Kurt Gödel : The Princeton Lectures on Intuitionism

Hämeen-Anttila, Maria

2021

Hämeen-Anttila, M & von Plato, J 2021, Kurt Gödel : The Princeton Lectures on Intuitionism. Sources and Studies in the History of Mathematics and Physical Sciences, Springer Nature Switzerland AG, Cham. https://doi.org/10.1007/978-3-030-87296-0

http://hdl.handle.net/10138/353333
10.1007/978-3-030-87296-0

submittedVersion

Kurt Gödel

The Princeton Lectures on Intuitionism

Edited by Maria Hämeen-Anttila and Jan von Plato

This is a preprint of the following work: M. Hämeen–Anttila and J. von Plato (eds.), Kurt Gödel: The Princeton Lectures on Intuitionism, 2021, Springer, reproduced with permission of Springer Nature Switzerland AG. The final authenticated version is available online at: https:/doi.org/10.1007/978-3-030-87296-0.

Preface

Gödel's Princeton Lectures on Intuitionism of 1941 are preserved in two notebooks written in longhand English. They contain a detailed presentation of his famous functional interpretation of arithmetic and have been studied in connection with the editing of Gödel's *Collected Works*, in particular for the light they shed on a lecture on intuitionistic logic he gave at Yale. The writing is on the whole quite clear, with occasional additions and remarks in German shorthand, and a gap toward the end, at pages 89–106. It turned out in 2017 that the missing pages were inside an envelope in another place, ten reels apart in the microfilm edition of Gödel's manuscripts. That discovery was the starting point of the present edition. Gödel's *Arbeitshefte* or mathematical workbooks, especially number 9, have close connections to the Princeton Lectures. This source and others, including the *Resultate Grundlagen* notebook series, are described in the introduction written by the first editor.

The reader may ask why Gödel didn't publish his lectures at the time, or at least their main results. The answer should be that he failed to achieve his central aim, clearly indicated by the mentioned sources, namely to extend the functional interpretation to the transfinite to obtain a proof of the consistency of analysis.

Bill Howard generously shared his knowledge of Gödel's functional interpretation with us, and told about his encounters with Gödel, as reported in the introduction. We are very glad to dedicate this little volume to him.

Acknowledgements

The editing of Gödel's two notebooks that contain his Princeton Lectures on Intuitionism turned out to be a clearly more demanding enterprise than we first thought. There are lots of minute formal details, quite often incompletely and at places even erroneously given by Gödel himself. We thank Annika Kanckos and Tim Lethen, our research team members, for their support, as well as Fernando Ferreira and Bill Howard for sharing their expertise on Gödel's functional interpretation.

The preparation of this book has been financed by the European Research Council Advanced Grant Godeliana (grant agreement No 787758).

Contents

Introduction: Gödel's functional interpretation in context

In the spring of 1941, Kurt Gödel held a lecture course on intuitionistic logic at the Institute for Advanced Study in Princeton. Two spiral notebooks labelled simply "Vorl." and two sets of loose notes contain handwritten notes for the lecture course. The lecture notes divide into two themes. The first part is an introduction to intuitionistic logic. The second part is a detailed presentation of Gödel's functional interpretation of Heyting Arithmetic and its applications.

The general aim of the lectures is to examine the constructivity of intuitionistic logic. In the first part of the lectures, Gödel focuses heavily on the interconnection between intuitionistic and classical logic. The standard *proof explanation* of the intuitionistic logic was, he believed, not adequate to show the constructive character of intuitionistic logic. By reinterpreting intuitionistic logic in a more precise way, Gödel wants to prove that Heyting Arithmetic is properly constructive in the sense that it has the existence property. This reinterpretation is Gödel's functional system Σ, and the Princeton course is the most detailed presentation of it.

The theme of the lectures was closely connected to Gödel's previous talks of 1933 and 1938, as well as a lecture given at Yale University in April 1941. In the lecture "The present situation in the foundations of mathematics" given in Cambridge, Massachusetts, in 1933, Gödel argues that intuitionistic logic is not an ideal basis for a constructive foundation of mathematics because of the nature of its logical operations and the proof explanation. In his "Zilsel lecture" of 1938, he mentions an alternative interpretation of the logical operations in terms of a system of primitive recursive functionals of higher types. Finally, the system is developed in detail in the Princeton course and the Yale lecture. These results – apart from the Princeton lectures – were published posthumously in Gödel's *Collected Works* in 1995; the first published article on the functional interpretation appeared 17 years after the Princeton course, in the journal *Dialectica* in 1958.

In what follows, I will give an overview of the lecture course, highlighting the features which are missing from the other works of the 1930s and early 1940s. Apart from higher level of detail, the new aspects include an alternative version of Gödel's negative translation between Peano and Heyting Arithmetic (Gödel 1933b), the "truth table theorem" that proves that classical and intuitionistic propositional logics coincide under the assumption of decidability of atomic formulas, and a presentation of applications of the functional system Σ only mentioned in the Yale lecture. However, even where Gödel considers themes already mentioned in the other works, we often gain new insight into his views on particular issues. In this sense, the Princeton lectures complement the shorter lectures and give a richer picture of Gödel's early views on intuitionism.

Content of the lectures

If Gödel's lecture course had a specific title, it is not known to us: the IAS Bulletin of October 1941 tells only that "Dr. Gödel lectured on some results concerning intuitionistic logic," and that in the academic year 1941-1942, "he will continue his researches on this subject and its connection with the continuum problem." The course consisted of at least nine lectures, although the notes are not divided into sections. However, Gödel seems to have started each lecture with a review of the previous lecture's contents; there are, in total, nine of this kind of "last time..." summaries. At the Institute for Advanced Study, the Spring Semester lasted from 1st February to 1st May, and Gödel probably gave his course around this time. For the most part, the notes are clearly written and easy to understand, although toward the end more advanced themes are introduced. In a letter of 4th May 1941 to his brother, Gödel wrote that there were only three students left at the end of his course.[1] The wartime circumstances were probably one cause for the lack of attendance – and perhaps Gödel's rigorous yet terse presentation had scared away some of the listeners.

The lectures divide into two main parts. The first part, p. 1–47 of the lecture notes, introduces intuitionistic propositional and predicate logic and studies the interconnections between intuitionistic and classical logic. The second part, p. 48–117, concerns the functional interpretation of Heyting arithmetic. Moreover, Gödel's mathematical notebooks, the *Arbeitshefte*, contain early sketches of proofs featured in the lectures. The notebooks 7–10 (030025–030028)[2] probably date from early 1941; *Heft 7* (030025) is dated 1.1.1941 and in *Heft 9* (030027) we find the date "Feb 1941."[3] The earliest drafts of the functional system Σ in *Hefte* 7 and 9 are all titled "Gentzen" or "Gentzen Bew⟦eis⟧." This probably refers to Gentzen's first consistency proof of 1935 (Gentzen 1935/ 1974), which

[1] The letter is quoted in (Van Atten 2015, 201).

[2] The items in Gödel's *Papers* are referred to by their document code.

