$$[s = t] = (s - t) + (t - s)$$

$$[\neg A] = 1 - sgn[A]$$

$$[A \wedge B] = [A] \times [B].$$

The preceding discussion of first order arithmetic is independent of whether we are speaking of CPA or IPA. The following, however, is significant only for CPA. The hierarchy of Π^0_n and Σ^0_n formulas of PA are defined as follows: The Π^0_0 are the formulas without quantifiers and not containing the predicate constant U. The Σ^0_n formulas are the complements of Π^0_n formulas. The Π^0_{n+1} formulas are those of the form $\forall x A(x)$, where $A(b)$ is Σ^0_n. Since the Π^0_0 formulas are equations, the formulas in Π^0_n are all of rank n. Every formula of PA is logically equivalent in CPL to a Π^0_n formula and the hierarchy is not degenerate: for every $n > 0$, there is a Π^0_n formula which is not Σ^0_n.

By $\Pi^0_n - PA$ we will mean PA but with mathematical induction **MI** restricted to induction formulas which are Π^0_n. By the discussion above, this means that each deduction in $\Pi^0_n - PA$ transforms into a deduction in PA^* of rank $< \omega \times \omega$ and cut-degree $n + 1$. So, noting that

$$2^{\omega \times \omega} = \omega^\omega, \quad 2_2^{\omega \times \omega} = 2^{\omega^\omega} = \omega^{\omega^\omega}$$

and in general for $n > 1$

$$2_n^{\omega \times \omega} = \omega_n^\omega$$

We have

Theorem 4 *If Γ is a set of sentences deducible in $\Pi^0_n - PA$, then Γ has a normal deduction in PA^* of rank $< \omega^\omega_n$.*

Corollary 5 *Let Γ be a set of equations in PA, i.e. in PRA, and let it be deducible in $\Pi^0_0 - PA$. Then*

a. Γ is deducible in PRA.

b. Γ has a normal deduction of rank $< \omega^\omega$ in PA^*.

3.5 Derivability of Induction

Let Σ be a deductive system which is a *faithful extension* of PA^*. This means, first, that every formula A of PA^* is a formula of Σ and that the components, complement, sign, and rank of A are the same in Σ as they are in PA^* and, secondly, that, for every set of formulas of PA^*, it is an axiom set of PA^* iff it is an axiom set of Σ.

Let a formula of PA with just the free variables a and b be given. We denote the formula by $a \prec b$. So for any close terms s and t, $s \prec t$ is a formula of PA^*. We assume that

$$\bar{m} \prec \bar{n} \in \Sigma_A$$

defines a well-ordering of the natural numbers, which we shall also denote by $\prec$. For every n, let $o(n)$ denote the ordinal of the set of predecessors of n in the ordering $\prec$.

Recall that U and $\overline{U}$ are the unary relation constants in PA. Let

$$H = H(U) = \forall x[\forall y(y \prec x \longrightarrow U(y)) \longrightarrow U(x)]$$

Thus, H expresses the fact that U is *hereditary* with respect to $\prec$ and, since the only axiom set containing formulas containing U or $\overline{U}$ are of the form $U(\bar{n}), \overline{U}(\bar{n})$,

$$\overline{H}, \forall x U(x)$$

expresses in Σ the principle of induction on $\prec$ and

$$\overline{H}, \forall x[x \prec \bar{n} \longrightarrow U(x)]$$

expresses induction on $\prec$ up to n. We will prove the

Theorem 5 (Induction Theorem) *Let $\omega > p > 0$. Then*

$$\vdash_\Sigma \overline{H}, U(\bar{n})[\alpha, p] \Rightarrow o(n) \leq \psi^p(\alpha)$$

Proof. We can assume that Σ is a classical system, since its deductions include those of the corresponding intuitionistic system. Let Σ' result from adding to Σ all inferences of the form

$$\mathbf{U}(\bar{\mathbf{n}}) \qquad \frac{\cdots \Gamma, \underline{U(\bar{m})} \cdots}{\Gamma, \underline{U(\bar{n})}} \qquad (m \prec n)$$

We can assume that all deductions in Σ and Σ' are contractions. As we noted above (see II), the Elimination Lemma and so the Cut Elimination Theorem remain valid for Σ'.

Lemma 10 *Let $p > 0$. Then for every Γ*

$$\vdash_\Sigma \overline{H}, \Gamma\ [\alpha, p]$$

implies

$$\vdash_{\Sigma'} \Gamma\ [2 \times \alpha, p]$$

The proof is by induction on α. If the final inference is other than one with $\overline{H}$ as pf, the result is immediate. Let the last inference have $\overline{H}$ as pf. Hence, for some m and $\beta < \alpha$

$$\vdash_\Sigma \overline{H}, \forall x[x \prec \bar{m} \longrightarrow U(x)] \wedge \overline{U}(\bar{m}), \Gamma\ [\beta, p]$$

By two applications of part a) of the Reduction Lemma and one application of part b)

$$\vdash_\Sigma \overline{H}, \bar{k} \not\prec \bar{m}, U(\bar{k}), \Gamma\ [\beta, p]$$

for each k and

$$\vdash_\Sigma \overline{H}, \overline{U}(\bar{m}), \Gamma\ [\beta, p]$$

For each $k \prec m$, $\{\bar{k} \prec \bar{m}\}$ is an axiom set and so by Part c) of the Reduction Lemma,

$$\vdash_\Sigma \overline{H}, U(\bar{k}), \Gamma\ [\beta, p]$$

for all $k \prec m$. So by the induction hypothesis

$$\vdash_{\Sigma'} U(\bar{k}), \Gamma\ [2 \times \beta, p]$$

for all $k \prec m$; and hence by $\mathbf{U}(\bar{\mathbf{m}})$

$$\vdash_{\Sigma'} U(\bar{m}), \Gamma\ [2 \times \beta + 1, p]$$

Again by the induction hypothesis

$$\vdash_{\Sigma'} \overline{U}(\bar{m}), \Gamma\ [2 \times \beta, p]$$

So, by a cut with cut-formula $U(\bar{m})$ of rank 0,

$$\vdash_{\Sigma'} \Gamma\ [2 \times \alpha, p]$$

The remainer of the proof of the Induction Theorem consists in showing that in Σ'

$$\vdash U(\bar{n})[\beta, 0] \Longrightarrow o(n) \leq \beta$$

This will follow from

$$\vdash_{\Sigma'} U(\bar{n}_0), \ldots, U(\bar{n}_k)[\beta, 0] \Longrightarrow Min_{i \leq k}\ o(\bar{n}_i) \leq \beta$$

But a normal deduction of $U(\bar{n}_0), \ldots, U(\bar{n}_k)$ in Σ' can only involve inferences $\mathbf{U}(\bar{\mathbf{n}})$, and so the conclusion immediately follows.

$\forall x \prec \bar{n} A(x)$ is an abbreviation for $\forall x(x \prec \bar{n} \longrightarrow A(x)$.

Corollary 6 *Let* $\omega > p > 0$.

$$\vdash_{\Sigma} \overline{H}, \forall x \prec \bar{n} U(x)[\alpha, p] \Longrightarrow o(n) \leq \psi^p(\alpha) + 1$$

We need only note that, by Reduction,

$$\vdash \overline{H}, \forall x \prec \bar{n} U(x)[\alpha, p] \Longrightarrow \vdash \overline{H}, U(\bar{m})[\alpha, p]$$

for all $m \prec n$.

Corollary 7 *In* PA

$$\vdash \overline{H}, \forall x \prec \bar{n} U(\bar{n}) \Longrightarrow o(n) < \epsilon_0$$

f

For the hypothesis implies that in PA^*

$$\vdash \overline{H}, \forall x \prec \bar{n} U(\bar{n})[< \alpha, 1]$$

for some $\alpha < \epsilon_0$. But then $\psi(\alpha) < \epsilon_0$.

Corollary 8 *In* $\Pi^0_k - PA$

$$\vdash \overline{H}, \forall x \prec \bar{n} U(\bar{n}) \Longrightarrow o(n) < \omega^{\omega}_{k+1}$$

The antecedent implies that in PA^*

$$\vdash \overline{H}, U(\bar{m})[\omega^2, k+1]$$

for all $m \prec n$.

By a *primitive recursive relation*, we mean a relation whose characteristic function is primitive recursive (pr). With certain countable limit ordinals α, we may associate a primitive recursive relation $\prec_\alpha$ which well-orders the set of natural numbers and such that the well-ordering has order type α. (Of course, there is no unique such pr well-ordering.) We will call such ordinals *primitive recursively represented* or *pr represented* or *prr*. So $a \prec_\alpha b$ can be expressed by a formula in PA with the free variables a and b. We denote this formula by $a \prec_\alpha b$, and it will have the property that, for all m and n, $PA \vdash \bar{m} \prec_\alpha \bar{n}$ or $PA \vdash \bar{m} \not\prec_\alpha \bar{n}$ (where $a \not\prec_\alpha b$ denotes the complement of $a \prec_\alpha b$.

We are interested in the question: for what prr ordinals α is transfinite induction on $\prec_\alpha$ a theorem of PA? We express induction on $\prec_\alpha$ applied to a formula $A(b)$ by

$$\mathbf{J}_x(A(x), \alpha) \ := \ \forall x[\forall y \prec_\alpha xA(y) \longrightarrow A(x)] \longrightarrow \forall x A(x)$$

In particular, $J_\alpha := J_x(U(x), \alpha)$ expresses induction on $\prec_\alpha$ for an arbitrary property of numbers: if it is derivable in PA, then so is $J_x(A(x), \alpha)$ for every formula $A(x)$ of PA.

PA_0 will denote PA but without the principle of mathematical induction. Since we are interested in what instances of mathematical induction are needed to derive induction on α, we want to consider only deductions in PA_0.

The least limit ordinal ω is of course prr, where for example $\prec_\omega$ can be taken to be the natural ordering $<$ of the numbers.

Proposition 13 *$J_x(A(x), \omega)$ is a theorem of $\Pi^0_n - PA$ for every Π^0_n formula $A(b)$.*

Let $B(b) = \forall x < bA(x) \longrightarrow A(b)$. The sets

$$B(0) \qquad \overline{\forall x[B(x) \longrightarrow B(S(x))]}, \overline{B(b)}, B(S(b))$$

are deducible in $\Pi^0_0 - PA$. Hence, by **MI** with induction formula $B(b)$, we have

$$\overline{\forall x[B(x) \longrightarrow B(S(x))}, B(b)$$

and hence

$$\overline{B(0)}, \overline{\forall x[B(x) \longrightarrow B(S(x))]}, \forall x B(x)$$

Every ordinal $\beta < 2^\alpha$ is uniquely of the form $2^{\beta_1} + \cdots + 2^{\beta_n}$ where $\beta_1 > \cdots > \beta_n$. Moreover, if $\gamma = 2^{\gamma_1} + \cdots + 2^{\gamma_m}$, where $\gamma_1 > \cdots > \gamma_m$, then $\beta < \gamma$ iff $\langle \beta_1, \ldots, \beta_n \rangle \prec \langle \gamma_1, \ldots, \gamma_m \rangle$, where $\prec$ is the lexicographical ordering defined in Chapter 2 (Definition 2). This suggests the following definition of $\prec_{2^\alpha}$.

The finite sequence $\langle k_1, \ldots, k_n \rangle$ can be coded by the number

$$2^{k_1+1} \times 3^{k_2+1} \times \cdots \times p_{n-1}^{k_n+1}$$

where $p_0 = 2, p_1 = 3, \ldots, p_{n-1}, \ldots$ is the sequence of all the prime numbers in increasing order. We are including the case of the null sequence of numbers ($n = 0$). In this case, the code is 0. When $k_n \prec_\alpha \cdots \prec_\alpha k_1$, we call the code an *$\alpha$-sequence number*. The function f which enumerates the α-sequence numbers in their natural order is pr. If $f(k)$ is the above α-sequence number, then we write

$$k = [k_1, \ldots, k_n]$$

Moreover, the function $k_* = n$, the length of the coded sequence, is pr as is the function defined by $k_{[i]} = k_{i+1}$ if $i < k_*$ and $k_{[i]} = 0$ if $i \geq k_*$. Now we define the pr representation $\prec_{2^\alpha}$ of 2^α from $\prec_\alpha$ by

$$[k_1, \ldots, k_n] \prec_{2^\alpha} [k'_1, \ldots, k'_{n'}]$$

iff

- $n < n'$ and $k_i = k'_i$ for all $i = 1, \ldots, n$, or
- there is an j with $0 < j \leq n$ such that $j \leq n'$, $k_i = k'_i$ for $0 < i < j$ and $k_j \prec_\alpha k'_j$.

In what follows, let $\prec$ be $\prec_{2^\alpha}$. If $k = [k_1, \ldots, k_n]$, then we write $k^{(m)} = [k_1, \ldots, k_p]$, where $p = Min(m, n)$. (So $k^{(0)} = 0$.) Let

$$B(a, b) = \forall y \prec b^{(a)} U(y) \longrightarrow U(b)$$

$$C(b) = \forall x u[u \leq x_* \wedge x_{[u]} = b \longrightarrow B(u, x)]$$

Proposition 14 *In PA_0 J_{2^α} can be deduced from $J_x(C(x), \alpha)$.*

Assume $J_x(C(x), \alpha)$ and

$$\forall x[\forall y \prec xU(y) \longrightarrow U(x)] \tag{3.5}$$

We need to derive $\forall x U(x)$ from these assumptions. It will suffice to deduce $\forall x C(x)$, since $C(b_{[0]}) \longrightarrow B(0, b)$ and $B(0, b) \longrightarrow U(b)$. (The latter follows from $b^{(0)} = 0$.) So assume

$$\forall v \prec_\alpha dC(v) \tag{3.6}$$

We need to deduce $C(d)$ from this. For then we will have $\forall x[\forall v \prec_\alpha xC(v) \longrightarrow C(x)]$ and $\forall x C(x)$ will follow using $J_x(C(x), \alpha)$. Assume

$$a \leq b_* \wedge d = b_{[a]} \tag{3.7}$$

We need to deduce $B(a, b)$. So assume

$$\forall y \prec b^{(a)} U(y) \tag{3.8}$$

From all of these assumptions we need to derive $U(b)$. Since $a \geq b_*$ implies $b^{(a)} = b$, it follows from (3.5) and (3.8) that

$$a = b_* \longrightarrow U(b) \tag{3.9}$$

Assume that $a < b_*$. If $a + 1 = b_*$, then $B(a + 1, b)$ follows from (3.5) and $b^{(a+1)} = b$. If $a + 1 < b_*$, then $b_{[a+1]} \prec_\alpha b_{[a]}$; and so $C(b_{[a+1]})$ by (3.6) and (3.7). So in any case $B(a + 1, b)$, i.e.

$$\forall y \prec b^{(a+1)} U(y) \longrightarrow U(b)$$

So we need to deduce $\forall y \prec b^{(a+1)} U(y)$. Let $e \prec b^{(a+1)}$. If $e \prec b^{(a)}$, then $U(e)$ follows by (3.8). So we can assume that $b^{(a)} \preceq e \prec b^{(a+1)}$ which means that

$$e^{(a)} = b^{(a)} \wedge e_{[a]} \prec_\alpha b_{[a]}$$

$C(e_{[a]})$ by (3.6) and (3.7). We are assuming $a < b_*$. So, since $e^{(a)} = b^{(a)}, a \leq e_*$. So $B(a, e)$ and therefore by (3.8), $U(e)$. So we have deduced $\forall y \prec b^{(a+1)} U(y)$ and therefore

$$a < b_* \longrightarrow U(b)$$

Thus, we have $U(b)$.

Now, backing up: $B(a,b)$ follows from (3.5), (3.6) and (3.7). So

$$\forall x[\forall v \prec_\alpha xC(v) \longrightarrow C(x)]$$

follows from (3.5). So $\forall x C(x)$ follows from (3.5) using $J_x(C(x),\alpha)$. So (3.5) implies $B(0,b)$, i.e (3.5) implies $U(b)$. QED

Definition 9 *The* proof-theoretic ordinal *of a system Σ is the least ordinal β such that for every prr $\alpha < \beta$, J_α is a theorem of Σ.*

Theorem 6 *a. The proof-theoretic ordinal of PA is ϵ_0.*

b. The proof-theoretic ordinal of $\Pi^0_n - PA$ is ω^ω_{n+1}.