[3] Gödel was not in the habit of writing down dates of his notebook entries; he often only marked the change of the year.

he chose not to publish because of Gödel's and Paul Bernays' critique, involving a reduction procedure reminiscent of the "no-counterexample" interpretation of Gödel's Σ. The later proofs do not mention Gentzen.

It is beyond the scope of this introduction to consider Gödel's shorthand notebooks in depth. Unlike the lectures, the *Arbeitshefte* do not contain finished proofs ready for publication, there are many unfinished sketches, trial and error, and long computations.[4] However, in *Arbeitsheft 9*, p. 2–3, we find a numbered list written in shorthand and titled "Vorl. 1941 Sommer," which is clearly a plan for the Princeton lectures. The plan contains twelve points. Item number $2'$, a later addition on p. 2, summarizes Gödel's general agenda:

> On the basis of the intuitionistic axioms formulated by Heyting, criticism against them [especially the availability of negative universal statements.] *What is a properly intuitionistic* system [in particular, existential statements superfluous]. *Then also classical number theory derivable. This would perhaps be a reason against ⟦Heyting's logic⟧, but not correct, because the Brouwerian concepts are expressible in a system where no such unclarities occur. That is the goal of the lectures. It results also in a consistency proof for number theory.* First, however, the intuitionistic Heyting system and its properties.[5]

Although Gödel's goal is philosophically motivated, the lectures are mostly formal in nature. Nevertheless, each proof or formal explanation seems carefully planned to support the overarching goal of demonstrating the problems of intuitionistic logic and then giving an alternative interpretation in order to prove that intuitionistic logic (or, at least, arithmetic) is properly constructive. The lack of philosophical remarks is not surprising, as Gödel's early style was in gen-

[4] The other mathematical notebook series, *Resultate Grundlagen*, contains the finished proofs, but only two of them (the "constructive negation translation" discussed below and an inductive proof of computability of Σ-functionals; see (Hämeen-Anttila 2020, 98–102)) are directly related to the Princeton lectures.

[5] Aufgrund dieser intuit⟦uitionistischen⟧ Axiome formuliert [von Heyting] Kritik dagegen [insbesondere Vorhandensein der Negationen von Allaussagen]. *Was ist ein wirklich intuit⟦ionistisches⟧ ⟦System⟧?* [Insbesondere Existenzaussagen überflüssig]. Daher auch klassische Zahlentheorie ableitbar. *Das ⟦wäre⟧ vielleicht ein Grund dagegen, aber nicht richtig, denn die Brouwer'schen Begriffe ⟦sind⟧ ausdrückbar in einem System, in welchem keine solchen Unklarheiten vorkommen. Das ist der Zweck der Vorlesungen. Ergibt auch Widerspruchsfreiheitsbeweis für Zahlentheorie.* Zunächst aber intuit⟦ionistisches⟧ Heyt⟦ing'sches⟧ System und seine Eigenschaften.

eral very concise and rather formal.[6] It is only in the 1958 article in *Dialectica* where we find Gödel's – now more mature – philosophical views on constructivity fully laid out.

Gödel's full plan (*Arbeitsheft 9*, p. 2–3) includes the following themes:

1. Definition of the logical connectives.
2. Basic intuitionistic logic, non-constructive existential statements and their origin, namely the axioms $A \vee \sim A$ and $\sim\sim A \supset A$.
3. The exclusion of these principles in intuitionistic logic and the definition of negation in terms of absurdity. The axioms concerning negation can thus be left out.
4. The intuitionistic predicate calculus.
5. Derivability and non-derivability in intuitionistic calculus; in particular, the addition of either of the two principles $A \vee \sim A$ and $\sim\sim A \supset A$ gives classical logic.
6. "System **S**"[7] has the properties of an intuitionistic system.
7. The interpretation Σ as well as the construction of existential statements.
8. Proof of the soundness of the intuitionistic axioms with respect to system Σ.
9. Consistency of number theory:
   (a) Formalization of classical number theory;
   (b) Interpretation of the aforementioned system;
   (c) The negative translation for the system **S**.
10. Proof of consistency of $\neg(p)(p \vee \neg p)$.

[6] Kreisel (1987, 144) describes the early works as "concise and cavalier, apparently scoffing [...] at the antics of the rhetoric." The later works, quite the contrary, are more sensitive to philosophical issues in particular.

[7] System Σ seems to refer to the quantified system $\overline{\Sigma}$ of the Princeton lectures here. **S**, on the other hand, probably refers to the quantifier-free version denoted by Σ in the Princeton lectures. At one place Σ is written as a mirror image, resembling number 3.

11. Computability of all functions in **S**.

12. Proof that consistency is not provable in any smaller system.

For the most part, the lectures proceed according to Gödel's plan; however, items 11 and 12 are not covered in the lectures. Of particular interest is the issue of computability of higher-type functions, which Gödel still thought he could prove successfully at this point. I will discuss this below in the section on the system Σ.

The more detailed overview of the Princeton lectures is divided into four themes. I will start with Gödel's presentation of intuitionistic logic and its properties, especially in relation to classical logic. The second theme is Gödel's criticism of intuitionism and the sources of this criticism. The third part discusses Gödel's presentation of the functional system and the features not covered in the Yale lecture of the same year. Finally, I will consider the last theme of Gödel's lecture, namely the applications of the quantified functional system $\overline{\Sigma}$.

Sources

The lecture notes can be found in two spiral notebooks (040407, 040408) and a dozen loose pages (040409) filed together in Gödel's papers. Elsewhere (030077) we can find an envelope with "Beweis d. Gültigkeit d. int. Ax" written on it which contains the soundness proof for the functional interpretation.[8] The original transcripts were made from microfilm copies of the original notes, which were later controlled against the originals at the Princeton University Library.

The pages in the envelope have originally been numbered from 1 to 16. The page numbers have then been erased and replaced by new ones continuing the page numbering in the second spiral notebook. The envelope also contains a slip explaining how the loose pages should be ordered.

The lecture notes are mainly written in longhand English, with some shorthand additions in German. Gödel was used to writing his personal notes in Gabelsberger shorthand; e.g., the *Arbeitshefte* are almost entirely written in this script. We have transcribed and translated these additions, and where there might be a possibility of misunderstanding or a longer shorthand passage, added the German transcription as well.

Because the Gabelsberger system is language-specific and Gödel was lecturing in English, he had to write, for the most part, in longhand. However, even

[8] As far as I know, these missing pages were first discovered by Van Atten (2015).

his longhand writing retains many characteristics common to shorthand writing. These include the frequent use of abbreviations and the lack of punctuation or capital letters, and occasionally, a shorthand German word can be found in the middle of an English sentence. E.g., a passage on p. 66 of Gödel's notes reads:

> to be more exact if T_i should contain some var diff from $x_1 \ldots x_n$ we form first terms T_i' by repl the *überflüssige* var by arb. const. and then these are correct Df. with T_i' inst of T_i For $n = 0$ we obtain the following special case $A(\overline{u_1}\_\_\overline{u_n}\,\underline{y_1} \ldots \underline{y_r})$ is dem in $\overline{\Sigma}$ if and only if there are const $\alpha_1 \ldots \alpha_n$ such that $A(\alpha_1 \ldots \alpha_n\, y_1 \ldots y_r)$ is dem in Σ

For someone accustomed to stenographic writing, the slow pace of longhand writing is surely frustrating, and this is probably one reason for Gödel's frequent use of abbreviations. To maintain readability, we have not indicated where an abbreviation has been completed or a comma or a full stop added. Only in cases where the interpretation is not completely straightforward have we indicated the completion of a word. For the most part, however, we felt that Gödel's (occasionally non-idiomatic) style of writing should be respected, and have avoided editing the text beyond those small completions and corrections, even where Gödel's grammar or choice of words could seem somewhat awkward.

Gödel's formal notation is not entirely uniform, and in this case, we have chosen to edit it more heavily. E.g., Gödel uses both brackets and dots to indicate order in formulas, so the formula $(A \rightarrow B) \rightarrow C$ might sometimes be written $A \rightarrow B\ .\rightarrow .\ C$. We have chosen to use the former notation which is easier to read. Gödel uses both $\cdot$ and . for conjunction, and sometimes he leaves the conjunction out altogether, so that $A \cdot B$ becomes AB. Here, too, we have opted for the symbol $\cdot$ which occurs most often in the original text. Gödel employs, as Heyting did in his 1930s works, two different sets of connectives for intuitionistic and classical logic: $\{\neg, \&, \mathrm{v}, \rightarrow, \leftrightarrows\}$ and $\{\sim, \cdot, \vee, \supset, \equiv\}$, respectively. (The quantifiers have no special symbols in intuitionistic logic.) These we have, of course, left untouched.

Gödel denotes arbitrary formulas by upper case $A, B, C \ldots$ and occasionally with P, Q; however, he sometimes uses what is known as Sütterlin-Schrift instead of Latin letters. For formulas, where Gödel alternates between the two notations, we have chosen to use latin letters. However, Gödel consistently denotes sequences of variables by Sütterlin letters $\mathfrak{x}, \mathfrak{y}, \mathfrak{z}, \ldots$ and individual variables by lowercase Latin $x, y, z \ldots$. A printer would have typeset the Sütterlin letters in Fraktur, and this is the convention we have adopted in this case.

As mentioned, Gödel did not divide the notes into sections. The start of a new lecture is indicated only by Gödel's "last time ..." summaries. These have been indicated in bold.

The intuitionistic viewpoint

Gödel starts with the question, "what is constructive reasoning in mathematics?" He first shows some examples of *non*-constructive reasoning, which is here defined as those ways of inference of classical mathematics which allow for non-constructive existence proofs, i.e., proofs of existential statements $(\exists x)\varphi(x)$ without a corresponding instance $\varphi(a)$. The task, then, is to formalize mathematics in a way that avoids these undesirable consequences. This means that we need to avoid the two principles known to lead to such non-constructive existence statements, namely the Principle of Excluded Middle $A \vee \sim A$ and the Double Negation Elimination $\sim\sim A \supset A$. Of course, there might be other axioms or rules that have the same effect, so we need to be careful in choosing the right axioms.

The principle by which the intuitionists have chosen their axioms, Gödel remarks, is that they are taken as primitive and based simply on evidence (p. 7). Gödel makes it clear that there is room for improvement, and indeed, giving a formal as opposed to an intuitive interpretation of the logical operations is his main objective in the second part of the lectures. For now, however, he simply introduces what is today known as the proof explanation or the BHK (Brouwer-Heyting-Kolmogorov) interpretation of the intuitionistic operators.

He then presents the rules of intuitionistic propositional logic, which he attributes to two sources: Gerhard Gentzen's "Untersuchungen über das logische Schliessen" (Gentzen 1934-35) and Arend Heyting's "Die formalen Regeln der intuitionistischen Logik" and "Die formalen Regeln der intuitionistischen Mathematik" (Heyting 1930a,b). Although Gödel's view of deduction was, as opposed to Gentzen's, axiomatic in nature, his axioms and rules resemble more closely Gentzen's simple system than Heyting's 1930 formalism, which has eleven axioms but rules only for Modus Ponens, propositional substitution, and conjunction introduction. The same holds for Gödel's formulation of intuitionistic predicate logic.

The interrelation between classical and intuitionistic logic is of particular interest to Gödel. His own negative translation of 1932 (Gödel 1933b) showed that Heyting Arithmetic is equiconsistent with Peano Arithmetic, settling the question of whether intuitionistic methods surpass the finitistic ones. On p. 24–27 and 40–45, Gödel presents two further results on the connection between intuitionistic and classical systems. The first shows that the conditions for classical truth tables for propositional logic can be modelled in intuitionistic propositional logic. The second is a variant of the negative translation for predicate logic, using, however, "a more constructive notion of negation" than intuitionistic absurdity.

Between intuitionistic and classical logic

The truth conditions for each classical formula can be expressed in what is called a truth table. Given any valuation of its constituents, the rules for evaluating a compound formula can be expressed in a table such as this one for $A \supset B$:

A	B	$A \supset B$
T	T	T
T	F	F
F	T	T
F	F	T

Any connective of classical propositional logic can be given a meaning in terms of its truth table. For intuitionistic logic, however, this is not possible.[9] From the intuitionistic point of view, one can say that classical logic is the logic of finitary domains. For a decidable sentence, the Principle of Excluded Middle is validated, as truth and provability will then coincide. Gödel's truth table theorem (p. 24–27) proves this for propositional logic.

Assume we are given a truth table for an arbitrary expression A the atomic components of which are $p_1, p_2, \ldots, p_n$. Its truth table then has 2^n rows to cover all possible valuations. Let p_i^j denote the proposition formed as follows: $p_i^j = p_i$ if p_i is T on the jth row of the table, and otherwise $p_i^j = \neg p_i$.

Denote by *primitive conjunction* of A a conjunction $C_k = \bigwedge_{i=1}^{n} p_i^k$ for some row R_k in the truth table of A. In other words, this is a conjunction that expresses one valuation in a truth table. What Gödel wishes to show is:

[9] Another result of Gödel's (Gödel 1932) shows that intuitionistic logic cannot be represented in any finite-valued truth semantics.

Theorem. *Whenever A is a classical tautology, there is an intuitionistic derivation of $C_j \to A$ for each primitive conjunction C_j, $1 \leq j \leq n$, obtainable from the truth table of A.*

The result follows from the fact that for any A and $1 \leq j \leq n$, $C_j \to A$ or $C_j \to \neg A$, depending on whether the valuation on row j makes A true or false, is provable in intuitionistic logic. This can easily be proven by induction on the structure of A. A tautology is never false, and therefore we always have $C_j \to A$.

Nevertheless, a further principle is needed to prove A. Notice first that since any row in the truth table of A verifies A, it does not matter whether or not a given atomic constituent of A is negated or not in one of its primitive conjunctions C_k. Therefore intuitionistic logic validates both $p_1 \& C'_k \to A$ and $\neg p_1 \& C'_k \to A$, where $C'_k = \bigwedge_{i=2}^{n} p_i^k$. If we were able to eliminate p_1 somehow to obtain $C'_k \to A$, we could arrive to A by iteration.