We have already proved that these are upper bounds on the α such that J_α is deducible. To prove that they are least upper bounds, it suffices to prove this for b. Let

$$J(\Pi^0_n,\alpha) \longrightarrow J(\Pi^0_m,\beta)$$

mean that for each Π^0_m formula $A(b)$ there is a Π^0_n formula $B(b)$ such that

$$J_x(B(x),\alpha) \longrightarrow J_x(A(x),\beta)$$

is deducible in $\Pi^0_0 - PA$.

i) We observe without proof that

$$J(\Pi^0_n,\omega) \longrightarrow J(\Pi^0_n,\omega^k)$$

for each $k < \omega$.

A fine proof of this is given by G. Mints.

ii) The result of substituting a Π^0_n formula $A(b)$ in $C(b)$ for $U(b)$ is equivalent in CPL to a Π^0_{n+1} formula. So

$$J(\Pi^0_{n+1},\alpha) \longrightarrow J(\Pi^0_n,2^\alpha)$$

Hence

$$J(\Pi^0_n,\omega^{k+1}) \longrightarrow J(\Pi^0_{n-1},\omega^{(\omega^k)}) \longrightarrow \cdots \longrightarrow J(\Pi^0_0,\omega^k_{n+1})$$

Chapter 4

Theory of Types and Natural Deduction

The second branch of proof theory stemming from [Gentzen, 1935] concerns the system of Natural Deduction. But I think that the significance of this system is best understood in a more general context: the theory of types of Curry and Howard.

4.1 Types

There are two leading ideas behind the theory of types. The first idea implies a kind of mathematical essentialism: objects of mathematics always exist as objects of some particular *type*. A type is given by specifying definite rules for constructing objects of that type (*Introduction Rules*) and reasoning about them (*Elimination Rules*). The notion of a type in something like our sense was first introduced in Bertrand Russell's *Principles of Mathematics*, Appendix B. The types that Russell considered are those built up from the type of individuals by passing from a type A to its *power type*

$$\mathbf{P}(A)$$

and from types A and B to their *binary product type*

$$A \times B$$

$\mathbf{P}(A)$ is the type of all sets of objects of type A and $A \times B$ is the type of all ordered pairs ('couples with sense', to use Russell's term) (s, t) with s of type A and t of type B.

The notion of a type is a primitive notion and should not be confused with the notion of a set. The notion of set is first introduced in the context *set of objects of type* A for some type A. The question of whether an object of type A is in a set s of objects of type A, i.e. whether or not it is an element of the object s of type $\mathbf{P}(A)$, can be non-trivial: the set s may be defined as the extension of some logically complex property of objects of type A. But the question of whether a given object is of a given type A is always a trivial question: the notion of 'object' is type-ambiguous: an object *is* an object of some type and to be given an object implies being given its type. Of course, associated with any type A is the set of type $\mathbf{P}(A)$ of all objects of type A; but to speak of this, or any, set is to presuppose the type of objects from which the set is to be constituted. The initial (Russell's) motivation for the theory of types was the realization that the anti-essentialist conception of Frege is incoherent: the totality of 'all' mathematical objects cannot itself be treated in mathematics and that, consequently, the easy answer to the question "set of what?" given by "set of any objects whatsoever" is inadequate.

It is also important to distinguish the primitive notion of an ordered pair (s,t) from its representations such as $\{\{s\},\{s,t\}\}$ in axiomatic set theory. When the types A of s and B of t are distinct, then there is no set $\{s,t\}$. There is indeed, as we shall see, a type $A+B$ with natural embeddings $s \mapsto s'$ and $t \mapsto t'$ of A and B into $A+B$, so that the 'set' $\{s,t\}$ can be represented by the object $\{s',t'\}$ of type $\mathbf{P}(A+B)$. But the construction of $A+B$ already involves the notion of an ordered pair.

A further type-forming operation was introduced in Hilbert's "Über das Unendliche" [Hilbert, 1926]. Namely, from given types A and B, we may introduce the *exponential type*

$$B^A$$

of all B-valued functions defined on A.

We may generalize the constructions A^B and $A \times B$. Let A be a type and let F be a type-values function defined on A. That is, for each object s of type A, $F(s)$ is a type. We introduce the *Product Type*

$$\Pi x\!:\!A\ F(x)$$

of all functions f defined on A such that, for each s of type A, $f(s)$ is of type $F(s)$. We also introduce the *Sum* or *Disjoint Union* type

$$\Sigma x\!:\!A\ F(x)$$

of all pairs (s,t) where s is of type A and t is of type $F(s)$. Notice that when F is a constant function $F(s) = B$ for all s of type A, then we obtain Exponentiation and Binary Product as special cases:

$$B^A = \Pi x{:}A\ F(x) \qquad A \times B = \Sigma x{:}A\ F(x)$$

We introduce the *Null Type*, which will be denoted by

$$\mathbf{0}$$

Of course, there will be other types which fail to have objects. (For example, there are no objects of type $\mathbf{0} \times A$.) But the notion of type is not extensional and, in particular, not every type which fails to have objects is identified with $\mathbf{0}$. $\mathbf{0}$ is distinguished from these other 'empty' types by being *specified* as null.

We also introduce the *One-Object Type*

$$\mathbf{1}$$

whose object we denote by $\mathbf{o}$, and the *Two-Object Type*

$$\mathbf{2}$$

whose objects, *Truth* and *Falsety*, are denoted respectively by

$$\top \qquad \bot$$

We also introduce the type-valued function $\mathbf{T}$ defined on $\mathbf{2}$ with

$$\mathbf{T}(\top) = \mathbf{1} \qquad \mathbf{T}(\bot) = \mathbf{0}$$

The power type can now be introduced by definition:

$$\mathbf{P}(A) = \mathbf{2}^A$$

For t of type A and s of type $\mathbf{P}(A)$, we define

$$t \epsilon_A s := \mathbf{T}(s(t))$$

So, when $s(t)$ is True, $t \epsilon_A s$ is the 'true' type $\mathbf{1}$; and when $s(t)$ is False, it is the 'false' type $\mathbf{0}$.

4.2 Propositions as Types (PAT)

Speaking of 'true' and 'false' types leads us to the other leading idea of type theory: we may regard a proposition as a type.

A proposition is the type of its proofs

A proposition is true when there is a proof of it. On this conception of propositions, the type-forming operations we have just introduced assume the character of familiar logical operations. This conception of mathematical propositions and of the logical operations goes back to Brouwer's intuitionism and, more particularly, to Heyting's analysis of intuitionistic logic. We shall not restrict ourselves to intuitionistic logic, however, but will regard this as just a restricted case of proofs in general.

On the PAT conception, the *implication*

$$A \longrightarrow B$$

simply becomes B^A. For a proof f of $A \longrightarrow B$ should yield, given any proof p of A, a proof of B; and moreover, there is no reason to require anything more of a proof of $A \longrightarrow B$. Hence, the functions of type B^A are precisely the proofs of $A \longrightarrow B$.

More generally, the *universal quantification*

$$\forall x : A F(x)$$

where A ia a type and F is a propositional function (i.e. a type-valued function) defined on A, is just the product type $\Pi x : A F(x)$. For any proof of $\forall x : A F(x)$ should yield, for any object t of type A, a proof of $F(t)$. And, conversely, nothing further should be required of an object of type $\Pi x : A F(x)$ for it to be a proof of $\forall x : A F(x)$.

The *conjunction*

$$A \wedge B$$

becomes the binary product $A \times B$. For a proof of $A \wedge B$ should be nothing more or less than a pair consisting of a proof of A and a proof of B, i.e. an object of type $A \times B$.

More generally, the *existential quantification*

$$\exists x : A\ F(x)$$

where F is a propositional function defined on A, becomes the sum $\Sigma x : A\ F(x)$. For a proof of $\exists x : AF(x)$ should be nothing more or less than a pair consisting of an object t of type A and a proof of $F(t)$.

From now on, expressing a certain bias, perhaps, we will use the logical notations $\forall, \exists, \wedge$, etc. instead of the corresponding $\Pi, \Sigma, \times$, etc.

On the PAT conception, the type $\mathbf{0}$ becomes the *Absurd Proposition*. Of course, as we noted above, this does not mean that there are no other false propositions, i.e. propositions without proofs, only that $\mathbf{0}$ is *introduced* as an absurdity. In the same sense, we may regard $\mathbf{1}$ as the *Necessary Proposition*, since a proof $\mathbf{o}$ of it is explicitly given. The connection between the truth values $\top$ and $\bot$ and the necessary and absurd propositions is given by the propositional function $\mathbf{T}$. Namely, $\mathbf{T}(\top)$ is $\mathbf{1}$ and $\mathbf{T}(\bot)$ is $\mathbf{0}$.

The *Negation* $\neg A$ of the proposition A is defined by

$$\neg A := A \longrightarrow \mathbf{0}$$

Once one accepts the PAT conception, this definitions seems forced. The minimal condition for a type A^* to be a negation of A would be that A^* and A together imply every proposition. But that will be so precisely if they together imply $\mathbf{0}$. So the minimal condition is $A^* \wedge A \longrightarrow \mathbf{0}$ or $A^* \longrightarrow (A \longrightarrow \mathbf{0})$. So $A \longrightarrow \mathbf{0}$ is the weakest type with the property in question.

But this may still seem an unreasonable analysis of negation. To say that A is true is to say that there is an object of type A. To say that it is false, therefore, is to say that there is no object of type A. But that is not an existence statement, whereas the assertion of $\neg A$ is. However, we may question whether the proposition that there is no object of type A has any precise mathematical meaning *other than* that expressed by $\neg A$. Thus, the argument may be turned around: to assert that there is no object of type A is nothing more than to assert $\neg A$.

We define *bi-implication* by

$$A \longleftrightarrow B := (A \longrightarrow B) \wedge (B \longrightarrow A)$$

Given propositions A and B, we define the propositional function $\langle A, B\rangle$ on $\mathbf{2}$ by

$$\langle A, B\rangle(x) = (\mathbf{T}(x) \longrightarrow A) \wedge (\neg\mathbf{T}(x) \longrightarrow B)$$

Thus $\langle A, B\rangle(\top)$ is $(\mathbf{1} \longrightarrow A) \wedge (\neg\mathbf{1} \longrightarrow B)$, which in ordinary propositional logic is equivalent to A. Similarly, $\langle A, B\rangle(\bot)$ is $(\mathbf{0} \longrightarrow A) \wedge (\neg\mathbf{0} \longrightarrow B)$, which is equivalent to B. So, assuming that we introduce the means for constructing the proofs of ordinary propositional logic, we will have

$$\langle A, B\rangle(\top) \longleftrightarrow A \qquad \langle A, B\rangle(\bot) \longleftrightarrow B. \tag{4.1}$$

So we may express the *Disjunction* $A \vee B$ by

$$A \vee B := \exists x{:}\mathbf{2}\ \langle A, B\rangle(x)$$

Considered as a type, we could denote $A \vee B$ by $A + B$. An object of type $A \vee B$ is a pair (s, t), where s is of type $\mathbf{2}$ and t is a proof of $\langle A, B\rangle(s)$. By (4.1), a proof t of A yields a proof $(\top, f(t))$ of $A \vee B$ and a proof t of B yields a proof $(\bot, g(t))$ of $A \vee B$, where f is of type $A \longrightarrow \langle A, B\rangle(\top)$ and g is of type $B \longrightarrow \langle A, B\rangle(\bot)$. So $t \mapsto (\top, f(t))$ and $t \mapsto (\bot, g(t))$ are the embeddings of A and B in $A + B$ mentioned in §1.

We could likewise have defined a different operation of conjunction $A\&B$ by

$$A\&B := \forall x{:}\mathbf{2}\ \langle A, B\rangle(x)$$

Remark 4 *There is certainly a difference between propositions and types in regard to our interests in them. In general, with a proposition our concern is simply whether or not it is provable, not with the different kinds of proofs that it might have; whereas with a type A, we are likely to be interested in what kinds of objects of type A there are. So perhaps we shouldn't think of all types as propositions. (For example,* $\mathbf{N}$ *is a pretty uninteresting proposition, proved by each of the natural numbers.) But the issue is not really important. For if within the (perhaps) more general domain of types we can prove a proposition A (that we agree* is *a proposition), even using types that are not recognized as propositions, then it is difficult to see on what grounds we would reject A as a logical truth. For example, one might want to reject the identification of* $\exists x{:}A\ F(x)$ *with the disjoint union* $\Sigma x{:}A\ F(x)$ *and adopt Frege's definition of it as* $\neg\forall x{:}A\neg F(x)$ *instead, on the grounds that the disjoint union is not really a proposition. Of course, this proposal only makes sense in classical logic, since* $\exists \longleftrightarrow \neg\forall\neg$ *is not valid intuitionistically. But in the classical theory of types there is an object of type* $\Sigma x : A\ F(x) \longleftrightarrow \neg\forall x : A\ \neg F(x)$*; and so any theorem of logic based on one of these definitions is a law of logic*

based on the other. An important example of this which we will discuss later is the Axiom of Choice in the form

$$\forall x{:}A\ \exists y{:}B\ F(x,y) \longrightarrow \exists f{:}B^A\ \forall x{:}A\ F(x,f(x))$$

which is a law of logic when $\exists$ is taken to mean disjoint union and so, by our argument above, is simply a law of logic.

Remark 5 *A striking illustration of the difference between the 'truth-functional' conception of logic and the PAT point of view concerns the matter of definitions. Suppose that we wish to introduce by nominal definition a propositional function R defined on some type A. From the truth-functional point of view propositions are determined by their truth values and so R is defined when we specify some equivalence*

$$\forall x{:}A\ [R(x) \longleftrightarrow F(x)] \tag{4.2}$$

where F is some given propositional function defined on A. But from the type theoretic point of view, this is entirely inadequate. For any object t of type A, this equivalence only tells us that $R(t)$ is true precisely when $F(t)$ is; but it does not identify the proposition $R(t)$—it doesn't tell us what the objects of type $R(t)$ are. Indeed, from this equation we can deduce

$$\forall x{:}A\ [R(x) \longleftrightarrow F(x) \wedge F(x)]$$

but $F(t)$ is a distinct proposition from $F(t) \wedge F(t)$. On the type theoretic conception, the appropriate form of a nominal definition is that $R(t)$ means *or is* definitionally equal to *$F(t)$, which, unlike (4.2), is not itself a mathematical proposition, but a* metamathematical *one.*

Remark 6 *Notice that on the PAT conception, logical truths are not 'empty', as they are sometimes said to be. The truth of*

$$A \longrightarrow [B \longrightarrow A]$$

or

$$[A \longrightarrow B] \wedge [A \longrightarrow C] \longrightarrow [A \longrightarrow B \wedge C]$$

for example, expresses something about the general notions of function and pair. Indeed, on the PAT conception, it is hard to see along traditional lines where logic leaves off and mathematics begins. Perhaps the point of demarcation is the introduction of infinite types such as $\mathbf{N}$.

Remark 7 *The identification of the relation* proposition/proof *with* type/object *reveals a fact about the latter. Proofs are things that we can in principle construct; so, when we speak of an object of type A, we are speaking of something that we can construct, in the sense that it is given by a term. (The question of whether or not two terms denote the same object, i.e. are definitionally equal, is decidable.) But it may appear to be a difficulty with this point of view that we can only construct a countable number of objects in this sense, whereas, once we introduce the infinite type $\mathbf{T}$, we can construct types such as $\mathbf{P}(\mathbf{N})$ which are uncountable. There are two remarks to be made about this. One is that the theoruy of types does not form a closed formal system. Once we introduce $\mathbf{N}$, we are lead to introduce further infinite types which lead to the construction of more and more objects of type $\mathbf{P}(\mathbf{N})$. Of course, though, it is true that, no matter what direction we might extend the theory of types, the collection of terms that we obtain will remain countable. But—and this is the second remark to be made about the countability of what we can construct—the uncountability of $\mathbf{P}(\mathbf{N})$ is not a statement external to mathematics, about what objects we can in the long run construct, an anthropological statement, so to speak. Rather, it is expressed by the proposition*

$$\forall f{:}\mathbf{N} \longrightarrow \mathbf{P}(\mathbf{N})\exists x{:}\mathbf{P}(\mathbf{N})\forall y{:}\mathbf{N}[x \neq f(y)]$$

which we can prove.

4.3 Formulas and Terms

So far we have been speaking of types and objects informally, relying on the intuitive notion of function and pair. But the point of logic is to give a foundation for these basic notions by specifying exactly what constructions of them we admit and what principles we admit in reasoning about them. The most natural way to present such a foundation is to shift from the material to the formal mode of speach and, instead of speaking of types and objects, to speak of *formulas* and *terms*.

The classes of formulas and terms must be defined simultaneously. Formulas without free variables will be called *sentences* and are intended to denote types. Each term will be assigned a class of formulas, called its *types*. A term of type A is intended to denote an object of type A. We shall write

$$t{:}A$$

to mean that t is of type A.