However, the principle needed to conclude $C'_k \to A$ is the inference

$$\frac{\begin{array}{l} A \to B \\ \neg A \to B \end{array}}{B}$$

which is equivalent to $(A \vee \neg A) \to B$. That $C_k \to A$ intuitionistically implies A is then equivalent to the derivability of

$$((p_1 \vee \neg p_1) \& (p_2 \vee \neg p_2) \& \ldots \& (p_n \vee \neg p_n)) \to A$$

that is to say, the decidability of every atomic proposition. This is indeed assumed by classical logic, but for intuitionistic logic, it holds generally only in finite domains. Classical logic appears now as the special case of intuitionistic logic where the basic predicates and relations are decidable. A prime example is equality between natural numbers.

The second small theorem that Gödel proves is a form of a negative translation. The possibility of a translation of classical logic into intuitionistic logic relies on the fact that the fragment of classical logic which contains only $\{\supset, \cdot, (\,)\}$ is identical with the corresponding fragment $\{\to, \&, (\,)\}$ of intuitionistic logic. Gödel's 1932 translation, discovered independently by Gentzen, interprets the connectives $\vee$ and $(\exists)$ in terms of the classical equivalences

$$(A \vee B) \equiv \neg(\neg A \cdot \neg B) \text{ and } (\exists x)A \equiv \neg(x)\neg A$$

Then if, for a certain theory, the translations of the axioms of the classical theory hold in the intuitionistic theory, the translation gives a mapping from classical to intuitionistic logic so that the translation of each classical theorem is validated in the intuitionistic theory. In particular, this holds for the classical and intuitionistic theories of arithmetic.

The proof for the equivalence of the existence- and disjunction-free fragments is particularly impressive, says Gödel, for it assumes only positive logic, the perfectly unobjectionable rules of modus ponens, syllogism, the axioms of export and import, and the rules for the universal quantifier. This is the same remark that Gödel makes in the Zilsel lecture, except that there he believes that an additional axiom $p \rightarrow q \equiv \neg p \vee q$ for atomic p, q is needed. However, Sieg and Parsons comment, the use of this axiom in the proof is not necessary, and minimal logic suffices (Sieg and Parsons 1995, 73). Gödel may have made the same discovery, although here he does not explain the base case for atomic sentences.

The translation is more complex for predicate logic than for propositional logic. As discovered by Kolmogorov, one can obtain the translation for the latter system simply by prefixing every classical formula with two negations. This is not possible for predicate logic. However, a variant of this theorem is provable: for a "more constructive kind of negation than absurdity," it holds that if A is a theorem of an appropriate classical theory, the (ordinary) negation of the constructive negated statement is provable.

We first define the constructive negation translation of A as follows:

$$A^{con} := \begin{cases} A & \text{for atomic } A \\ \neg A_{con} & \text{otherwise} \end{cases}$$

$$A_{con} := \neg A \text{ for atomic } A$$

$$(B \& C)_{con} := B_{con} \vee C_{con}$$

$$(B \vee C)_{con} := B_{con} \& C_{con}$$

$$(B \rightarrow C)_{con} := (\neg B)_{con} \& C_{con}$$

$$((x)B)_{con} := (\exists x)A_{con}$$

$$((\exists x)B)_{con} := (x)A_{con}$$

$$(\neg B)_{con} := \begin{cases} \neg\neg B & \text{for atomic } B \\ (B_{con})_{con} & \text{otherwise} \end{cases}$$

We then have:

Theorem. *If A is derivable in a classical first-order theory, then* A^{con} *is derivable in its corresponding intuitionistic theory.*

The theorem holds, of course, under the same assumptions on the theory in question as the original negative translation.

The constructive negation translation is different from the standard translations of Gödel-Gentzen, Kuroda (1951), and Krivine (1990). It is, nevertheless, equivalent in the sense that it maps classical logic into the negative fragment of intuitionistic logic. It does have the special property that apart from the first negation which binds the whole formula, negations occur only in front of atomic propositions or negated atoms. This results in a rather complicated rule for evaluating negated expressions, as the negation must first be translated into constructive form which will then be transformed again by a second application of the constructive negation.

Gödel states on p. 41 that this A^{con} is "in a sense the most constructive statement equivalent to A." This is not entirely correct, as by Gödel's translation positive atomic propositions translate into double-negated ones, and a more constructive interpretation would be to drop the double negation. However, it is possible to show that for a compound formula B, $(B_{con})_{con} = B_{con^2}$ is equivalent to $B[p_i/\neg\neg p_i]$ for any non-negated atomic component p_i of B.[10] From this, it follows that for n greater than 2, B_{con^n} reduces to $B_{con^{n-2}}$.

The result is a translation which is not optimal in the sense that it need not contain the least possible amount of negations (see Ferreira and Oliva 2010), but one can say that it has a low negation complexity (in the sense that except for the first negation binding the whole formula, negations bind only atomic propositions).

For Gödel, the interconnections between intuitionistic and classical calculus are not only of formal interest, but they reveal a foundational relationship between the classical and intuitionistic theories which are intertranslatable in

[10] For the case of implication $C \to D$, one needs to treat it as $\neg C \vee D$.

this sense. He finishes the constructive negation translation with the following remark on p. 45.1:

> The results obtained have been pretty much surprising in so far as they show that in a sense the whole classical logic is contained in the intuitionistic logic. Of course it is contained only formally i.e. the same formulas can be proved but the meaning of these formulas is completely different (e.g. $\neg(x)\varphi(x)$ ⟦and⟧ $\sim(x)\varphi(x)$). But this difference of meaning makes the result still more surprising since this means that the non-constructive classical logic has a constructive interpretation. And this makes one doubtful whether intuitionistic logic really is constructive or if not perhaps some non-constructive elements are hidden in the axioms, which is quite possible regarding the great complicatedness in the primitive terms.

It was this worry that drove Gödel into redefining intuitionistic connectives in a way that he believed was more secure. Before examining his solution, however, I will briefly discuss his general criticism of intuitionistic logic and the proof interpretation.

Vagueness and absurdity: Gödel's critique of intuitionism

The passage on p. 45.1 suggests that there are two grounds on which, according to Gödel, intuitionistic logic is suspicious. The first he already mentioned at the beginning of the lectures: the intuitionistic connectives, understood through the concept of a proof, lack clarity and well-definedness. The second is the intimate connection of classical and intuitionistic predicate logic. These two objections are already raised in the Cambridge lecture of 1933, and repeated in the lectures of 1938 and 1941.

The proof explanation is the source of vagueness in intuitionistic logic. Compared to the relatively sharp critique of 1933, 1938, and the Yale lecture of 1941, Gödel stays quite neutral, saying that the notion of a proof is "perhaps not so absolutely clear." The main idea is that a proof in the intuitionistic sense is, as Gödel noted already in 1933 (Gödel 1933a, 53), understood as absolute, but such a notion of proof is no longer enumerable.

Gödel's main source for the proof explanation was Heyting. In the Princeton lectures, Gödel refers to Heyting's 1930 works (Heyting 1930a,b). Here the proof explanation is not yet mentioned. We also know that Gödel heard Heyting's Königsberg lecture in 1930 (Heyting 1931). Finally, he had seen an early version of Heyting's 1934 book (Heyting 1934), which grew out of a joint project of his and Heyting's. Although the 1930 papers do not discuss the proof explanation,[11] the meaning of the connectives in terms of provability is mentioned both in the Königsberg lecture and the book. However, what exactly counts as an intuitionistic proof is never defined exhaustively by Heyting or Brouwer; not because of sloppiness, but because from the Brouwerian viewpoint, there is no exhaustive definition. Intuitionistic mathematics is, in principle, incompletable, and intuitionistic logic as a description of intuitionistic mathematics shares the same property.

Gödel's second line of thought is in fact already stated in the Menger Colloquium talk, delivered in June of 1932. In the last paragraph of the paper, Gödel remarks that

> Theorem 1 ⟦of intertranslatability of HA and PA⟧ [...] shows that the system of intuitionistic arithmetic and number theory is only apparently narrower than the classical one, and in truth contains it, albeit with a somewhat deviant interpretation. The reason for this is to be found in the fact that the intuitionistic prohibition against restating negative universal propositions as purely existential propositions ceases to have any effect because the predicate of absurdity can be applied to universal propositions, and this leads to propositions that formally are exactly the same as those asserted in classical mathematics. Intuitionism appears to introduce genuine restrictions only for analysis and set theory; these restrictions, however, are due to the rejection, not of the principle of excluded middle, but of notions introduced by impredicative definitions [...] (Gödel 1933b, 295)

These are, in essence, the same remarks that Gödel makes on the three loose pages 39.1, 39.2, 39.3, probably ripped off from the first notebook and stacked between pages 63^{iv} and 64 of the second notebook. It seems that the three pages where Gödel discusses the relationship between intuitionistic and classical theories were originally meant to continue the discussion on p. 38. Apparently,

[11] Heyting later expressed his dissatisfaction with his early works, as they "diverted the attention from the underlying ideas to the formal system itself" (Heyting 1978, 15).

Gödel then changed his mind and decided to introduce the constructive negation translation instead.

A point that Gödel makes several times in his works of the 1930s and early 1940s is that intuitionistic logic allows to prove negated universal statements $\neg(x)\varphi$ without exhibiting an instance $\neg\varphi[a/x]$. This is because a negated sentence is seen as a hypothetical statement, and this is something that both Brouwer and Heyting saw clearly. However, not everyone agreed with this interpretation of intuitionistic negation. In 1926, Hermann Weyl stated the law of the excluded middle as follows: either all numbers have a property P or else there is a number which has the property $\neg P$, i.e., $(x)P(x) \vee (\exists x)\neg P(x)$ (Weyl 1926, 42). From this then follows $\neg(x)P(x) \rightarrow (\exists x)\neg P(x)$. A negation of a universal statement must then be forbidden, for that would be an existential statement. Weyl expressed the opinion that quantified statements are not proper judgments but rather "judgment-instructions" (*Urteilsanweisungen*, in the case of the universal quantifier) or "judgment-abstracts" (*Urteilsabstrakte*, existential quantifier) (Weyl 1921, 71). In this way, his view was in fact much stricter than that of Brouwer and in fact very close to Hilbert's finitism, where the acceptance of excluded middle leads to the rejection of quantifiers.

Gödel stated in 1975 that he first read Brouwer as late as in 1940 (Gödel 2003a). He wrote to his brother in 21st September 1941 asking Rudolf Gödel to obtain a copy of Brouwer's dissertation for him.[12] He probably attended one of Brouwer's two lectures in Vienna in 1928 (Wang 1989, xx) and seems to have known of Brouwer's earlier articles.

However, it is more likely that Gödel's view of intuitionism and its logic came from Weyl, whose works he had read in the 1930s,[13] as well as Hilbert and Bernays. Indeed, Hilbert, too, stated in his influential "Über das Unendliche" – which Gödel knew – that universal statements are not capable of being negated (*nicht negationsfähig*) (Hilbert 1926, 173). This has the consequence, Hilbert says, that it does not hold that every equation must be either satisfied for all numbers or have a numerical counterexample. This he interprets, as Weyl does, as a case of the Principle of Excluded Middle.

[12] Quoted in (Van Atten 2015, 189–190).

[13] We found bibliographic notes mentioning (Weyl 1926), which was quoted above, in a notebook titled *Altes Excerpten Heft I* (1931–) (030079), most of which can be found in (von Plato 2021). The same notebook also contains some brief notes related to Weyl's "Die heutige Erkenntnislage in der Mathematik" (Weyl 1925), which Gödel borrowed from the University of Vienna library in November 1932. Since both Gödel and Weyl were affiliated with the IAS at Princeton, they also knew each other personally.

Apparently, Gödel saw the negative translation between classical and intuitionistic arithmetic as a source of doubt for the constructivity of intuitionistic quantifiers. Otherwise, it would have made little sense to him to mention it in each of the 1932, 1933, 1938, and 1941 (Yale) papers in connection with the problem of negated universals. In general, it seems as though his views on intuitionistic logic had not changed much between 1932 and 1941.

It goes without saying that this interpretation of the intuitionistic negation and the quantifiers is, at the very least, rather unfair to Brouwer's original conception of intuitionism. Already in his dissertation in 1907, Brouwer made a clear distinction between an existential statement as a construction and a statement expressing that a construction is blocked or incompatible with some other proven fact. One could say, following Bernays (1935) as well as the first volume of *Grundlagen der Mathematik* (see Hilbert and Bernays 1934, 43) – both of which Gödel had read carefully – that the point at which intuitionism surpasses finitism is that it allows for ideal elements to appear in presuppositions. Thus, even if $(x)A$ were a non-finitary statement, one could presuppose it in a proof of an implication. The permissibility of ideal presuppositions is the additional abstract element – which Gödel does acknowledge in 1958 – that intuitionistic logic has and finitary systems have not.

At this point, Gödel was still more focused on the logical than the philosophical issues, and this is perhaps a reason why he took his negative translation to be so important, even though he admitted it to be purely formal. In any case, these considerations led Gödel to develop his own, formal and system-specific, interpretation of intuitionistic logic in terms of functionals of higher types. Whereas Gödel mentions the idea already in 1938, the Princeton course is the earliest source where Gödel considers the system in full detail.

The constructive system Σ and the calculability question

Despite his criticism, Gödel's goal is not to prove that intuitionistic mathematics is in general defective. Rather, it is the proof explanation of the logical connectives that is ill-defined. With a more precise interpretation of the logical connectives, intuitionistic logic can be shown to be properly constructive at least in the case of specific formal systems.