We shall introduce relations

$$s \equiv t \qquad A \equiv B$$

of *definitional equality* between terms and formulas. Definitionally equivalent terms of formulas are understood to have the same meaning. We will show that this relation is decidable. The first step in our definition is to specify that

$$t\!:\!A, A \equiv B \Longrightarrow t\!:\!B$$

It will turn out that all the types of a term are definitionally equal and that definitionally equal terms have the same types.

We want to allow for types other than the ones we can obtain by means of the type-forming operations we consider here, and so we assume given some number of symbols, including $\mathbf{2}$, which we call *type constants*. For each sentence A, we may have zero or more *predicate constants* of *sort* A. The predicate constants will include $\mathbf{T}$ which is of sort $\mathbf{2}$.

The *atomic formulas* are the type constants and the expressions Rt, where R is a predicate constant of sort A and $t\!:\!A$. So in particular, $\mathbf{0}, \mathbf{2}$ and $\mathbf{T}s$ are atomic formulas whenever $s\!:\!\mathbf{2}$.

With each formula A we introduce the *free variables*

$$v_n(A)$$

of *sign* A ($n < \omega$). We sometimes denote free variables of sign A by $v(A)$ or, when the type is given by the context, by u or v. But it must be understood that a variable of sign A contains A as a syntactical part. Thus, the free variables occuring in $v(A)$ are itself along with the free variables occuring in A. A free variable of sign A is a term of type A:

$$v(A)\!:\!A$$

It follows that $v(A) : B$ for any B definitionally equivalent to A. The free variables of type A are intended to range over the objects of type A.

Besides free variables, we also introduce an infinite supply of *bound variables*, which we denote by $x, x, z, \ldots x_0, \ldots$. These are syntactically atomic and are not terms.

The formulas are built up from atomic formulas by means of the following operation: Let A and $F(v(A))$ be formulas such that, for all free variables

$u(C)$ occuring in $F(v(A))$ $v(A)$ does not occur in C. Let x be a bound variable not occuring in A or $F(v(A))$. Then

$$\forall x\!:\!A\ F(x) \qquad \exists x\!:\!A\ F(x)$$

are formulas.

The restriction that x not occur in A or in $F(v(A))$ is just a matter of convenience. For example, consider $\forall x\!:\!B\ F(x, v(A))$. If we now violate the restriction and form $\exists x\!:\!A \forall x\!:\!B\ F(x,x)$ we have lost track of which occurrences of x refer to which quantifier. We could make conventions governing this; but we lose nothing by imposing our simple restriction.

The point of the restriction on the free variables should be clear. The meaning of a formual or term should be determined once we assign values to the free variables it contains. Call a free variable v *unfettered* in a formula B or term t iff, for all free variables $u(C)$ in B or t, respectively, v does not occur in C. Suppose that v is fettered in $F(v)$, e.g. $F(v)$ is $F(v, u(C(v)))$, containing the variable $u(C(v))$. Then in $\forall x\ F(x) = \forall x\ F(x, u(C(x)))$ the expression $u(C(x))$ is no longer a variable, since $C(x)$ is not a type. The assignment of values to the free variables which occur in $\forall x\ F(x, u(C(x)))$ therefore does not determine its meaning.

From the point of view of logic, formulas denote propositions and a term t of type A denotes a proof of A. Let the free variables in t which are not in A have signs $B, \ldots, C$. Then we say that t is a *deduction* of A from the *premises* $B, \ldots, C$. For example, the variable $v(A)$ is a deduction of A from the premise A (since $v(A)$ cannot occur in A.) Note that when a variable $v(B)$ occurs in both t and in A, we do not *on that account* take B to be a premise; for in this case, it is not the proof that depends upon the value of $v(B)$, but the proposition A itself. Of course, it could happen that there are other variables of sign B in t which do not occur in A, so that B is nevertheless a premise of the deduction.

Our definition of a deduction from premises has a somewhat odd consequence which may be seen from the example of predicate logic. If D is the domain of individuals, R is a predicate constant of sort D and $v = v(D)$ is a free individual variable, then there is a term $t = \lambda x\!:\!Rv\ x$ (to be introduced below) of type $Rv \longrightarrow Rv$. Since t contains no free variables other than v, which also occurs in $Rv \longrightarrow Rv$, it is a deduction of $Rv \longrightarrow Rv$ with no premises. On the other hand, (v, t) is of type $\exists x : D\ [Rx \longrightarrow Rx]$ and, according to our definition, is a deduction of this sentence from the premise

D. But this deduction is generally regarded, except by the advocates of 'free logic', to be an absolute deduction, i.e. without premises. So the theory of types sides with free logic on this issue. On the other hand, on the model theoretic conception of logic, we assume the domain of individuals to be non-empty; and this amounts to taking D as a premise. In the general case, the point to be made is that the inference from $F(s)$ to $\exists x\!:\!A\ F(x)$, when $s\!:\!A$, may introduce a premise B. Namely this will happen if a free variable $v(B)$ in s is also in the deduction t of $F(s)$ but no free variable of type B is in $\exists x\!:\!A\ F(x)$. On the other hand, the inference to $\exists x\!:\!A\ F(x)$ is not from $F(s)$ alone, as we are used to thinking in the case of predicate logic, but from A and $F(s)$. Since, by hypothesis, $v(B)$ does not occur in $\exists x : A\ F(x)$, it *a fortiori* does not occur in A. Hence, B is already a premise of s. So no new premises are really being introduced.

There is another oddity in the treatment of logic within the theory of types that is connected with the fact, already noted, that in logic we are generally not interested in the kinds or multiplicity of proofs that a proposition has, but only in whether or not it has a proof. So from the point of view of deductions, there is no need for more than one variable of each type. A deduction of a sentence A containing two distinct variables of sign B is simply a deduction of A from the premise B, and we might as well substitute one of the variables of sign B for all of the others in the deduction. On the other hand, when we think of the formulas as denoting types of objects and the terms as denoting objects, then there isa significant difference between the term $t(u, v)$ containing two variables of sign B which can vary independently of one another and the term $t(v, v)$. We could eliminate this difference between deductions in the usual sense and deductions in the theory of types by requiring the premises of deductions in the former sense to be *labelled*—to be indexed by numbers. Then the premise A with index n would correspond to the variable $v_n(A)$.

By choosing a variable $v(A)$ not occuring in B, we obtain the formulas

$$A \longrightarrow B := \forall x\!:\!A\ B \qquad A \wedge B := \exists x\!:\!A\ B$$

Of course this notation hids the particular choice of the bound variable x occuring in the formula; but this won't matter, since *we will not distinguish between terms or formulas which can be obtained from one another by changes ('renaming') of bound variables.* A term will be of type $\forall x\!:\!A\ F(x)$ just in case it is of type $\forall y\!:\!A\ F(y)$, providing that both are formulas; and similarly

for $\exists$. It follows that we will pass from the formula $F(v(A))$ to $\forall x : A\ F(x)$ or $\exists x : A\ F(x)$ without mentioning that x must not occur in $F(v(A))$, since, if it did, we could always first rename the bound variables to comply with the restriction.

Similarly, given the term $t(v(A))$ and formula $F(v(A))$, we may substitute a term s of type A for $v(A)$, yielding $t(s)$ and $F(s)$, respectively. If s contains bound variables, we simply may first rename them to be different from any bound variables in $t(v(A))$ or $F(v(A))$ before making the substitution.

Our definitions in the last two sections of $\mathbf{P}(A)$, $s \in_A t$, $A \longleftrightarrow B$, $\langle A, B\rangle$ and $A \vee B$ will carry over to the formal mode; but now they must be understood, not as the definition of relations and operations, but as abbreviations of formulas. For example, $\mathbf{P}(A)$ is just an abbreviation for $A \longrightarrow \mathbf{2}$, i.e. for $\forall x : A.\mathbf{2}$.

We turn now to the construction of terms other than free variables. The rules of construction of terms have the form of introduction and elimination rules in the sense of Natural Deduction.

The Type **2**

The **2**-*Introduction Rule* is

$$\top : \mathbf{2} \qquad \bot : \mathbf{2}$$

The **2**-*Elimination Rule* is

$$s : F(\top), t : F(\bot), r : \mathbf{2} \Rightarrow [s, t, r]_{F(x)} : F(r)$$

This rule expresses the fact that $\top$ and $\bot$ are the only objects of type **2**. So, if $F(\top)$ and $F(\bot)$ are true, then so is $F(r)$ for all $r : \mathbf{2}$. For the sake of brevity, we will generally drop the subscript $F(x)$. Clearly we should have $[s, t, s]$ mean the same as s and $[s, t, t]$ mean the same as t. We will express this by means of the *Conversion Rule*

$$[s, t, \top]\ CONV\ s \qquad [s, t, \bot]\ CONV\ t$$

We now introduce the abbreviations

$$\mathbf{0} := \mathbf{T}\bot \qquad \mathbf{1} := \mathbf{T}\top$$

Now we can introduce the abbreviation $\neg A$ of $A \longrightarrow \mathbf{0}$.

The Type $\mathbf{0}$

There are no introduction rules for $\mathbf{0}$; but there is the $\mathbf{0}$-*Elimination Rule*

$$t : \mathbf{0} \Rightarrow N(A, t) : A$$

for every formula A. So, for every formula A and every terms t of type $\mathbf{0}$, there is a term $N(A, t)$ of type A.

The Type $\mathbf{1}$

The $\mathbf{1}$-*Introduction Rule* is

$$\mathbf{o} : \mathbf{1}$$

The $\mathbf{1}$-*Elimination Rule* is

$$s : F(\mathbf{o}), t : \mathbf{1} \;\Rightarrow\; [s, t]_{F(x)} : F(t)$$

for each formula $F(v(\mathbf{1}))$. Again, we shall usually drop the subscript $F(x)$ when denoting this term. Clearly $[s, \mathbf{o}]$ should just be s. So we have the conversion rule

$$[s, \mathbf{o}]\ CONV\ s$$

Universal Quantification

The $\forall$-*Introduction Rule* is this: Let $V = v(A)$ be unfettered in $F(v)$ and in the term $t(v)$. Then

$$t(v) : F(v) \Rightarrow \lambda x : A\ t(x)\ :\ \forall x : A\ F(x)$$

We have already discussed the necessity for requiring v to be unfettered in $F(v)$ in order to form $\forall x : A\ F(x)$. The requirement that v be unfettered in $t(v)$ is precisely the familiar requirement that v not occur in any premise in the deduction $t(v)$ of $F(v)$. For example, without this restriction, we could construct the nonsense 'deduction' $\lambda x : A\ u(F(x))$ from the deduction $u(F(v))$ of $F(v)$.

In the case of implication $A \longrightarrow B$, $B = F(v)$ doesn't contain v. $t(v)$ is a deduction of B and, if v actually occurs in $t(v)$, then it is a deduction of B from the premise A (among others possibly). Passing to the deduction $\lambda x : A\ t(x)$ of $A \longrightarrow B$ expresses the usual rule of $\longrightarrow$-*Introduction* or *Deduction Theorem*: given a deduction of A from the assumption v of A, we may infer $A \longrightarrow B$, 'discharging' the assumption v. Discharging the assumption v is expressed for us by binding the variable v.

Let f be a term of type $\forall x : A\ F(x)$ and let $s : A$. Then f is intended to denote a function defined on A and s an object of type A. We form the term (fs) to denote the value of the function f for the argument s. Thus, we write (fs) instead of the more usual $f(s)$. When there will be no confusion, we drop the parentheses and write fs. The rule of $\forall$-*Elimination* then is

$$f : \forall x : A\ F(x), s : A\ \Rightarrow fs : F(s)$$

The type $F(s)$ might itself be of the form $\forall y : B\ G(y)$, so that fs denotes a function defined on B. For t of type B we then write fst for $(fs)t$. In general,

$$fst \cdots r \text{ abbreviates } (\cdots((fs)t)\cdots r)$$

In the special case of $A \longrightarrow B$, in which $F(v) = B$ does not contain v, $\forall$-Elimination takes the form of *Modus Ponens* or $\longrightarrow$-*Elimination*

$$f : A \longrightarrow B, s : A\ \Rightarrow fs : B$$

When $f = \lambda x{:}A\ t(x)\ {:}\forall x{:}A\ F(x)$ and $s{:}A$, then clearly fs is to denote the deduction of $F(s)$ that we obtain from $t(v)$ by substituting the object s of type A for the variable v of type A. Thus, the meaning of fs is given by the Conversion Rule

$$(\lambda x{:}A\ t(x))s\ CONV\ t(s)$$

This is called the rule of λ-*Conversion*.

Existential Quantification

The rule of $\exists$-*Introduction* is

$$s{:}A, t{:}F(s)\ \Rightarrow (s,t)_B{:}\exists x{:}A\ F(x)$$

where $B = \exists x : A\ F(x)$. The subscript is necessary to avoid ambiguity. For example, we might have a formula $G(v,v)$, where $v = v(A)$ and $F(v)$ could be any of $G(s,v), G(v,s)$ or $G(v,v)$. $s{:}F(s)$ would hold in each case if it held in any of them. So, without the subscript, the type of (s,t) is not determined. Nevertheless, when the context determines which formula is involved, we shall drop the subscript and write simply (s,t).

A special case of $\exists$-Introduction is $\wedge$-*Introduction*

$$s{:}A, t{:}B\ \Rightarrow\ (s,t){:}A \wedge B$$

Let $p{:}\exists x{:}A\ F(x)$. Then p stands for a pair whose first member is an object s of type A and whose second member t is of type $F(s)$. If we denote s by $(p1)$ and t by$(p2)$, then the rule of $\exists$-*Elimination* is simply

$$p{:}\exists x{:}A\ F(x)\ \Rightarrow (p1){:}A, (p2){:}F((p1))$$

From the explanation just given, we have the Conversion Rules

$$((s,t)1)\ CONV\ s \qquad ((s,t)2)\ CONV\ t$$

When it will cause no confusion, we will write $p1$ and $p2$ for $(p1), (p2)$, respectively. Also, we will extend the abbreviation

$$fst\cdots r \text{ abbreviates } (\cdots((fs)t)\cdots r)$$

to the case in w hich some of the $s, t, \ldots, r$ are not terms, but 1's and 2's.

In the special case of $A \wedge B$, we obtain the usual rule of $\wedge$-Elimination:

$$p{:}A \wedge B\ \Rightarrow\ p1{:}A, p2{:}B$$

But in the general case, our version of $\exists$-Elimination is not the usual one. For example, suppose that A is the domain of individuals and that p is a deduction of the first-order formula $\exists x\!:\!A\ F(x)$. Then $p1\!:\!A$. In other words, from the deduction p we obtain an individual term $p1$. Moreover, $p2$ is a deduction of $F(p1)$. But this will not in general be a first-order formula. The same observation applies to higher order predicate logic as well. The theory of types does not fit in the framework of predicate logic. On the other hand, as we shall see, Gentzen provided another form of $\exists$-Elimination which does not lead outside the framework of predicate logic and such that any first-order formula deducible in the theory of types will be deducible using Gentzen's form of $\exists$-Elimination.

There is one more rule of construction that we need, which occurs only in classical logic and not in intuitionistic logic; namely the rule of $\neg\neg$-*Elimination*

$$t\!:\!\neg\neg A\ \Rightarrow D(t)\!:\!A$$

Terms t in which all occurrence of D are in parts of t of the form $v(B)$, will be called *constructive* or *intuitionistic* terms or deductions.

For handy reference, here is a list of the rules for constructing terms:

$$\begin{gathered}
v(A)\!:\!A \\
t\!:\!\mathbf{0} \Rightarrow N(A,t)\!:\!A \\
s\!:\!F(\mathbf{o}), t\!:\!\mathbf{1}\ \Rightarrow\ [s,t]\!:\!F(t) \\
\top\!:\!\mathbf{2} \qquad \bot\!:\!\mathbf{2} \\
r\!:\!\mathbf{2}, s\!:\!F(\top), t\!:\!F(\bot) \Rightarrow [s,t,r]\!:\!F(r) \\
t(v(A))\!:\!F(v(A)) \Rightarrow \lambda x\!:\!At(x)\ :\ \forall x\!:\!AF(x) \\
f\ :\ \forall x\!:\!AF(x), s\!:\!A\ \Rightarrow\ fs\!:\!F(s) \\
s\!:\!A, t\!:\!F(s)\ \Rightarrow\ (s,t)\!:\ \exists x\!:\!AF(x) \\
p\ :\ \exists x\!:\!AF(x)\ \Rightarrow\ p1\!:\!A, p2\!:\!F(p1) \\
t\!:\!\neg\neg A \Rightarrow D(t)\!:\!A \\
t\!:\!A, A \equiv B \Rightarrow t\!:\!B
\end{gathered}$$

where, in the case of $\lambda x\!:\!At(x)$, it is required that $v(A)$ be unfettered in both $t(v(A))$ and $F(v(A))$.