Gödel now gives precise criteria for a "strictly" constructive system. There are three requirements:[14]

[14] An interesting detail, compared to the Yale lecture as well as the Dialectica paper of 1958,

1. The propositional connectives bind only quantifier-free expressions.

2. The theory contains no existential quantifiers.[15]

3. Primitive relations are decidable and primitive functions are calculable.

This is a slight reformulation of the conditions given in the lectures of 1933 and 1938 and the Yale lecture of 1941. Condition 1 is replaced in the Cambridge and Yale lectures with the less restrictive prohibition of negated universal statements. Given that Gödel's primary worries concern the use of the implication, and more specifically, absurdity, the broadening of the conditions makes little difference in practice. Condition 3 is repeated in all of the lectures. Obviously, intuitionistic arithmetic does not satisfy conditions 1 and 2.

What is missing, compared to the 1933 and 1938 lectures, is the condition (expressed slightly differently in the two papers) that the basic objects of the theory should be somehow graspable or finitely generated. In 1938 he demands that our basic objects be "surveyable (*überblickbar*), that is, denumerable" (Gödel 1938, 91). This condition disqualifies, in particular, *the proof explanation* of intuitionistic logic as constructive in a way that cannot necessarily be overcome.[16]

Gödel presents a system Σ as a system that satisfies the three conditions of strict constructivity. The system Σ is essentially equivalent to Primitive Recursive Arithmetic with the addition of primitive recursive functionals of higher types. The types form a hierarchy of levels depending on the level of the highest-level argument or the value. The level of an integer (in Gödel's notation, type I) is 0. The level of a functional of type $t_1 \tau t_2$ (or $t_2 \mapsto t_1$) equals $\max(lev(t_1), lev(t_2)) + 1$.[17]

The atomic expressions of Σ are of the form $a = b$, where both a and b have the type I. For compound expressions, we can use the classical connectives

is that in this series of lectures, Gödel never uses the notion "finitistic" in connection with this properly constructive system. In the other works, he does mention that the "lowest level", i.e., his constructive system Σ restricted to functionals of type 1 and lower, is probably what Hilbert thought of as finitist mathematics.

[15] Except perhaps as a defined notion: $\exists x A := A[\alpha/x]$ for some constant expression α.

[16] However, Gödel's remark in 1938 that this condition is problematic "because of the concept of function" (1938, 91) brings Gödel's own interpretation by higher-type functionals into question (Sieg and Parsons 1995, 70). In any case, this is not an issue that Gödel discusses here or in the Yale lecture.

[17] Conventionally, the level of a functional of type $t_2 \mapsto t_1$ would be equal to $\max(lev(t_2)+1, lev(t_1))$.

$\{\sim, \cdot, \vee, \supset\}$, as there is no danger of applying propositional connectives to quantified statements. Higher-level equality is a defined notion and denoted by $\doteq$. For A, B of type other than I, $A \doteq B$ iff for any complete argument series[18] $\mathfrak{x}$, $A(\mathfrak{x}) = B(\mathfrak{x})$.

Gödel's axioms consist of the axioms of classical propositional logic, which can now be applied without restriction as Σ is quantifier free, the usual axioms for successor as well as for explicit definition and definition by primitive recursion extended to functionals of all finite types. As rules, we have Modus Ponens, the rule of substitution, and complete induction. Moreover, there is a rule of (weak) extensionality, formulated as follows (note that P is automatically quantifier free):

$$\frac{P \supset S(x) \doteq T(x)}{P \supset \varphi(S) \equiv \varphi(T)}\ \text{Ext}$$

Whereas Gödel does not mention identity and its treatment in the Yale lecture, it seems that at this point, he did not (seriously; see below) consider the intensional definition of identity adopted in the 1958 *Dialectica* article. In the Princeton lectures, he was particularly concerned with the question of computability of higher-type functionals, something that he only mentions in passing in the Yale lecture.

Σ satisfies the first two requirements simply because it is quantifier free. The primitive relation $=$ is obviously decidable. However, the status of higher-type functionals is not quite clear. In the Yale lecture, Gödel mentions the question of calculability but says that a proof is "pretty complicated," and will not be discussed in the lecture (Gödel 1941, 195). The truth is that Gödel did not have a satisfactory proof at hand. The Yale lecture took place on 15th April 1941, but given that Gödel's Princeton course had started already in February, his discussion of the calculability issue here probably preceded the Yale lecture. His sketch of a list of contents in *Arbeitsheft 9* suggests that he thought he could obtain a proof, and still in the middle of the lectures (p. 63^{iv}) he refers to a proof that will be given later. The fact that Gödel never returned to that proof seems to indicate that he simply did not get it ready in time.

However, Gödel does give two alternatives for a proof, which are also mentioned by Anne Troelstra in his introduction to the Yale lecture (Troelstra 1995).

[18] Gödel defines a complete argument series for an expression A as a sequence of terms of appropriate types $B_1, \ldots, B_n$ such that $A(B_1)(B_2)\ldots(B_n)$ reduces to type I.

The first option is to show, by simple induction, that for any complete argument series $\mathfrak{a}$ for a term f, there is an integer k such that $f(\mathfrak{a}) = k$ is provable. For functions of level L_1 this is clear. Because $f(\mathfrak{a})$ of level L_k has an argument type less than L_k, we can apply the inductive hypothesis to obtain the result. As Troelstra mentions, this resembles the approach in (Tait 1967). However, it is not good enough for Gödel, as it requires the use of the full intuitionistic logic, in having to apply complete induction to a *quantified* statement of the form $(\ )(\exists)$.

The first remarks on the inductive computability proof are followed by a cancelled passage (p. 61–62) which reads:

> So our attitude must be this that the axioms of Σ (in part⟦icular⟧ the schemes of definition) must be admitted as constructive without proof and it is shown that the axioms of intuitionistic logics can be deduced from them with suitable definitions. This so it seems to me is a program

This is the course that Gödel took in the *Dialectica* paper, allowing, in a sense, for the same kind of vagueness that appears in the justification of intuitionistic axioms on the basis of "intuitive evidence." Perhaps for this very reason, Gödel rejected this alternative in 1941.

In the beginning of the next lecture, Gödel returns to the calculability question with a new suggestion, remarking that there is another proof which does not rely on HA. This strategy utilizes transfinite induction up to ε_0. The idea is to show that one can define a sequence of substitutions to a given functional f that reduces to an integer term after a bounded number of steps. This would involve two steps; associating an ordinal $< \varepsilon_0$ with each term and showing that the ordinal is diminished by every replacement into the term (the "complete argument series"), and then appealing to the well-foundedness of ordinals below ε_0.

This method, employed before in Gentzen's and Wilhelm Ackermann's consistency proofs for Peano Arithmetic (Gentzen 1936; Ackermann 1940), presupposes that the use of transfinite induction can be "justified" in some sense; however, as Gödel notes (p. 63^{iv}), it does not look simpler than the system Σ itself in any obvious way. In the Zilsel lecture, Gödel states that Gentzen's method of transfinite induction, even if not strictly constructive, has "a high degree of intuitiveness" (Gödel 1938, 107). The main problem for Gödel in 1938 is that the property "α is an ordinal" is impredicative; this is not specific to Gentzen's proof but to transfinite induction in general. Therefore, it is not clear how he would justify the use of the principle in his own proof.