We have been glossing over a point that should now be made explicit: Let $v = v(A), s\!:\!A$ and let $t(v)\!:\!F(v)$. We have been assuming that i) $F(s)$, the result of substituting s for v in $F(v)$ is a formula and ii) $t(s)$, the result of substituting s for v in $t(v)$, is a term of type $F(s)$. The proof of this, by

induction on $F(v)$ and $t(v)$, is routine.

4.4 Definitional Equality

Here is a list of the conversion rules:

$$[s, \mathbf{o}]\ CONV\ s \tag{4.3}$$

$$[s, t, \top]\ CONV\ s \qquad [s, t, \bot]\ CONV\ t \tag{4.4}$$

$$\lambda x\!:\!At(x)s\ CONV\ t(s) \tag{4.5}$$

$$(s, t)1\ CONV\ s \qquad (s, t)2\ CONV\ t \tag{4.6}$$

The right- and left-hand terms in any instance of (4.4) have the same type $F(\top)$ or $F(\bot)$. This is true in the case of (4.5, too. If $x\!:\!At(x)\ :\ \forall x\!:\!AF(x)$ and $s : A$, then both $\lambda x : At(x)s$ and $t(s)$ are of type $F(s)$. In the case of (4.6), $(s, t)1$ and s have the same type A, where $(s, t)\ :\ \exists x : AF(x)$. But $(s, t)2$ has the type $F((s, t)1)$ and t has the type $F(s)$. So if they are to have the same types, we must ensure that $F((s, t)1) \equiv F(s)$. However, that will follow immediately from the following

Definition 10 (Definitional Equality) *The relation $\equiv$ is the least equivalence relation between terms or formulas such that*

$$s\ CONV\ t\ \Rightarrow\ s \equiv t$$

$$s \equiv t, A \equiv B\ \Rightarrow\ N(A, s) \equiv N(B, t)$$

$$s \equiv s', t \equiv t', F(v) \equiv G(v)\ \Rightarrow\ [s, t]_{F(x)} \equiv [s', t']_{G(x)}$$

$$s \equiv s', t \equiv t', r \equiv r', F(v) \equiv G(v)\ \Rightarrow\ [s, t, r]_{F(x)} \equiv [s', t', r']_{G(x)}$$

$$f \equiv g, s \equiv t\ \Rightarrow\ fs \equiv gt$$

$$s(v) \equiv t(v), A \equiv B\ \Rightarrow\ \lambda x\!:\!A\ s(x) \equiv \lambda y\!:\!B\ t(y)$$

$$s \equiv s', t \equiv t', F(v) \equiv G(v)\ \Rightarrow\ (s, t)_{F(x)} \equiv (s', t')_{G(x)}$$

$$s \equiv t\ \Rightarrow\ s1 \equiv t1, s2 \equiv t2$$

$$s \equiv t \ \Rightarrow D(s) \equiv D(t)$$

$$s \equiv t \ \Rightarrow \ Rs \equiv Rt$$

and, if v is a variable of type A

$$A \equiv B, F(v) \equiv G(v) \ \Rightarrow \forall x\!:\!AF(x) \equiv \forall y\!:\!BG(y)$$

and

$$A \equiv B, F(v) \equiv G(v) \ \Rightarrow \exists x\!:\!AF(x) \equiv \exists y\!:\!BG(y)$$

Note that $v_n(A) \equiv v_n(B)$ holds only when A and B are identical.

It easily follows from the definition that

Proposition 15

- *If v is unfettered in $s(v)$ and $t(v)$, $s(v) \equiv t(v)$, and $p \equiv q$, then $s(p) \equiv t(q)$.*
- *If v is unfettered in $F(v)$ and $G(v)$, $F(v) \equiv G(v)$, and $p \equiv q$, then $F(p) \equiv G(q)$.*
- *Let $s\!:\!A$ and $t\!:\!B$. Then $s \equiv t$ implies $A \equiv B$.*

It follows that each equivalence class C of terms is associated with an equivalence class C' of formulas such that the types of each $t \in C$ are precisely the formulas in C'.

Our aim now is to obtain an algorithm for deciding whether or not two terms or formulas are definitionally equal.

Let $v = v(A)$. We call a term *v-less* if it does not contain v. By a *maximal v-less part* of a term t we mean an occurrence of a v-less term in t which is not a proper part of an occurrence of another v-less term in t. We may write $t = T(t_1, \ldots, t_n)$, where the $t_i : A_i$ are maximal v-parts of t and $T(v(A_1), \ldots, v(A_n))$ contains no v-less parts and each variable $v(A_i)$ has just one occurrence in it. (So the t_i are not necessarily distinct.) We call $T(t_1, \ldots, t_n)$ the *v-structure* of t. If v does not occur in t, then $T(v(A_1), \ldots, v(A_n))$ is just a variable, since t is the only maximal v-less part of t.

Definition 11 *The relations s RED t and s $n-RED$ t between terms is defined for $n > 0$ by*

- *s $1-RED$ s.*

- *If s CONV t, then $s\ 1-RED\ t$.*
- *If $s\ 1-RED\ s', t\ 1-RED\ t'$, and $r\ 1-RED\ r'$, then $N(A,s)\ 1-RED\ N(A,s'), [s,t]\ 1-RED\ [s',t'], [s,t,r]\ 1-RED\ [s',t',r']$ and $(s,t)\ 1-RED\ (s',t')$.*
- *Let $T(s_1,\ldots,s_n)$ be the v-structure of $s(v)$, let $s_i\ 1-RED\ t_i$ for $0<i\leq n$ and let $t(v)=T(t_1,\ldots,t_n)$. Then $\lambda xs(x)\ 1-RED\ \lambda yt(y)$.*
- *If $f\ 1-RED\ g$ and $s\ 1-RED\ t$, then $fs\ 1-RED\ gt$.*
- *If $r1-REDs$ and $sn-REDt$, then $rn+1-REDt$.*
- *If $s\ n-RED\ t$ for some n, then $s\ RED\ t$.*

It is immediate that $s\ REDt$ implies that $s\equiv t$.

By a *simple term* we shall mean one which is not of the form st, where s is a term and t is either a term or is 1 or 2. So, using our convention for dropping parentheses, every term is uniquely of the form

$$fst\cdots r$$

where f is a simple term and $s,t,\ldots,r$ are all either 1's, 2's or terms.

In what follows, we just write $N(t)$ for $N(A,t)$ and $\lambda xt(x)$ for $\lambda x{:}At(x)$.

Lemma 11 *If s' is obtained from s by simultaneously replacing terms t_i by terms t_i', where $t_i\ 1-RED\ t_i'$, then $s\ 1-RED\ s'$.*

The proof is by induction on s.

Case 1. $s=s_0s_1\cdots s_n$ and $s'=s_0's_1'\cdots s_n'$, where $n>0$ and each substitution of a t_i is in one of the s_j. Then the result is immediate from the induction hypothesis. So we can assume then that s is simple.

Case 2. $s=\lambda xp(x)$ and $s'=\lambda xp'(x)$. Let $T(p_1,\ldots,p_k)$ be the v-structure of $p(v)$. Then the substitutions of the t_i in s must be substitutions in the p_j. So $p'=T(p_1',\ldots,p_k')$, where by the induction hypothesis, $p_j\ 1-RED\ p_j'$. Hence, by definition, $s\ 1-REDs'$.

Case 3. s is $N(p),[p,q],[p,q,r],(p,q),p1,p2$ or $D(p)$. The result follows immediately from the induction hypothesis.

Lemma 12 *Let r CONV s and $r\ 1-RED\ t$. Then there is a term u such that $s\ 1-RE\ u$ and $t\ 1-RED\ u$.*

The proof is by induction on r.

If $s = t$, we have nothing to prove, so we assume that s and t are distinct. It follows that t is not obtained from r by conversion.

Case 1. $r = [s, \mathbf{o}]$ and $t = [s', \mathbf{o}]$. Then $u = s'$.

Case 2. $r = [s, p, \top]$ and $t = [s', p'\top]$. Then $u = s'$.

Case 3. $r = [p, s, \bot]$ and $t = [p', s'\bot]$. Then $u = s'$.

Case 4. $r = \lambda x p(x)q$, $s = p(q)$ and $t = \lambda x p'(x)q'$, where $\lambda x p(x)$ $1 - RED$ $\lambda x p'(x)$ and q $1 - RED$ q'. Then t $1 - RED$ $p(q')$. We need to show that s $1 - RED$ $p'(q')$ Let $T(p_1, \ldots, p_k)$ be the v-structure of $p(v)$. Then $p'(v) = T(p'_1, \ldots, p'_k)$, where p_j $1 - RED$ p'_j. The p_j are all v-less and so $p'(q')$ is obtained from $p(q)$ by simultaneously replacing zero or more terms by terms to which they 1-reduce. So by Lemma 11, s $1 - RED$ $p'(q')$.

Case 5. $r = (s, p)1$ and $t = (s', p')1$, Then $u = s'$ suffices.

*Case 6.*Similarly for $r = (p, s)2$ and $t = (p', s')2$.

Lemma 13 *Let r $1 - RED$ s and r $1 - RED$ t. Then there is a term u such that s $1 - RED$ u and t $1 - RED$ u.*

$$\begin{array}{ccc} r & \xrightarrow{1-RED} & s \\ {\scriptstyle 1-RED}\downarrow & & \downarrow{\scriptstyle 1-RED} \\ t & \xrightarrow[1-RED]{} & u \end{array}$$

Proof by induction on r. Let $r = r_0 \cdots r_n$, where r_0 is simple. The 1-reduction of r to s is called *internal* iff $s = s_0 \cdots s_n$, where s_0 is simple and r_i $1 - RED$ s_i for each i. If the 1-reduction is not internal, we call it *external*.

Case 1. The 1-reductions of r to s and t are both internal, then $s = s_0 \cdots s_n$ and $t = t_0 \cdots t_n$, where r_i 1-reduces to both s_i and t_i for each i. Assume $n > 0$. Then the induction hypothesis applies to yield, for each i, a u_i such that both s_i and t_i 1-reduce to u_i. Then s and t 1-reduce to $u = u_0 \cdots u_n$.

If $n = 0$, then $r = r_0$ is a simple term $\mathbf{o}, \top.\bot, N(r'), [r', r''], [r', r'', r'''], (r', r'')$ or $\lambda x r'(x)$ and the result follows easily by the induction hypothesis.

Case 2 Both 1-reductions are external. Then $r_0 \cdots r_k$ converts to some r', $s = r's_{k+1} \cdots s_n$ and $t = r't_{k+1} \cdots t_n$. By the induction hypothesis, the is a u_i for each $i > k$ such that s_i and t_i 1-reduce to u_i. So s and t 1-reduce to $r'u_{k+1} \cdots u_n$.

Case 3. One of the 1-reductions, say the 1-reduction of r to s, is external and the other is internal. So $r_0 \cdots r_k CONV r'$, $s = r' s_{k+1} \cdots s_n$, where r_i1-reduces to s_i for $i > k$, and $t = t_0 \cdots t_n$, where r_i 1-reduces to t_i for each i. By Lemma 12, r' and $t_0 \cdots t_k$ 1-reduce to some u'. By the induction hypothesis, s_i and t_i 1-reduce to some u_i for $i > k$. So s and t 1-reduce to $u' u_{k+1} \cdots u_n$.

Lemma 14 *If r $m-RED$ s and r $n-RED$ t then there is a u such that s $n-RED$ u and t $m-RED$ u.*

Proof:

$$\begin{array}{ccccccc}
r = r_{11} & \longrightarrow & r_{12} & \longrightarrow & \cdots & \longrightarrow & r_{1m} = s \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
r_{21} & \longrightarrow & r_{22} & \longrightarrow & \cdots & \longrightarrow & r_{2m} \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
\vdots & \longrightarrow & \vdots & \longrightarrow & \vdots & \longrightarrow & \vdots \\
\downarrow & & \downarrow & & \downarrow & & \downarrow \\
t = r_{n1} & \longrightarrow & r_{n2} & \longrightarrow & \ldots & \longrightarrow & r_{nm} = u
\end{array}$$

Corollary 9 (Church-Rosser Theorem) *If r RED s and r RED t then there is a u such that s RED u and t RED u.*

Definition 12 *The relation of* subterm *is the least reflexive relation between terms such that*

- *s is a subterm of $[s,t]$, $[t,s]$, $[s,r,t]$, $[r,s,t]$, $[r,t,s]$, (s,t), (t,s), $s1$ and $s2$.*
- *s is a subterm of $\lambda x\!:\!At(x)$ if it is a subterm of $t(v)$, where $v = v_n(A)$ for the least n such that $v_n(A)$ does not occur in $t(x)$.*

A term is called *normal* and is said to be in *normal form* iff it contains no convertible subterms. If s RED t and t is normal, then t is called a *normal form of* s. We can change the 'a' to 'the':

Corollary 10 (Uniqueness of normal form) *Every term has at most one normal form.*

We will prove later on that every term has a normal form.

Remark 8 *Lemma 14 and the Church-Rosser Theorem do not depend on type structure. Indeed, both were first proved for the abstract Lambda-Calculus and the Theory of Combinators. The existence of normal form, however, depends very much on type structure. For example, let* $t = \lambda x(xx)$*, a term which could not carry any type structure.* tt *has no normal form, since it converts and only converts to itself.*

Definition 13 *The class of* terms in *the formula* A *is defined by induction on* A*.*

- *There are no terms in a type constant.*
- *The terms is* Rt *are the subterms of* t*.*
- *The terms in* $\forall x : AF(x)$ *or* $\exists x : AF(x)$ *are the terms in* A *or* $F(v)$ *(where* $v = v_n(A)$ *for the least* n *such that* $v_n(A)$ *does not occur in* $F(x)$*).*

A formula is said to be *normal* or to be in *normal form* iff it has no convertible terms.

Definition 14 *The relation* $A\ RED\ B$ *between formulas is defined by*

- $A\ RED\ A$
- $s\ RED\ t\ \Rightarrow\ Rs\ RED\ Rt$
- $A\ RED\ B, F(v(A))\ RED\ G(v(A)\ \Rightarrow\ \forall x{:}AF(x)\ RED\ \forall x{:}BG(x), \exists x{:}AF(x)\ RED\ \exists x{:}BG(x)$

It is immediate that $A\ RED\ B$ implies $A \equiv B$. B is called a *normal form of* A iff B is normal and $A\ RED\ B$. It follows by induction on A from the uniqueness of normal form for terms that

Corollary 11 *Every formula has at most one normal form.*

We now want to prove that every term has a normal form. The relation

$$s > t$$

is defined exactly as is $s\ 1-RED\ t$, except that the first clause (viz. $s\ 1-RED\ s$) is dropped. In otherwords, $s > t$ implies that some conversion takes place in obtaining t from s. This of course does not mean *prima facie* that $s > s$ cannot ever hold. We just considered an example where it does hold in the abstract Lambda-Calculus. But we will prove that it does not happen in type theory. Call a term t *well-founded* iff there are no infinite sequences

$$t = t_0 > t_1 > t_2 > \cdots$$

We will prove that every term is well-founded.

Definition 15 *We define the notion of a* computable term *of type A by induction on A.*

- *A term of atomic type is computable iff it is well-founded.*
- *A term f of type $\forall x : AF(x)$ is computable iff ft is a computable term of type $F(t)$ for every computable term t of type A.*
- *A term p of type $\exists x : AF(x)$ is computable iff $p1$ is a computable term of type A and $p2$ is a computable term of type $F(p1)$.*

We define the notion of a *c-extension* of a term t of type A by induction on A. If A is atomic, then t is the only c-extension of t. If $A = \forall x : BF(x)$, then the c-extensions of t are precisely the c-extensions of ts for all computable terms $s : B$. If $A = \exists x : BF(x)$, then the c-extensions of t are precisely the c-extensions of $t1$ and $t2$. So the c-extensions of t are all the terms of atomic type of the form $ts_1 \cdots s_n$ where the s_i are either 1, 2 or are computable. t is computable iff all of its c-extensions are computable and hence well-founded.

The *rank*$|A|$ of a formula A is defined as follows: For atomic formulas A, set $|A| = 0$ and set $|\forall x : BF(x)| = |\exists x : BF(x)| = Max\{|B|, |F(v)|\} + 1$.

Lemma 15 *For each type A*

a) Every variable v of type A is computable.

b) Every computable term of type A is well-founded.

c) If s is a computable term of type A and $sREDt$, then t is computable.

The proof is by induction on $|A|$.

a) We need to show that all c-extensions $s = vs_1 \cdots s_n$ of v are well-founded. The s_i which are terms are of types of rank $< |A|$ and so, by the induction hypothesis, are well-founded. It immediately follows that s is well-founded.

b) Let s be computable and consider the c-extension $t = ss_1 \cdots s_n$ of s. t is computable and hence well-founded. So s must be well-founded.

c) We need to show that all c-extensions $tt_1 \cdots t_n$ of t are well-founded. But if such a term were not well-founded, then neither would the c-extension $st_1 \cdots t_n$ of s be, contradicting the computability of s.

It follows from b) that in order to prove that every term is well-founded, it suffices to prove that every term is computable. By an *c-instance* of a term t we will mean any result of replacing the variables in t by computable terms of the same type. (So by a) above, t is a c-instance of itself.)

Theorem 7 (Computability Theorem) *Every c-instance of a term t is computable.*

The proof is by induction on t. If t is a variable, this is immediate. If it is a constant, then it is of atomic type and, being well-founded, is computable. If $t = rs$, then by the induction hypothesis, the c-instances of r and s are computable. So any c-instance of t is computable by definition. Similarly, if $t = s1$ or $t = s2$, where s is computable. So we need only consider terms t of the form $[r, s], [p, r, s], (r, s)$ or $\lambda x\!:\!As(x)$.

Let t' be a c-instance of t. Consider any c-extension $p = t't_1 \cdots t_n$ of t'. We need to show that p is well-founded. Consider a sequence

$$p = p_0 > p_1 > \cdots \tag{4.7}$$

We have to show that (4.7) is finite. In order to simplify notation, we will agree to write $1 > 1$ and $2 > 2$.

Let $t = N(A, s)$ and $t' = N(A', s')$. Then every term in the sequence is of the form $N(A', q)q_1 \cdots q_n$, where s' *RED* q and t_i *RED* q_i for $0 < i \leq n$. By the induction hypothesis, s' is computable and the terms among the t_i are computable; Hence the sequence is finite.

Let $t = [r, s]$ and $t' = [r', s']$. By the induction hypothesis, r' and s' are computable. Suppose that every term in the sequence is of the form $[q, q']q_1 \cdots q_n$, where r' *RED* q, s' *RED* q' and t_i *RED* q_i for $0 < i \leq n$.

Then the sequence is finite because r', s' and all the terms in the list $t_1 \cdots t_n$ are computable. The other alternative is that some term in the sequence is of the form $[q, \mathbf{o}]q_1 \cdots q_n$, where $r' > q, s'$ RED $\mathbf{o}$ and t_i RED q_i for $0 < i \leq n$, and the next term is $qq'_1 \cdots q'_n$, where q_i RED q'_i for $0 < i \leq n$. . Then the sequence is finite because $q, q'_1, \ldots q'_n$ are.

The argument in the case of $t = [r, s, t]$ is exactly the same.

Let $t = (r, s)$ and $t' = (r', s')$. Suppose that every term in the sequence is of the form $(q, q')q_1 \cdots q_n$, where $r' > q, s' > q'$ and $t_i > q_i$ for $0 < i \leq n$. Then the sequence is finite because r', s' and all the terms in the list $t_1 \cdots t_n$ are computable. The other alternative is that $n > 0$, $t_1 = q_1$ is either 1 or 2 and some term in the sequence is of the form $(q, q')q_1 \cdots q_n$, where $r' > q, s' > q'$ and $t_i > q_i$ for $0 < i \leq n$, and the next term is $qq'_2 \cdots q'_n$ or $q'q'_2 \cdots q'_n$ (depending on whether t_1 is 1 or 2), where $q_i > q'_i$ for $1 < i \leq n$. Then the sequence is finite because $q, q', q'_1, \ldots q'_n$ are.

Finally, let $t = \lambda x s(x)$ (dropping the type for the sake of brevity). Then $t' = \lambda x s'(x)$. Suppose that every term in the sequence is of the form $\lambda x q(x) q_1 \cdots q_n$, where $s'(v) > q(v)$ and $t_i > q_i$ for $0 < i \leq n$. Then the sequence is finite because s' and all the terms in the list $t_1 \cdots t_n$ are computable. Otherwise, $n > 0$ and there will be a term $\lambda x q(x) q_1 \cdots q_n$ in the sequence, where $s'(v) > q(v)$ and $t_i > q_i$ for $0 < i \leq n$, and where the next term in the sequence is $s'(q_1)q'_2 \cdots q'_n$, where $q_j > q'_j$ for $1 < j \leq n$. In that case the sequence is finite by the induction hypothesis because $s(q_1)$ is a c-instance of $s(v)$.

Corollary 12 (Well-foundedness Theorem) *Every term is well-founded and, therefore, has a (unique) normal form.*

Moreover, we can effectively compute the normal form of any term or formula. We now want to use these facts to obtain the promised algorith for deciding whether or not two terms of formulas are definitionally equal.

Definition 16 *We define the* strong normal form *of a term or formula.*

$$SNF(t) \qquad SNF(A)$$

will denote, respectively, the strong normal form of t and the strong normal form of A. For t and A non-normal, we define $SNF(t)$ and $SNF(A)$ to be the strong normal forms of their normal forms, respectively. So in the following, we are assuming the terms and formulas to be in normal form

- *If t is a variable or constant, $SNF(t) = t$.*
- $SNF(N(A,t)) = N(SNF(A), SNF(t))$.
- $SNF([s,t]_{F(x)} = [SNF(s), SNF(t)]_{SNF(F(x)}$, *where $SNF(F(x))$ denotes the result of replacing v by x in $SNF(F(v))$.*
- $SNF([r,s,t]_{F(x)} = [SNF(r), SNF(s), SNF(t)]_{SNF(F(x)}$.
- $SNF((s,t)_{F(x)} = (SNF(s), SNF(t))_{SNF(F(x)}$.
- $SNF(\lambda x{:}At(x)) = \lambda x{:}SNF(A)SNF(t(x))$.
- $SNF(p1) = SNF(p)1$ *and* $SNF(p2) = SNF(p)2$.
- $SNF(st) = SNF(s)SNF(t)$
- *If A is a type constant, $SNF(A) = A$.*
- $SNF(Rt) = RSNF(t)$.
- *If Q is a quantifier, $SNF(Qx{:}AF(x)) = Qx{:}SNF(A)SNF(F(x))$.*

It is easy to prove that $SNF(t)$ and $SNF(A)$ are normal. When $SNF(t) = t$ and $SNNF(A) = A$, we say that t and A are in *strong normal form.*

Now, let $s \sim t$ mean $SNF(s) = SNF(t)$ and let $A \sim B$ mean $SNF(A) = SNF(B)$. It is easy to show that $s \sim t$ implies $s \equiv t$ and that $A \sim B$ implies $A \equiv B$. Conversly, $\sim$ satisfies the conditions

$$s\ CONV\ t\ \Rightarrow\ s \sim t$$
$$s \sim t, A \sim B\ \Rightarrow\ N(A,s) \sim N(B,t)$$
$$s \sim s', t \sim t', F(v) \sim G(v)\ \Rightarrow\ [s,t]_{F(x)} \sim [s',t']_{G(x)}$$
$$s \sim s', t \sim t', r \sim r', F(v) \sim G(v)\ \Rightarrow\ [s,t,r]_{F(x)} \sim [s',t',r']_{G(x)}$$
$$f \sim g, s \sim t\ \Rightarrow\ fs \sim gt$$
$$s(v) \sim t(v), A \sim B\ \Rightarrow\ \lambda x{:}As(x) \sim \lambda y{:}Bt(y)$$
$$s \sim s', t \sim t', F(v) \sim G(v)\ \Rightarrow\ (s,t)_{F(x)} \sim (s',t')_{G(x)}$$
$$s \sim t\ \Rightarrow\ s1 \sim t1, s2 \sim t2$$
$$s \sim t\ \Rightarrow D(s) \sim D(t)$$

$$s \sim t \ \Rightarrow \ Rs \sim Rt$$

and, if v is a variable of type A

$$A \sim B, F(v) \sim G(v) \ \Rightarrow \forall x{:}AF(x) \sim \forall y{:}BG(y)$$

and

$$A \sim B, F(v) \sim G(v) \ \Rightarrow \exists x{:}AF(x) \sim \exists y{:}BG(y)$$

So, since $\equiv$ is defined to be the least equivalence relation satisfying these conditions, $\equiv$ and $\sim$ are the same relation. It immediately follows that

Proposition 16 *The relation of definitional equality between terms or formulas is decidable.*

Since the rules of conversion $s \ CONV \ t$ do not introduce variables in t which are not in s, every variable in the strong normal form of t or A already occured in t or A, respectively. It easily follows by induction on t that

Proposition 17 *If $t{:}A$ and A is in strong normal form, then every variable in A is in t.*

If t is a constant, then A contains no variable. If $t = v(B)$, then $SNF(B) = A$ and so every variable in A is in B and so in t.

If $t = [r, s]_{F(x)}$, then $A = SNF(F(s))$ and the result is immediate from the induction hypotheis. The case of the other simple terms is exactly the same. Let $t = rs$, where $r : \forall x{:}BF(x)$ and $s{:}B$. Then $A = SNF(F(s)$ and again the result follows immediately by the induction hypothesis.

It follows, in particular, that every closed term has a closed type.

Remark 9 *Definitional equality is a* metamathematical *relation between terms (or formulas). Two definitionally equal terms are intended to denote the same object; so that, put in the material mode of speach, definitional equality becomes the identity relation between objects. Better, when we speak of objects or types, we are speaking of closed terms or formulas (i.e. without free variables) modulo the relation of definitional equality. Of course, equivalently we could be speaking about closed terms and formulas in strong normal form. Definitional equality differs from the mathematical relation of extensional equality which we will later define. Distinct objects may be extensionally equal. Definitional equality is a decidable relation between terms,*

extensional equality is not. Frege's problem, discussed in his "Sense and reference," arises from his (characteristic) confusion of identity with extensional equality. Identity is not a relation between terms (as he supposed in his Begriffsschrift); rather it is obtained by abstraction from a relation between terms, namely that of definitional equality.) But the question of the identity of objects presented by two terms is trivially answered, or at least is decidable, since it is just the question of whether the two terms are definitionally equal. But the question of whether the objects presented by the two terms are mathematically *equal, i.e. extensionally equal, may be a nontrivial mathematical question.*

Nevertheless, a small question remains about the identity relation. Definitional equality between two terms is determined by the conversion rules. But what is special about these particular rules? We could, for example introduce the further conversion rules that reflect on the intended meaning of the operations. For example, that $\top$ and $\bot$ are the only objects of type $\mathbf{2}$ may be expressed by the fact that a function f of type $\forall x{:}\mathbf{2}F(x)$ is determined by its values $f\top$ and $f\bot$. We may express this by

$$[fT, fT, r]\ CONV\ fr$$

Again, let f be of type $\forall x : AF(x)$. Then $\lambda x : Afx$ is of the same type; and clearly we should have them extensionally equal. Why not the conversion rule

$$\lambda x{:}Afx\ CONV f$$

Again, when p is of type $\exists x : AF(x)$, then the pair $(p1, p2)$ has this type also, and clearly should be extensionally equal to p. So why not

$$(p1, p2)\ CONV\ p$$

The addition of these conversions would preserve the validity of the Church/Rosser Theorem and the Well-foundedness Theorem. Is there something special about the conversion relations thast we have introduced or is the relation of identity really a somewhat arbitrary notion? Of course, the question is irrelevant to mathematical practice, since it is the relation of extensional equality rather than identity that plays a role there.

It should be noted that the option does not exist to take identity in the widest possible sense and treat extensionally equal terms as denoting the same

object. Otherwise, if t is a term of type $\mathbf{T}s$ and r is extensionally equal to s, then we would have to regard t as denoting an object of type $\mathbf{T}r$. But then the question of whether or not the object denoted by t is of type $\mathbf{T}r$ would depend upon a proof that s and r are definitionally equal. But as we have noted, the 'proposition' that t is of type A is to be singular and outside the domain of mathematical propositions in the proper sense of requiring proof.

4.5 Some Theorems of Logic

Recall that a term of type A whose free variables other than those in A are of signs $B, \ldots, C$ is called a deduction of A from the premises $B, \ldots, C$. (Of course, some of these premises may be definitionally equal.) From now on, we shall want to speak of deductions of A simpliciter, i.e. one which has no premises, i.e. contains no variables not in A. We call such deductions *absolute*.

From now on, we shall assume that the signs of all variables are in strong normal form. This is of course no real restriction: Simple replace distinct variables $V(A_i)$ in terms or formulas by distinct variables whose signs are the strong normal forms of A_i.

Let t be a deduction of A, where A is in strong normal form. If t contains any variables, then it contains an unfettered one; choose one v of sign C, where, by our recent agreement, C is in strong normal form. Since every variable occuring in A occurs in t, v is unfettered in A. (Of course, it might not occur in A.) Write $t = t(v)$ and $A = F(v)$, displaying the occurrences of v. Then $t' = \lambda y : Bt(y)$ is a deduction of $A' = \forall y : CF(y)$ and is absolute if t is. Note that A' is in strong normal form. If t' contains a variable, then it contains an unfettered one, u of some sign B in strong normal form. Writing $t' = \lambda y : C(u)t(u, y)$ and $A' = \forall y : C(u)F(u, y)$ to display occurrences of u, we have $t'' = \lambda x : B\lambda y : C(x)t(x, y)$ of type $\forall x : B\forall y : C(x)F(x, y)$, and t'' is absolute if t is.

A list of variables $v_1, \ldots, v_n$ of signs $A_1, \ldots, A_n$, respectively, is said to be in *good order* iff $0 < i < j \leq n$ implies that v_j does not occur in A_i. Suppose that this list included all the free variables in the term $t = t(v_1, \ldots, v_n$ of type $A = F(v_1, \ldots, v_n)$, where A is in strong normal form and we are displaying all the occurrences of the variables in t and so in A. Remembering that the list of v_i is in good order, we may write $A_i = A_i(v_1, \ldots, v_{i-1})$ displaying all

the occurrences of variables in A_i. Then $t' =$

$$\lambda x_1 : A_1 \lambda x_2 : A_2(x_1) \cdots \lambda x_n : A_n(x_1, x_2, \ldots, x_{n-1}) t(x_1, x_2, \ldots, x_n)$$

is a deduction of the *universal closure*

$$\forall x_1 : A_1 \forall x_2 : A_2(x_1) \cdots \forall x_n : A_n(x_1, x_2, \ldots, x_{n-1}) F(x_1, x_2, \ldots, x_n)$$

of A. t' is absolute if t is.

The rule of $\forall$-Introduction is exactly the generalization of the rules of $\forall$-Introduction and $\longrightarrow$-Introduction in Natural Deduction for first-order predicate logic to the theory of types. The rule of $\forall$-Elimination differs from the Natural Deduction formulation in that in the latter case, there is just the one premise $f : \forall x F(x)$ and the conclusion $ft : F(t)$ for a given individual term t. Of course the type of x needn't appear, since there is only one type of variables in first-order logic: the type D of individuals. Also, the individual terms ('deductions' of D) are given once and for all, so that our second premise $t : D$ is not necessary. Of course, in the special case of $\longrightarrow$-Elimination, our form of the rules is exactly the Natural Deduction form, namely the rule of Modus Ponens.

$\exists$-Introduction is again exactly the same as in type theory as in Natural Deduction, both for the quantifier $\exists$ and for $\wedge$. On the other hand, our rule of $\exists$-Elimination, namely the operations $p1$ and $p2$, although the same as in Natural Deduction for the case of $\wedge$, is radically different in the case of $\exists$. Indeed, if $p : \exists x : AF(x)$, where v actually occurs in $F(v)$, then even when $\exists x : AF(x)$ is a first-order formula (with A the type of individuals), the type $F(p1)$ of $p2$ is not a first-order formula. Existential quantifiers elimination in Natural Deduction is essentially

$$f : \forall x : D[F(x) \longrightarrow B], p : \exists x : DF(x) \Rightarrow \{f, p\} : B$$

where does not occur in B. We can of course define the operation $\{f, p\}$. Namely, $\{f, p\} = f(p1)(p2)$.

But the type theoretic form of $\exists$-Elimination yields a principle of logic not derivable in predicate logic of any order, viz. the *Principle of Choice*

$$\forall x : A \exists y : BG(x, y) \longrightarrow \exists z : (A \longrightarrow B) \forall x : AG(x, zx)$$

For let v be a variable of type $\forall x : A \exists y : BF(x, y)$. Set $f(v) = \lambda x : Avx1$ and $g(v) = \lambda x : Avx2$. Then $(f(v), g(v))$ is a deduction of $\exists z : (A \longrightarrow B) \forall x : AG(x, zx)$ and so

$$\lambda x' : [\forall x : A \exists y : BF(x, y)](f(x'), g(x'))$$

is the required absolute deduction.