In his notebooks from the early 1940s, we can see that Gödel was very interested in Gentzen's method of transfinite induction. However, it seems that he never came up with a proof that would satisfy him.[19] In 1958, Gödel no longer mentions the proof of calculability, instead returning to his previous idea of assuming computability without a formal proof. Later, William Howard gave a reductive proof using transfinite ordinals (Howard 1970), but at the time, Gödel did not seem to like it.[20]

Interpretation of intuitionistic arithmetic in system $\overline{\Sigma}$

The system $\overline{\Sigma}$, which extends the quantifier-free system Σ with existential and universal quantifiers, functions like what is called simply the system Σ in the Yale lecture of April 1941. The notation, however, is rather unique. Instead of quantifiers, two new types of *variables* are introduced: universal variables $\underline{x}$ and existential variables $\overline{x}$. Thus, e.g., a statement conventionally written as

$$(\exists y)(x)A(x, y)$$

expressed in the $\overline{\Sigma}$ notation as

$$A(\underline{x}, \overline{y})$$

A matrix $M[\xi_1, \ldots \xi_n]$ is an expression in which the terms that occur are not specified as to their type or variable type (free, universal, or existential).

[19] The *Arbeitshefte*, in particular, contain dozens of pages of notes on transfinite ordinals and functionals of finite types. These notes have so far not been investigated in detail but at a first glance, it seems that, as one would expect, they do not contain a finished proof.

[20] In an email message to Jan von Plato on 8th June 2007, Howard refers to a conversation between him and Gödel in 1972:

> GÖDEL: "You must work out an assignment of ordinals to terms for primitive recursive functions of finite type, such that a calculation step lowers the assigned ordinal."
> BH: "But I did this in my paper."
> GÖDEL: "It is not satisfactory; it is too complicated; one cannot see why it works."

It appears that still in the 1970s, over thirty years later, Gödel had not let go of the idea of finding a proof.

We represent a matrix completed with arguments as $M[t_1, \dots t_n]$. This corresponds, in the standard notation, to a quantified expression

$$(\exists y_1) \dots (\exists y_k)(x_1) \dots (x_l) M[t'_1, t'_2, \dots t'_n]$$

where all overlined or underlined variables in $t_1, t_2 \dots t_n$ have been turned into normal ones, $y_1 \dots y_k$ are all the overlined variables and $x_1 \dots x_l$ all the underlined variables in the original expressions.

The motivation for using this notation over the standard one is not explicitly stated in the lectures. In the early presentations of the functional interpretation (1938 and 1941), as well as in the *Dialectica* article, Gödel uses quantifier notation; many of his preparatory notes for the Princeton course are also written in the ordinary notation. In fact, the only place we seem to find this style of notation in Gödel's notes are in parts of *Arbeitsheft* 9 and in *Arbeitsheft* 10. One advantage is brevity, especially in the longer proofs which we find on p. 89–106. Furthermore, Gödel may have wanted to emphasize the difference between intuitionistic quantifiers and those of his functional system. Indeed, on p. 81 he suggests using Σ, Π to talk about existential and universal quantifiers in the quantified system $\overline{\Sigma}$.

The goal, then, is to define a translation A' for each intuitionistic formula A, such that it maintains the validity of the (translations of the) intuitionistic axioms in system $\overline{\Sigma}$. The soundness of the interpretation gives a relative consistency proof for HA with respect to $\overline{\Sigma}$.

Save for notation, the interpretation of intuitionistic logic in the system $\overline{\Sigma}$ does not differ from that given in the Yale lecture or the *Dialectica* article of 1958.[21] For atomic sentences, the $\overline{\Sigma}$-interpretation is the formula itself. Take $A' = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$ and $B' = N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$, where $\mathfrak{a}$, $\mathfrak{b}$ are sequences of free variables, and where all the variables of A and B are assumed to be mutually disjoint. The interpretation is defined as follows:

1. $(A \& B)' := M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \cdot N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$
2. $(A \text{ v } B)' := (M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \cdot \overline{u} = 0) \vee (N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}] \cdot \overline{u} = 1)$
3. $(A \rightarrow B)' := M[\mathfrak{a}, \underline{\mathfrak{f}}, \bar{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{y}})] \supset N[\mathfrak{b}, \bar{\mathfrak{p}}(\underline{\mathfrak{f}}), \underline{\mathfrak{y}}]$

[21] Except for the erroneous disjunction clause, which gives the translation of AvB as $M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \vee N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$ in the Yale lecture. Gödel originally makes the same mistake of defining the translation of disjunction similarly to that of conjunction in the Princeton lectures, but corrects it soon after.

4. $(\exists x A)' := M[\overline{x}, \mathfrak{a}, \overline{\mathfrak{f}}, \underline{\mathfrak{x}}]$

5. $(\forall x A)' := M[\underline{x}, \mathfrak{a}, \overline{g}(\underline{x}), \underline{\mathfrak{x}}]$

In 2, u must not occur in A, B; similarly for $\mathfrak{q}, \mathfrak{p}$ in 3. In 4 and 5, x is a variable free in A. Negation $\neg A$ is here interpreted as $A \rightarrow \perp$

Pages 89 to 106 contain fairly detailed proofs of the derivability of the intuitionistic axioms and rules; Gödel probably did not go through every one of them during the lectures. As mentioned before, these pages were ripped off from a notebook and filed elsewhere, probably because the proofs were mostly routine and very formal in nature.

The main difference to the Yale lecture is the level of detail: in Yale, obviously, Gödel does not go through all of the proofs. Two points are worth mentioning in the context of the Princeton lectures and Gödel's aim of giving a proof of constructivity for intuitionistic logic. Whereas the desired proof is obtained by the interpretation (see next section), one can ask what actually is needed for this proof and whether those methods are acceptable from a constructive viewpoint. The first question, then, is whether the translation is intuitionistically acceptable and the second whether everything needed to *prove* the validity of the intuitionistic axioms in $\overline{\Sigma}$ is acceptable.

As for the first question, the equivalence of the intuitionistic formulas and their translations is not necessarily itself provable in Heyting Arithmetic extended to finite types, i.e., Σ_I (or as it is nowadays denoted, HA^ω). We do have

$$\vdash_{\Sigma_I} A \;\Rightarrow\; \vdash_{\Sigma_I} A'$$

but in general, not the other way around. Problems arise with the translation of ( ) and $\supset$. In the case of the universal quantifier, what is needed for extracting $M[\underline{x}, \mathfrak{a}, g(\underline{x}), \underline{\mathfrak{x}}]$ from $M[x, \mathfrak{a}, \overline{\mathfrak{f}}, \underline{\mathfrak{x}}]$ is the Axiom of Choice (for finite types). AC is usually accepted when x is of type I, but it is not necessarily intuitionistically acceptable in general.