Remark 10 *The Principle of Choice was absolutely deduced constructively, i.e. without using Double Negation Elimination. Also we can constructively and absolutely deduce*

$$\forall x{:}A\neg F(x) \longleftrightarrow \neg\exists x{:}AF(x)$$

and so by contraposition

$$\neg\forall x{:}A\neg F(x) \longleftrightarrow \neg\neg\exists x{:}AF(x)$$

So, in the classical system, using $\neg\neg$-Elimination, we have $\exists \longleftrightarrow \neg\forall\neg$. So from Choice we have

$$\forall x{:}A\neg\forall y{:}B\neg G(x,y) \;\longrightarrow\; \neg\forall z{:}(A \longrightarrow B)\neg\forall x{:}AG(x,zx)$$

in the classical system. Providing that $G(u,v)$ does not contain $\exists$, this quantifier does not occur in the formula at all: it is expressed entirely in terms of $\forall$ and $\mathbf{0}$ and the logical constants in $G(u,v)$. Yet the formula cannot be deduced in the fragement of the theory of types obtained by dropping $\exists$.

The principle of double negation elimination is derivable in constructive logic in some cases.

Definition 17 *A formula A is called* stable *iff $\neg\neg A \longrightarrow A$ is constructively absolutely deducible.*

Proposition 18 *a) $\mathbf{2}$ is stable.*

b) $\mathbf{T}s$ is stable for all terms $s{:}\mathbf{2}$.

c) $\neg\neg\forall x{:}AF(x) \longrightarrow \forall x{:}A\neg\neg F(x)$ is constructively absolutely deducible. Hence, if $F(v)$ is stable, then so is $\forall x{:}AF(x)$.

d) $\neg\neg(A \wedge B) \longrightarrow (\neg\neg A$ and $\neg\neg(A \wedge B) \longrightarrow \neg\neg B)$ are constructively absolutely deducible. So $A \wedge B$ is stable if both A and B are.

Proof. First, we note that:
I. If $f : A \longrightarrow B$, then $f' = \lambda x{:}\neg B\lambda y{:}Ax(fy)$ is a deduction of $\neg B \longrightarrow \neg A$. Hence, iterating once more, f'' is a deduction of $\neg\neg A \longrightarrow \neg\neg B$.

II. If $f:A \longrightarrow B$ and $g:B \longrightarrow C$, then $g \circ f = \lambda x:Ag(fx)$ is a deduction of $A \longrightarrow C$.
III. If $f:B \longrightarrow \forall x of type A F(x)$ and $g:\forall x:A[F(x) \longrightarrow G(x)]$ then $\lambda z:B\lambda x:A[gx \circ fx]$ is a deduction of $B \longrightarrow \forall x:AG(x)$.

Now we prove the proposition.

a) $\lambda x:\neg\neg\mathbf{2}\top$ is a deduction of $\neg\neg\mathbf{2} \longrightarrow \mathbf{2}$.

c) The second part of c) follows from the first using III. To prove the first part: Let v be of type A. $\lambda z[\forall x:AF(x)]zv$ is of type $\forall x:AF(x) \longrightarrow F(v)$. So $\neg\neg\forall x:AF(x) \longrightarrow \neg\neg F(v)$ has a deduction $t(v)$ by I. So $\lambda z:[\neg\neg\forall x:AF(x)]\lambda y:Af(y)z$ is a deduction of $\neg\neg\forall x:AF(x) \longrightarrow \forall x:A\neg\neg F(x)$. b) $s = \lambda x:\neg\neg\mathbf{T}\top.\mathbf{o}$ is a deduction of $\neg\neg\mathbf{T}\top \longrightarrow \mathbf{T}\top$. $t = \lambda x:\neg\neg\mathbf{0}[x(\lambda y:\mathbf{0}N(\mathbf{0},y)$ is a deduction of $\neg\neg\mathbf{T}\bot \longrightarrow \mathbf{T}\bot$. So, if $F(v) = \neg\neg\mathbf{T}v \longrightarrow \mathbf{T}v$, then $[s,t,r]_{F(x)}$ is a deduction of $\neg\neg\mathbf{T}r \longrightarrow \mathbf{T}r$.

d) $\lambda x:A \wedge Bx1$ is a deduction of $A \wedge B \longrightarrow A$ and $\lambda x:A \wedge Bx2$ is a deduction of $A \wedge B \longrightarrow B$. So by I we have the deductions of $\neg\neg(A \wedge B) \longrightarrow (\neg\neg A$ and $\neg\neg(A \wedge B) \longrightarrow \neg\neg B)$. By II, if A and B are stable, we then obtain deductions s and t of $\neg\neg(A \wedge B) \longrightarrow A$ and $\neg\neg(A \wedge B) \longrightarrow B)$. So $\lambda x:\neg\neg(A \wedge B)(sx,tx)$ is the required deduction of $\neg\neg(A \wedge B) \longrightarrow (A \wedge B)$.

Remark 11 *Note that the proof of d) that the stability of A and B implies the stability of $\exists x:AB$ does not extend to the general case $\exists x:AF(x)$, where v occurs in $F(v)$. A and $F(v)$ may be stable without $\exists x:AF(x)$ being stable. Indeed, that is why there is adifference between classical and constructive logic.*

If $s:A$ and $t:B$, then $(\lambda x:At,\lambda x:Bs)$ is a deduction of $A \longleftrightarrow B$ and $(\lambda x:\neg Axs,\lambda x:\mathbf{0}N(\neg A,x))$ is a deduction of $\neg A \longleftrightarrow \mathbf{0}$. In particular, there are deductions of

$$\neg\mathbf{0} \longleftrightarrow \mathbf{1} \qquad \neg\mathbf{1} \longleftrightarrow \mathbf{0} \tag{4.8}$$

$(\lambda x:A\lambda y:\mathbf{1}x,\lambda x:(\mathbf{1} \longrightarrow A)x\mathbf{o})$ is a deduction of

$$A \longleftrightarrow (\mathbf{1} \longrightarrow A) \tag{4.9}$$

Recall the abbreviation

$$\langle A,B\rangle s := (\mathbf{T}s \longrightarrow A) \wedge (\neg\mathbf{T}s \longrightarrow B)$$

when $s:\mathbf{2}$. Thus

$$\langle A,B\rangle\top := (\mathbf{1}\longrightarrow A)\wedge(\neg\mathbf{1}\longrightarrow B)$$

$$\langle A,B\rangle\bot := (\mathbf{0}\longrightarrow A)\wedge(\neg\mathbf{0}\longrightarrow B)$$

By (4.8) and (4.9) therefore, there are deductions

$$I_{AB}:\langle A,B\rangle\top\longleftrightarrow A\qquad II_{AB}:\langle A,B\rangle\bot\longleftrightarrow B$$

We defined

$$A\vee B := \exists x:\mathbf{2}\langle A,B\rangle x$$

So $\lambda x:A(\top,I_{AB}2x)$ and $\lambda x:B(\bot,II_{AB}2x)$ are, respectively, deductions of the usual $\vee$-Introduction rules

$$A\longrightarrow A\vee B\qquad B\longrightarrow A\vee B$$

$\lambda x:A\lambda y:\neg B\lambda z:Ay(xz)$ is a deduction of

$$(A\longrightarrow B)\longrightarrow(\neg B\longrightarrow\neg A)$$

So by $\vee$-Introduction we have deductions s and t of

$$\neg(A\vee\neg A)\longrightarrow\neg A\qquad\neg(A\vee\neg A)\longrightarrow\neg\neg A$$

So $r=\lambda x:\neg(A\vee\neg A)(tx)(sx)$ is a deduction of

$$\neg\neg(A\vee\neg A)$$

Note that this deduction is constructive. But now, using Double Negation Elimination, we obtain the *Law of Excluded Middle*

$$A\vee\neg A$$

So in classical logic, for any formula $F(v)$, where v is a variable of type A, there is a deduction of $F(v)\vee\neg F(v)$ and so there is a deduction

$$p\ :\ \forall x:A[F(x)\vee\neg F(x)]$$

Let $f=\lambda x:A(px1)$ and $g=\lambda x:A(px2)$. Then $f:A\longrightarrow\mathbf{2}=\mathbf{P}(A)$. Let $u:A$. Then gu is a deduction of $\langle F(u),\neg F(u)\rangle(fu)=$

$$[\mathbf{T}(fu)\longrightarrow F(u)]\wedge[\neg\mathbf{T}(fu)\longrightarrow\neg F(u)$$

From the second conjunct, in classical logic, we obtain a deduction of $F(u) \longrightarrow \mathbf{T}(fu)$ Recalling that $u\epsilon_A f$ is $\mathbf{T}(fu)$, we thus have a deduction $r(u)$ of

$$u\epsilon_A f \longleftrightarrow F(u)$$

So $(f, \lambda x{:}Ar(x))$ is a deduction of the *Comprehension Principle*

$$\exists z{:}\mathbf{P}(A)\forall x{:}A[x\epsilon z \longleftrightarrow F(x)$$

We can denote f by

$$\{x{:}A \mid F(x)\}$$

Thus every formula $F(t)$ is equivalent in classical logic to one of the form $t\epsilon\{x{:}A \mid F(x)\}$, i.e. $\mathbf{T}(\{x{:}A \mid F(x)\}t)$.

Remark 12 *The Comprehension Principle follows from the Law of Excluded Middle. The converse is also true: Let B be any formula. Using the Comprehension Principle, there is a $b : \mathbf{P}(\mathbf{2})$ such that $\mathbf{T}(b\bot) \longleftrightarrow B$. Hence, $\neg\neg\mathbf{T}(b\bot) \longleftrightarrow \neg\neg B$. But by Proposition* **??**, *$\neg\neg\mathbf{T}(b\bot) \longrightarrow \mathbf{T}(b\bot)$ and so $\neg\neg B \longrightarrow B$.*

4.6 The Theory of Arithmetical Types

We introduce the constant $\mathbf{N}$ to denote the type of the natural numbers. The rules of $\mathbf{N}$-*Introduction* are

$$0{:}\mathbf{N}$$

$$t{:}\mathbf{N} \;\Rightarrow\; S(t){:}\mathbf{N}$$

0 is intended to denote the least natuiral number and $\lambda x{:}\mathbf{N}S(x)$ the successor function. We will write $1 = S(0), 2 = S(1)$, etc. The rule of $\mathbf{N}$-*Elimination* is

$$s{:}F(0), g{:}\forall x{:}\mathbf{N}[F(x) \longrightarrow F(S(x))], t{:}\mathbf{N} \;\Rightarrow\; R(s,g,t){:}F(t)$$

Note that $R(s,g,0)$ and s have the same type $F(0)$ and $gtR(s,g,t)$ is of the same type as $R(s,g,S(t))$. In fact, the *primitive recursion operator* $\mathbf{R} =$

$$\lambda x{:}F(0)\lambda y{:}(\forall u{:}\mathbf{N}[F(u) \longrightarrow F(S(u))])\lambda z{:}\mathbf{N}R(x,y,z)$$

is defined by the *conversion rules*

$$R(s,g,0)\ CONV\ s \qquad R(s,g,S(t))\ CONVgtR(s,g,t)$$

These rules express the principle of *primitive recursive definition*. Note that $\mathbf{R}$ is of type

$$F(0) \longrightarrow [\forall x{:}\mathbf{N}[F(x) \longrightarrow F(S(x))] \longrightarrow \forall x{:}\mathbf{N}F(x)]$$

i.e. the axiom schema of *Mathematical Induction* for $F(x)$. Of corse, $\mathbf{R}$ is not in general closed, since it contains all free variables in $F(x)$.Taking $F(x)$ to be $x \in_{\mathbf{N}} v$, where v is a variable of type $\mathbf{P}(\mathbf{N})$, $\lambda y{:}\mathbf{P}(\mathbf{N})\mathbf{R}$ is a deduction of the second-order *Axiom of Mathematical Induction*:

$$\forall y{:}\mathbf{P}(\mathbf{N})\{0 \in y \longrightarrow [\forall x{:}\mathbf{N}[x \in y \longrightarrow S(x) \in y] \longrightarrow \forall x{:}\mathbf{N}x \in y]\}$$

Using the Comprehension Principle, each instance of the axiom schema of induction can be derived from the axiom.

We extend the notion of definitional equality to arithmetic types by adding the clause

$$s \equiv s', g \equiv g', t \equiv t' \ \Rightarrow\ R(s,g,t) \equiv R(s',g',t')$$

Let $s : B$ and $g : B \longrightarrow B$. Then taking $h = \lambda x : \mathbf{N}g$, $F(x) \equiv B$ and $f = \lambda x{:}\mathbf{N}R(s,h,x)$, we have the principle of *definition by iteration*

$$f0 \equiv s \qquad fS(t) \equiv g(ft)$$

We want to deine the formula $v =_{\mathbf{N}} u$, where v and u are variables of type $\mathbf{N}$. First, we define the predecessor function Pr by iteration:

$$Pr0 \equiv 0 \qquad PrS(t) \equiv t$$

Again by iteration we define the function $s - t$, which denotes subtraction of t from s when $t \leq s$ and has the value 0 otherwise:

$$s - 0 \equiv s \qquad s - S(t) \equiv Pr(s - t)$$

Addition is defined as usual by the iteration

$$s + 0 \equiv s \qquad s + S(t) \equiv S(s + t)$$

The meaning of $s =_{\mathbf{N}} t$ should be $(s - t) + (t - s) \equiv 0$. So define

$$\theta 0 \equiv \top \qquad \theta S(t) \equiv \bot$$

Then we may define

$$u =_{\mathbf{N}} v := \theta[(u - v) + (v - u)]$$

The usual axioms of identity

$$v =_{\mathbf{N}} v$$

$$v =_{\mathbf{N}} u \wedge v =_{\mathbf{N}} w \longrightarrow v =_{\mathbf{N}} w$$

as well as Dedekind's axioms

$$\neg S(t) =_{\mathbf{N}} 0$$

$$S(s) =_{\mathbf{N}} S(t) \longrightarrow s =_{\mathbf{N}} t$$

are deducible using mathematical induction. We will drop the subscript $\mathbf{N}$ in $=_{\mathbf{N}}$ from now on when no consusion will result (i.e. when it is clear that we are talking about a formula and not about two terms being identical). Note that a deduction of $s = s$ is at the same time a deduction of $s = t$ when $s \equiv t$. And so we also have *as formulas* the principle of definition by primitive recursion:

$$R(s, g, 0) = s \qquad R(s, g, S(t)) = gtR(s, g, t)$$

The proofs of the Church-Rosser Theorem and the Well-Foundedness Theorem extend easily to the theory of arithmetical types. The definition of 1-reduction needs to be extended by adding the clause

$$s\ 1 - RED\ s', g\ 1 - RED\ g', t\ 1 - RED\ t' \ \Rightarrow\ R(s, g, t)\ 1 - RE\ R(s', g', t')$$

Otherwise the definitions of $n - RED$ and RED are the same. To extend the proof of the Church-Rosser Theorem, it suffices to extend the proof of Lemma 12 to the theory of arithmetic types. So let $r\ CONV\ s$ and $r\ 1 - RED\ t$, where s and t are distinct. We need a u such that s and $t\ 1 - RED\ u$. The only new case is where r is $R(s, q, 0)$ or $R(p, q, S(o)$ and in the second case s is $qoR(p, q, o)$. In the first case, t is $R(s', q', 0)$ and so u can be s'. In the second case, t is $R(p', q', S(o'))$ and so u can be $q'o'R(p', q', o')$. The remainder of the

proof of the Church-Rosser Theorem and the consequence that every term and formula has at most one normal form goes through unchanged.

The definition of a computable term of type A is unchanged. The only modification of the proof of the Computability Theorem is that we have to add the case of terms of the form $R(s, g, t)$, where we are assuming that all c-instances of s, g and t are computable. Clearly it suffices to prove that $R(s, g, t)$ is computable if s, g, and t are. So assume the latter. So t is well-founded and therefore has a unique normal form t'. t' has the form $S(S(\cdots(S(r))\cdots))$, with some $m \geq 0$ nested initial occurrences of S's, which we call the *S-number* of t. We prove by induction on m that $R(s, g, t)$ is computable. Let $p = R(s, g, t)t_1 \cdots t_n$ be a c-extension of $R(s, g, t)$. We need to show that every sequence

$$p = p_0 > p_1 > \cdots$$

is finite.

Case 1. Every term in the sequence is obtained by internal reduction of the preceding term and so is of the form $R(s", g', t')t'_1 \cdots t'_n$, the result is immediate from the computability and hence well-foundedness of $s, g, t, t_0, \ldots, t_n$.

Case 2 Some term in the sequence is $R(s', g', 0)t'_1 \cdots t'_n$ and the next term is $s't''_1 \cdots t''_n$. Since s' and therefore s'' is computable, then sequence then has to be finite.

Case 3. Some term of the sequence is $R(s', g', S(t'))t'_1 \cdots t'_n$ and the next term is $g't'R(s', g', t')t''_1 \cdots t''_n$. But t' has the S-number $m - 1$ and so by the induction hypothesis, $s', g', t', R(s', g', t'), t''_1, \ldots, t''_n$ are all computable. The finiteness of the sequence follows.

The definition of strong normal form extends to arithmetical terms and formulas by adding the clause

$$SNF(R(s, g, t)) = R(SNF(s), SNF(g), SNF(t))$$

when $R(s, g, t)$ is in normal form. The proof that the relation of definitional equality is decidable now goes through as before.

4.7 Variable-Free Formalization of Type Theory

The reduction of the lambda calculus to the theory of combinators in [Schönfinkel, 1924] applies to *positive implicational logic*, i.e. to the typed lambda cal-

culus, where the types are built up from atomic types by means of the operation $A \longrightarrow B$, to show that the lambda operator can be eliminated in favor of combinators K and S of each type $A \longrightarrow (B \longrightarrow A)$ and $(A \longrightarrow (B \longrightarrow C)) \longrightarrow ((A \longrightarrow B) \longrightarrow (A \longrightarrow C))$, respectively.(This observation is essentially contained in the discussion of the so-called theory of functionality in Chapters 9 and 10 of [Curry and Feys, 1958].) We will extend that result to the theory of types. To extend the treatment of $\longrightarrow$ to $\forall$ we shall need a generalized form of the combinators K and S, and to deal with $\exists$ we will need to introduce a new form of the combinator S (whose type turns out to be a general form of the Axiom of Choice). But also in the present context, if we are to eliminate variables, then not only the lambda operator for forming terms, but also quantification as a variable-binding operation for forming formulas must be analyzed away; so we will need an analogue of the combinators for formulas.

As usual, we shall write st for the value $s(t)$ of the function s for the argument t; so rst is $(r(s))(t)$, etc.

Let v be a free variable of type A. We wish to rewrite the formulas $B(v)$, $\forall x\!:\!A.B(x)$ and $\exists x\!:\!A.B(x)$, respectively, as $B'v$, $\forall B'$ and $\exists B'$, where B' is a type-valued function on A. If $t(v)$ is a term of type $B(v)$, which we express by

$$t(v)\!:\!B(v)$$

then $\lambda x\!:\!A.t(x)$ is a term of type $\forall x\!:\!A.B(x)$, denoting a function on A whose value for $s\!:\!A$ is $t(s)\!:\!B(s)$. We wish to rewrite the terms $t(v), \lambda x\!:\!A.t(x)$, respectively, as $t'v\!:\!B'v$ and $t'\!:\!\forall B'$. Thus, a two-quantifier formula

$$Q_1 x\!:\!AQ_2 y\!:\!B(x).C(x,y)$$

where Q_1 and Q_2 are quantifiers, is to be rewritten as

$$Q_1 x\!:\!AQ_2 y\!:\!B(x).C(x)'y$$

or

$$Q_1 x\!:\!AQ_2 y\!:\!B'x.C''xy$$

or simply

$$Q_1 Q_2 C''$$

C'' is a function defined on A such that $C''s$ is a type-valued function defined on $B's$ for all $s:A$. Let u and v be free variables of types A and B(u),

respectively. A term $t(u, v)$ of type $C(u, v)$ should be rewritten as $t''uv$, where t'' is of type $\forall\forall C''$.

To discuss the general case, we need a definition.

Definition 18 *The notion of a* base of functionals *is defined by induction:*

- *The null sequence is a base.*
- *If A is a type and $\psi_1, \ldots, F_n$ are functions defined on A such that, for each $t : A$, $\langle F_1 t, \ldots, F_n t\rangle$ is a base, then the sequence $\langle A, F_1, \ldots, F_n\rangle$ is a base.*

When $\langle A, F_1, \ldots, F_n\rangle$ is a base, the base $\langle A, F_1, \ldots, F_{n-1}\rangle$ is uniquely determined by the functional F_n. As an example, in the two-quantifier example above, $\langle A, B', C''\rangle$ is a base. More generally, an n-quantifier formula

$$Q_1 x_1 : A_1\ Q_2 x_2 : A_2(x_1) \cdots Q_n x_n : A_n(x_1, \ldots, x_{n-1}).B(x_1, \ldots, x_n) \tag{4.10}$$

is to be rewritten as

$$Q_1 x_1 : A_1\ Q_2 x_2 : A'_2 x_1 \cdots Q_n x_n : A_n^{(n-1)} x_1 \cdots x_{n-1}.B^{(n)} x_1 \cdots x_n$$

where $\langle A, A'_2, \ldots A_n^{(n-1)}, B^{(n)}\rangle$ is a base, or simply as

$$Q_1 \cdots Q_n B^{(n)}.$$

If $v_1, v_2, \ldots, v_n$ are free variables of types $A_1, A_2(v_1), \ldots, A_n(v_1, \ldots, v_n)$, respectively, then a term $t(v_1, v_2, \ldots, v_n)$ of type $B(v_1, v_2, \ldots, v_n)$ is to be rewritten as $t^{(n)} v_1 v_2 \cdots v_n$, where $t^{(n)}$ is of type $\forall\forall \cdots \forall B^{(n)}$.

In order to carry out this analysis, we need to introduce a formalism in which we can represent functionals and objects which depend upon free variables.

4.7.1 The Calculus

We must simultaneously define three notions:

- The notion of a *base of formulas.*
  - Bases are finite sequences whose members are called *formulas.*

- If $\langle \vec{F}, G \rangle$ is a base, then $\vec{F}$ is called the *base* of G and denoted by $Base(G)$.
- When $\langle A \rangle$ is a base, A is called a *(formula) type*.
- A base of formulas is intended to denote a base of functionals for suitable values of the free variables.
- With a formula we may associate a *rule of conversion*, which specifies the meaning of the formula. $FCONVG$ means that the formula F converts to the formula G according to the rules of conversion.

- The notion of a *term of type* A, where A is a type.
  - That t is a term of type A is expressed by $t\!:\!A$.
  - With a term we may associate a *rule of conversion*, which specifies the meaning of the term. $sCONVt$ means that the term s converts to the term t according to the rules of conversion.

- The notion of *definitional equality* between two terms or between two functionals.
  - We denote this relation by $\equiv$.
  - We may specify at once that, for terms s and t, $s \equiv t$ is defined to mean $s\ RED\ r \wedge t\ RED\ r$ for some r, where the relation RED is defined in terms of the rules of conversion: call an occurrence of a formula or term X in a formula or term U *external* if it is not in a part of U of the form $v_n(A)$. (When A is a formula, $v_n(A)$ will be introduced as a variable of type A.) For formulas or terms U and V, $U > V$ will mean that V is obtained by replacing some external occurrence X of U by Y, where $X\ CONV\ Y$. RED is the least reflexive and transitive relation which includes $>$.
  - For formulas F and G, $F \equiv G$ will mean that the base of F and the base of G are of the same length $n \geq 0$ and, for some distinct new symbols $x_1, \ldots, x_n$, $Fx_1 \cdots x_n$ and $Gx_1 \cdots x_n\ RED$ to a common expression.[1]

[1]Notice that, on our definition, variables $v_n(A)$ are always in *normal form*, where a formula or term X is in normal form iff there is no Y such that $X > Y$. Thus, even when the distinct types A and B are $\equiv$, $v_n(A) \not\equiv v_n(B)$.

– We may also specify at once that the type of a term is to be determined only to within definitional equality. Thus, as a part of the definition of the type relation we specify that

$$t{:}A \wedge A \equiv B \longrightarrow t{:}B.$$

It will follow that

$$s \equiv t \wedge s{:}A \longrightarrow t{:}A.$$

If $\vec{Y} = \langle Y_1, \ldots, Y_n \rangle$, then $\langle X, \vec{Y} \rangle$ will denote $\langle X, Y_1, \ldots, Y_n \rangle$, $\langle \vec{Y}, Z \rangle$ will denote $\langle Y_1, \ldots, Y_n, Z \rangle$, $\vec{Y}t$ will denote $\langle Y_1 t, \ldots, Y_n t \rangle$, etc. *Atomic Formulas*

If $\vec{F}$ is a base of formulas none of which contains free variables, then $R_n(\vec{F})$ is an atomic formula with base $\vec{F}$ for each n. There may be conversion rules associated with an atomic formula. *Instantiation* If G has base $\langle A, \vec{F} \rangle$ and $t{:}A$, then Gt is a formula with base $\vec{F}t$. *Quantification*

If H has base $\langle \vec{F}, G \rangle$, then $\forall H$ and $\exists H$ are formulas with base $\vec{F}$.

- If $\vec{F}$ is not null and Q is a quantifier, then we have the conversion rule

$$(QH)t \; CONV \; Q(Ht)$$

- The *(universal) closure* of a formula H is

$$H^* = \forall \cdots \forall H$$

where the number of $\forall$'s is the length of the base of H. Thus, H^* is a type.

Dummy Argument Places

If $\langle \vec{F}, G \rangle$ and $\langle \vec{F}, H_1, \ldots, H_k \rangle$ are bases, then so is $\langle \vec{F}, G, H_1[G], \ldots, H_k[G] \rangle$.

- The rules of conversion for $H_i[G]$ $(i = 1. \ldots, k)$ are:
  - If $\vec{F} \neq \emptyset$

$$H_i[G]t \; CONV \; H_i t[Gt]$$

  - If $\vec{F} = \emptyset$

$$H_i[G]t \; CONV \; H$$

- Abbreviations: Let $Base(G) = Base(H)$

$$G \longrightarrow H = \forall(H[G])$$

$$G \wedge H = \exists(H[G])$$

Transposition of Argument Places

If $\langle \vec{E}, F, G, H_1, \ldots, H_k \rangle$ is a base, then so is $\langle \vec{E}, \forall G, F[\forall G], H_1\{1\}, \ldots, H_k\{k\} \rangle$.

The subscript 'i' in H_i is meta-notation, marking which formula in the base we are refering to; the '$\{i\}$', on the other hand, is part of the syntax of the formula $H_i\{i\}$. The rules of conversion are:

- If $\vec{E} \neq \emptyset$

$$H_i\{i\}t \; CONV \; H_it\{i\}$$

- If $\vec{E} = \emptyset$

$$H_i\{i\}st \; CONV \; H_it(st)$$

REMARK. In the second case, note that s must be a term of type $\forall G$ and t must be of type $F[\forall G]s$, i.e. of type F. Since G has base F, st is defined and is of type Gt, by the principle of $\forall$ Elimination in §1.8 below. So $H_it(st)$ is defined. *Variables*

For each type A and $n \geq 0$

$$v_n(A) : A$$

$v_n(A)$ is called a *free variable* of basic type A. Note that A is a syntactical part of $v_n(A)$. A variable of basic type A may be denoted by $v(A)$, $v'(A)$, etc. *Constants*

If A is a type containing no variables, zero or more constant terms of type A may be introduced.

Quantifier Elimination

Let $\langle A, F \rangle$ be a base.

- $\forall$ Elimination

$$s : A, \; t : \forall F \Longrightarrow ts : Fs$$

- $\exists$ Elimination

$$p : \exists F \Longrightarrow (p1) : A, \; (p2) : F(p1)$$

Existential Quantifier Introduction
Let H have base $\langle \vec{F}, G \rangle$.

$$P(H) : (H \longrightarrow \exists(H[G]))^*$$

The conversion rules for $\exists$ are

- If $\vec{F} \neq \emptyset$

$$P(H)t \ CONV \ P(Ht)$$

- If $\vec{F} = \emptyset$

$$P(H)st1 \ CONV \ s$$
$$P(H)st2 \ CONV \ t$$

The Combinator K
Let G and H have base $\vec{F}$.

$$K(G, H) : (G \longrightarrow (H \longrightarrow G))^*$$

The conversion rules associated with K are

- If $\vec{F} \neq \emptyset$

$$K(G, H)t \ CONV \ K(Gt, Ht)$$

- If $\vec{F} = \emptyset$

$$K(G, H)st \ CONV \ s$$

The Combinators $S_\forall$ and $S_\exists$
Let H have base $\langle \vec{E}, F, G \rangle$ and let Q be a quantifier $\forall$ or $\exists$. Then

$$S_Q(H) : (\forall QH \longrightarrow Q\forall(H\{1\})^*$$

The conversion rules are

- If $\vec{E} \neq \emptyset$

$$S_Q(H)t \ CONV \ S_Q(Ht)$$

- Assume that $\vec{E} = \emptyset$ and let $r : \forall QH$. So $H\{1\}$ has base $\langle \forall G, F[\forall G] \rangle$. Let $s : \forall G$ and $t : F[\forall G]s$. So $t : F$ and

$$S_Q(H)r : \forall Q(H\{1\})$$

– Let $Q = \forall$.

$$S_\forall(H)r : \forall\forall(H\{1\})$$

$S_\forall(H)rst$ must be defined to be of type $H\{1\}st$, i.e. of type $Ht(st)$. But $rt : \forall(Ht)$, $st : Gt$ and so $rt(st) : Ht(st)$. Thus, we may define $S_\forall(H)rst$ by the conversion rule

$$S_\forall(H)rst \; CONV \; rt(st)$$

– Let $Q = \exists$.

$$S_\exists(H)r : \exists\forall(H\{1\}$$

Thus

$$S_\exists(H)r1 : \forall G$$

$$S_\exists(H)r2 : \forall H\{1\}(S_\exists(H)r1)$$

So

$$S_\exists(H)r1t : Gt$$

$$S_\exists(H)r2t : Ht(S_\exists(H)r1t)$$

But $rt : \exists Ht$ and so $rt1 : Gt$ and $rt2 : Ht(rt1)$. So we may define $S_\exists(H)$ by the conversion rules

$$S_\exists(H)r1t \; CONV \; rt1$$

$$S_\exists(H)r2t \; CONV \; rt2$$

We have completed the description of the calculus.

Notice that the type of $S_\exists(H)$ is a general form of the Axiom of Choice: for example, let H have base $\langle A, B[A]\rangle$. Then $H\{1\}$ has base $\langle A \longrightarrow B, A[A \longrightarrow B]\rangle$ and the type $\forall\exists H \longrightarrow \exists\forall H\{1\}$ may be written as

$$\forall x : A \exists y : B Hxy \longrightarrow \exists f : A \longrightarrow B \forall x : A Hx(fx)$$

4.7.2 Some properties of the calculus

Let $Var(X)$ denote the set of variables in the formula or term X.

Definition 19 *The type B of the term t is* suitable *for t iff $Var(B) = Var(t) - \{t\}$.*

Lemma 16 *The following facts are easily derived.*

1. *Every variable in a formula in the base of F is in F.*
2. *Every term has a suitable type.*
3. *If G and H have bases $\vec{E}$ and $\langle \vec{E}, F \rangle$, respectively, then*

$$\forall(H[G]) \equiv (\forall H)[G]$$

.

4. *If G and H both have base $\langle \vec{E}, F \rangle$, then*

$$H[G]\{1\} \equiv H[\forall G]$$

5. *Let F, G and H all have base $\vec{E}$. Then*

$$H[F][G[F]] \equiv H[G][F]$$

.

Assuming that there are no further conversion rules, we may prove in the usual way

Theorem 8 *Church-Rosser Theorem If the formula or term X reduces to Y and to Z, then Y and Z reduce to some U. In particular, every term or formula has at most one normal form.*

Theorem 9 *Well-foundedness Theorem If X is a formula or term, then every sequence $X > Y > \cdots$ is finite. In particular, every formula or term has a normal form.*

In view of these two theorems, the relation $\equiv$ between formulas and terms is decidable. We will not discuss general conditions on extensions of the calculus obtained by adding new conversion rules under which the Church-Rosser and Well-foundedness Theorems are preserved, since the main result of this paper, the Explicit Definition Theorem below, will be preserved by any such extension.

4.7.3 Identity Function

Let G and H have base $\vec{F}$ and let $S = S_\forall(H[G])$. Then S is of type

$$(\forall(G \longrightarrow H) \longrightarrow \forall\forall(H[G]\{1\}))^*$$

which, by 3 and 4 of the Lemma is $\equiv$ to

$$(\forall(G \longrightarrow H) \longrightarrow (\forall G \longrightarrow \forall H))^* \tag{4.11}$$

Let G be $B[A]$ and let H be $C[A]$. By 5 of the Lemma, (4.11) is $\equiv$ to

$$(A \longrightarrow (B \longrightarrow C)) \longrightarrow ((A \longrightarrow B) \longrightarrow (A \longrightarrow C)) \tag{4.12}$$

So

$$S : (A \longrightarrow (B \longrightarrow C)) \longrightarrow ((A \longrightarrow B) \longrightarrow (A \longrightarrow C))$$

Set $B = A \longrightarrow A$, $C = A$, $K_1 = K(A, B)$ and $K_2 = K(A, A)$. Then $K_1 : A \longrightarrow (B \longrightarrow C)$ and $K_2 : A \longrightarrow B$. Set

$$I_A = SK_1K_2$$

Then $I_A : A \longrightarrow C$, i.e. $I_A : A \longrightarrow A$. Let $t : A$.

$$I_At = SK_1K_2t \equiv K_1t(K_2t) \equiv t.$$

Thus I_A is the identity function on A.

Notice that the combinators for positive implicational logic really are a special case of $K(G, H)$ and $S_\forall(H)$. Namely, they are $K(A, B)$ of type $A \longrightarrow (B \longrightarrow A)$ and $S_\forall(C[A][B[A]])$ of type (4.12).

4.7.4 Explicit Definition Theorem

Definition 20 *A variable v is* unfettered *in the term t (formula F) iff for every variable $v(A)$ occuring in t (F), v does not occur in A.*

Note: If B is a suitable type for the term t, then v is unfettered in t iff it is unfettered in B.

Theorem 10 (Explicit Definition Theorem) *Let $v = v(A)$.*

- *If* $\langle F_1, \ldots, F_n \rangle$ *is a base and* v *is unfettered in* F_n*, then there is a base* $\langle A, F_1', \ldots, F_n' \rangle$ *such that* $Var(F_i') \subseteq Var(F_i) - \{v\}$ *and*

$$F_i' v \ RED \ F_i$$

- *If* $t : B$ *and* v *is unfettered in* t *and in* B*, then there is a* $t' : \forall B'$ *such that* $Var(t') \subseteq Var(t) - \{v\}$ *and*

$$t' v \ RED \ t$$

Note: If $B \equiv C$, then $B' \equiv C'$. So, in particular, given a term t in which v is unfettered, we need only find one type C of t in which v is unfettered and construct $t' : \forall C'$. If B is another type of t in which v is unfettered, then t' will be of type $\forall B'$ as well.

Proof. The proof is by induction on the definition of the base or term.
Case 1. Assume that v does not occur in F_n. Then it does not occur in any F_i. Set $F_i' = F_i[A]$.
Case 2. Assume that v is not in t and let B be a suitable type for t. Then v is not in B and so $B' = B[A]$. Set $t' = K(B, A)t$, which is of type $\forall B' = A \longrightarrow B$ and $t' v \ CONV \ t$.

In the remaining cases, we may assume that v occurs in the formula or term in question.
Case 3. Let us assume that F' is defined for $F = G$, $F = H$ and for every formula F in the base of G or H. Then we may clearly set

$$(QH)' = QH'$$

$$H[G]' = H'[G']$$

$$H\{n\}' = H'\{n\}$$

$$P(H)' = P(H')$$

$$K(G, H)' = K(G', H')$$

$$S_Q(H)' = S_Q(H')$$

For example, $H[G]' v = H'[G'] v \, CONV \, H' v[G' v] \, RED \, H[G]$. And $K(G, H)' v = K(G', H') v \ CONV \ K(G' v, H' v) \ RED \ K(G, H)$. Note that $K(G, H)'$ is of

type $(G' \longrightarrow (H' \longrightarrow G')^*$, which is $\forall(G \longrightarrow (H \longrightarrow G)^*)'$, so the type is right.

Case 4. Let $F_i = \chi_i s$, where $s : C$ and $\langle C, \chi_1, \ldots, \chi_n\rangle$ is a base. Then $\langle A, C', G'_1, \ldots, G'_n\rangle$ is a base and $t' : \forall C'$. Set $F'_i = G'_i\{i\}t'$. Then

$$F'_i v \ CONV \ G'_i v(t'v) \ RED \ \chi_i t = F_i$$

Case 5. Let H have base $\langle B\rangle$, $f : \forall H$, and $t : B$. We need to define $ft)'$. $f' : \forall\forall \ H'$, $t' : \forall B'$ and $S_\forall(H')f' : \forall\forall H'\{1\}$. $H'\{1\}$ has base $\langle \forall B', A[\forall B'\rangle$. So $S_\forall(H')f't'$ is defined and is of type $\forall H'\{1\}t' \equiv \forall(Ht)'$. So set $(ft)' = S_\forall(H')f't'$. For

$$S_\forall(H')f't'v \ RED \ f'v(t'v) \ RED \ ft$$

Case 6. Let $p : \exists H$, where H has base B. We need to define $(p1)'$ and $(p2)'$. $p' : \forall\exists H'$, where H' has base $\langle A, B'\rangle$. $H'\{1\}$ has base $\langle \forall B', A[\forall B']\rangle$. So $S_\exists(H')p' : \exists\forall(H'\{1\})$. Set $(p1)' = S_\exists(H')p'1$ and $(p2)' = S_\exists(H')p'2$.

$$(p1)'v \ RED \ p'v1 \ REDp1$$

$$(p2)'v \ RED \ p'v2 \ REDp2.$$

The proof is completed.

We may now take $\forall x : A.B(x)$ to be an abbreviation for $\forall B'$, providing the free variable $v = v(A)$ is unfettered in $B(v)$. If v is fettered in B, then B has the form $B(v, u(C(v)))$, where $u(C(v))$ is a variable and v is unfettered in $C(v)$. But in this case, $\forall x : A.B(x)$, i.e. $\forall x : A.B(x, u(C(x)))$ doesn't make any literal sense: $u(C(x))$ does not denote a variable of any particular type. Rather we can only think of it as a dependent variable, depending on the value of x. But then we may more accurately replace $u(C(v))$ by $u(\forall C')v$, eliminating at least one context which fetters v. Iterating this proceedure, we finally transform $B(v)$ into a type $D(v)$ in which v is unfettered and such that $\forall x : A.D(x)$ expresses the only reasonable meaning of $\forall x : A.B(x)$. Similarly, we may restrict $\lambda x : A.t(x)$ to the case in which v is unfettered in $t(v)$; and in that case it is an abbreviation for t'. In this case, the restriction that v be unfettered in $t(v)$ is precisely Gentzen's restriction on his rule $\forall - I$ in the system of natural deduction.

Now we return to the initial discussion of the n-quantifier form. Let $B = B(v_1, \ldots, v_n)$ be a formula and $v_1, \ldots, v_n$ a list of variables including all the variables in B, $v_i = v_i(A_i)$. Assume that the list of variables is in *good order*, i.e. $i < j$ implies that v_j does not occur in A_i. So we may

write $A_i = A_i(v_1, \ldots, v_{i-1})$, displaying all the free variables. Then v_n is unfettered in B and we may apply the Explicit Definition Theorem to obtain B' with base $\langle A_n \rangle$, containing at most the variables v_i for $i < n$ and such that $B'v_n \equiv B$. v_{n-1} is unfettered in B' and so we may construct B'' with base $\langle A_{n-1}, A'_n \rangle$, containing at most the variables v_i for $i < n-1$, such that $B''v_{n-1}v_n \equiv B$. Iterating n times, we obtain the variable-free formula $B^{(n)}$ with base $\langle A_1, A'_2, \ldots, A_n^{(n-1)} \rangle$ such that $B^{(n)}v_1 \cdots v_n \equiv B$. Then (4.10) is precisely $Q_1 \cdots Q_n.B^{(n)}$. We denote B^n by

$$\lambda x_1 : A_1 \cdots \lambda_n x_n : A_n(x_1, \ldots, x_{n-1}).B(x_1, \ldots x_n).$$

Moreover, if $t = t(v_1, \ldots, v_n)$ is a term of type $B(v_1, \ldots, v_n)$, then n applications of the Explicit Definition Theorem yields $t^{(n)} : \forall \cdots \forall B^{(n)}$ with $t^{(n)}v_1 \cdots v_n \equiv t$. We denote t^n by

$$\lambda x_1 : A_1 \cdots \lambda_n x_n : A_n(x_1, \ldots, x_{n-1}).t(x_1, \ldots x_n).$$

For future reference, when $\vec{F} = \langle F_1, \ldots, F_n \rangle$ is a base and G a formula, we write

$$\lambda \vec{x} : \vec{F} = \lambda x_1 : F_1 \cdots \lambda x_n : F_n x_1 \ldots x_{n-1}.$$

and

$$G[\vec{F}] = \lambda \vec{x} : \vec{F}.G.$$

4.8 The Completeness of Intuitionistic First-Order Predicate Logic

We return to the original formalism for type theory, involving bound variables. We will be considering only the constructive system, without the rule

$$p : \not A \Longrightarrow D(p) : A$$

of Double Negation Elimination. Formulas of first-order predicate logic can be regarded as formulas in the theory of types. Choose a type constant D to be the type of individuals. Individual constants are just constants of type D, n-ary function constants are constants of type $D^D = \wedge \cdots \wedge D \longrightarrow D$ (with $n-1$ occurrences of $\wedge$), and n-ary relation constants are constants of sort

D^n. The quantifications $\forall x F(x)$ and $\exists x F(x)$ are taken to be abbreviations for $\forall x : D.F(x)$ and $\exists x : D.F(x)$, respectively.

In §4.5, we showed that every theorem of intuitionistic predicate logic is derivable from D in the theory of types. This follows from the fact that every rule of inference of the Intuitionistic system of natural deduction is derivable in the theory of types. But the converse does not hold. For example, if $p : \exists x F(x)$, then $p1$ is a term of type D but not a term of first-order predicate logic or in general even definitionally equal to one. So, if x occurs in $F(x)$, then $F(p1)$ is not a first-order formula. However, we are able to prove this:

Theorem 11 *Let Γ be a set of first-order formulas and let $t : A$, where A is a first-order formula, t is a constructive term, and all the variables in t are of sign D or are of some sign in Γ. Then A is derivable from Gamma in intuitionistic first-order predicate logic.*

A constructive term all of whose variables are of sign D or whose sign is in Γ will be called a Γ-deduction. When s is a term, $|s|$ denotes the number of occurrences of expressions of the form $p1$ or $p2$ in it. (p need not be a term, since it may contain bound variables which are not bound in p, though of course they are bound in s.) A deduction s is called *special* iff $|s| = 0$.

Lemma 17 *Let t be a normal Γ-deduction of A. Then there are normal special Γ- deductions $t_1, \ldots, t_n$ of first-order formulas $A_1, \ldots, A_n$, respectively, containing only the free variables of t , such that A is first-order deducible from $A_1, \ldots, A_n$.*

The proof is by induction on $|t|$ and, within that, by induction on (the complexity of) t. We can assume that $|t| > 0$.

CASE 1. Assume that $t = s(p1, p2)$ contains a part $p1$ or $p2$, where p is a term. CASE 1a. Let $p : \exists x F(x)$. Then $t' = \lambda x : D \lambda y : F(x).s(x, y)$ is a normal deduction of $\forall x[F(x) \longrightarrow A]$. But A is first-order deducible from $\exists x F(x)$ and $\forall x[F(x) \longrightarrow A]$ and $|p|, |t'| < |t|$. CASE 1b. Let $p : B \wedge C$. Then $t' = \lambda x : B \lambda y : C.s(x, y)$ is a normal deduction of $B \longrightarrow [C \longrightarrow A]$. A is first-order deducible from $B \wedge C$ and $B \longrightarrow [C \longrightarrow A]$ and $|p|, |t'| < |t|$. CASE 1c. Let $p : B \vee C$. $p1 : \mathbf{2}$ and $p2 : \langle B, C \rangle (p1)$. Let u be a new variable of sign B and v a new variable of sign C. There are special terms $r(u)$ and $r'(v)$ of types $\langle B, C \rangle \top$ and $\langle B, C \rangle \bot$, respectively. So $t' = \lambda x : B.s(\top, r(x))$ and $t'' = \lambda y : B.s(\bot, r(y))$ are deductions of $B \longrightarrow A$ and $C \longrightarrow A$, respectively.

But A is first-order deducible from $B \vee C$, $B \longrightarrow A$ and $C \longrightarrow A$. It thus suffices to note that $|p|, |t'|, |t''| < |t|$.

CASE 2. t contains no parts $p1$ or $p2$ where p is a term. CASE 2a. If t is any of $N(A, s), fs, (s, s')$

Bibliography

Browder, F. (ed.) [1976]. *Mathematical Developments arising from Hilbert's Problems*, Proceedings of Symposia in Pure Mathematics, Vol. 28, Providence: American Mathematical Society.

Cantor, G. [1883]. *Grundlagen einer allgemeinen Mannigfaltigheitslehre. Ein mathematisch-philosophischer Versuch in der Lehre des Unendlichen*, Leipzig: Teubner. A separate printing of [**?**] with a subtitloe, preface and some footnotes added. A translation *Foundations of a General Theory of Manifolds: A Mathmatico-Philosophical Investigation into the Theory of the Infinite* by W. Ewald is in [**?**, pp. 639-920].

Curry, H. and Feys, R. [1958]. *Combinatory Logic I, Studies in Logic and the Foundations of Mathematics*, Amsterdam: North-Holland. 2nd edition 1968.

Dedekind, R. [1872]. *Stetigkeit und irrationale Zahlen*, Braunschweig: Vieweg. in [Dedekind, 1932]. Republished in 1969 by Vieweg and translated in [Dedekind, 1963].

Dedekind, R. [1887]. *Was sind und was sollen die Zahlen?'*, Braunschweig: Vieweg. In Dedekind (1932). Republished in 1969 by Vieweg and translated in [Dedekind, 1963].

Dedekind, R. [1932]. *Gesammelte Werke, vol. 3*, Braunschweig: Vieweg. Edited by R. Fricke, E. Noether, and O. Ore.

Dedekind, R. [1963]. *Essays on the Theory of Numbers*, New York: Dover. English translation by W.W. Berman of [Dedekind, 1872] and [Dedekind, 1887].

Fodor, G. [1956]. Eine bemerkung zur theorie der regressiven funktionen, *Acta Scientiarum Mathematicarum Szeged* **17**: 139–142.

Gentzen, G. [1935]. Untersuchungen über das logische Schliessen I, II, *Mathematisce Zeitschrift* **39**: 176–210,405–431.

Gentzen, G. [1936]. Die Widerspruchfreiheit der reinen Zahlentheorie, *Mathematische Annalen* **112**: 493–565.

Gödel, K. [1931]. Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I, *Monatshefte für Mathematik und Physik* **38**: 173–198.

Hilbert, D. [1899]. *Grundlagen der Geometrie. Festschrift zur Freier der Enthüllung des Gauss-Weber Denkmals in Göttingen*, Leipzig: Teubner.

Hilbert, D. [1900]. Mathematische Probleme. Vortrag, gehalten auf dem internationalen Mathematiker-Kongress zu Paris 1900, *Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen* pp. 253–297. English translation by M. W. Newson in *Bulletin of the American Mathematical Society* 8 (1902). Reprinted in [Browder, 1976]. Excerpt in [**?**, volume 1, pp. 1089-1096].

Hilbert, D. [1905]. Über die Grundlagen der Logik und der Arithmetik, *Verhandlungen des Dritten Internationalen Mathematiker-Kongress* .

Hilbert, D. [1926]. Über das Unendliche, *Mathematische Annalen* **95**: 161–90. Translated by Stefan Bauer-Mengelberg in [van Heijenoort, 1967, 367-92].

Mancosu, P. (ed.) [1998]. *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920's*, Oxford: Oxford University Press.

Schönfinkel, M. [1924]. Über die Bausteine der mathematischen Logik, *Mathematische Annalen* **92**: 305–316.

Skolem, T. [1923]. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre, *Matematikerkongressen in Helsingfors 4-7 Juli 1922, Den femte skandinaviske matematikerkongressen, Redogörelse*, pp. 217–232.

Tait, W. [1981]. Finitism, *Journal of Philosophy* **78**: 524–556.

van Heijenoort, J. (ed.) [1967]. *From Frege to Gödel: A Source Book in Mathematical Logic*, Cambridge: Harvard University Press.

Veblen, O. [1908]. Continuous increasing functions of finite and transfinite ordinals, *Transactions of the American Mathematical Society* **9**: 280–292.

Weyl, H. [1921]. Über die neue Grundlagenkrise der Mathematik, *Mathematische Zeitschrift* **10**: 39–79. Translated by P. Mancosu in [Mancosu, 1998].