The case for $\supset$ is more complicated. In the Princeton lectures, as well as in Yale, Gödel gives a fairly detailed informal explanation for the justification of the interpretation of implication. However, as Spector has shown (Spector 1962), the full *formal* demonstration of the equivalence of $A \supset B$ and its $\overline{\Sigma}$-translation requires two principles that go beyond HA: extended versions of Markov's Principle and Independence of Premise. The version of Markov's Principle that is needed is:

(MP′)

$$\neg(x)A(x) \supset (\exists x)\neg A(x)$$

where A is quantifier free and x of any finite type; and moreover, we need the Independence of Premise

(IP′)

$$(\forall x A \supset (\exists x)B(x)) \supset (\exists x)(\forall x A \supset B(x))$$

where x is of any finite type.

Neither of the principles is valid in intuitionistic logic. Even if not intuitionistically acceptable, MP can be interpreted constructively as unbounded search, where x is assumed to be a variable of type Int: if $A(n)$ is indeed decidable for any natural number, then one can run through all the natural numbers until one finds a counterexample a such that $A(a) \supset \perp$ is constructible (Avigad and Feferman 1998, 337). However, whether this generalizes to higher types depends on whether we consider $\doteq$ in general decidable, and as was noted previously, Gödel did not think that this should be accepted without proof. IP is likewise suspicious because of the nature of the intuitionistic implication. Intuitionistically, the antecedent of IP is read, "given a proof of A, one can construct a witness a such that $B(a)$," whereas the consequent has the dependence the other way around: "one can construct an a such that given a proof of A . . .".

In the case of the soundness proof, the crucial part is the proof of $A \to (A \& A)$. It has been noted (Troelstra 1990, 227) that the proof of this needs to assume the existence of characteristic functions for formulas. I.e., when one chooses the proper substitution in (see p. 90 of the lectures)

$$M[\mathfrak{a}, \mathfrak{f}, \varrho(\mathfrak{a})(\mathfrak{f}\,\mathfrak{x}\,\mathfrak{y})] \supset M[\mathfrak{a}, \sigma_1(\mathfrak{a})(\mathfrak{f}), \mathfrak{x}] \cdot M[\mathfrak{a}, \sigma_2(\mathfrak{a})(\mathfrak{f}), \mathfrak{y}]$$

for $\varrho(\mathfrak{a})(\mathfrak{f}\,\mathfrak{x}\,\mathfrak{y})$ one needs a characteristic function c_M for $M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}]$, where $c_M = 0$ and $\varrho(\mathfrak{a})(\mathfrak{f}\,\mathfrak{x}\,\mathfrak{y}) = \mathfrak{y}$ or $c_M \neq 0$ and $\varrho(\mathfrak{a})(\mathfrak{f}\,\mathfrak{x}\,\mathfrak{y}) = \mathfrak{x}$.

As Urquhart (2016, 508) notes, it was indeed the classical form of this axiom, $(A \vee A) \supset A$, which caused the most trouble for Russell when he proved the validity of the propositional axioms for predicate logic in the *Principia*. In terms of sequent calculus, it is the corresponding rule of contraction that complicates any proof of consistency for arithmetic. Contraction-free arithmetic has a proof-theoretic ordinal of only ω^ω (see Petersen 2003). The analogy seems to be that in the presence of contraction, one needs to consider the length of a derivation of a formula as a parameter in cut elimination exactly because we have to choose which of the originally contracted formulas the cut was applied to. Gödel writes to Bernays as late as in July 1970 that he does not understand how characteristic functions are needed for the axiom $A \to (A \& A)$ (Gödel 2003a, 282). However, he does go through the full proof for this axiom using characteristic functions in the Princeton Lectures.

Does any of this affect the credibility of Gödel's interpretation? One could answer that $\overline{\Sigma}$ has the properties of a constructive system, and it in itself validates AC, MP′ and IP′, whether or not the proof explanation – which Gödel considered in certain ways defective – can make sense of them. But as Troelstra notes in his introduction to the *Dialectica* article, this can also be seen as a reason to be suspicious of whether $\overline{\Sigma}$ is indeed more constructive than HA. He remarks that "Markov's schema is false for some perfectly coherent intuitionistic theories such as the theory of lawless sequences [...] while Gödel himself [...] regards choice sequences as coming close to being finitistic" (Troelstra 1990, 232). Moreover, there are equally constructive interpretations of intuitionistic logic which do not validate MP. One example is Kreisel's modified realizability interpretation (Kreisel 1959), based on Kleene's work on numerical realizability (Kleene 1945).

It is highly unlikely that, as opposed to the problem of computability in higher types, Gödel was aware of any of these issues in 1941, or even in 1958. We cannot, then, but speculate how he would have answered the challenges. What is certain is that none of this diminishes the formal and pragmatic advantages of the functional interpretation in extracting constructive content from apparently non-constructive proofs, which is in itself a very useful property. It is thus only from the philosophical point of view that these issues could be seen as obstacles; and even then, there is no obvious way of drawing the conclusion that $\overline{\Sigma}$ is *not* a more constructive way of interpreting intuitionistic logic than the proof explanation.

Applications of the $\overline{\Sigma}$-translation

In the Yale lecture, Gödel briefly mentions several applications of his interpretation (Gödel 1941, 199–120). First of all, it can be used to show that despite $\sim(A \vee \sim A)$ being unprovable for any A, $\sim(x)(\varphi(x) \vee \sim\varphi(x))$ is consistent in HA. Moreover, the functional translation gives a relative consistency proof for HA, and via the negative translation theorem, one for PA, as well. Gödel mentions that the functional interpretation also demonstrates the existence property for HA. These applications are investigated in detail in the last part of the Princeton lectures.

Gödel's main goal, stated in the beginning of the lectures, is to show that intuitionistic logic is indeed constructive; in particular, that every existential theorem of HA can be instantiated. This turns out to be not possible to show via Gödel's translation. However, he presents the following argument (p. 84):

From the way existential variables were introduced in $\overline{\Sigma}$ – essentially, we have only an introduction rule for the existential quantifier – it follows that for any existential statement $(\exists x)\varphi$, if $\vdash_{HA} (\exists x)\varphi$, then (since Σ validates $((\exists x)\varphi)'$) it follows that $\vdash_{\Sigma} \varphi'[\alpha/x]$, where α is a constant term. Now, one can extend Σ by quantifiers and the (intuitionistic) axioms for them to obtain the system Σ_I, of which the Heyting Arithmetic is a subsystem. As Σ is a subsystem of Σ_I, we get the same result in Σ_I.

Although Gödel states that this argument gives "the desired proof for constructivity of intuitionistic logic," it does not show the existence property for HA. Namely, the constant term α might be of some type higher than I and thus not translate into a HA term, and A' not necessarily into A. Gödel does not notice this until at the end of the lecture course (p. 116), where he corrects himself. The only case where the existence property does hold is when $\varphi = A(x)$ for a quantifier-free A, and in this case, the proof translates to similar property for PA, as well (p. 117): because $(\sim(x)\sim A(x))'$ and $((\exists x)A(x))'$ are Σ-equivalent, if $\sim(x)A(x)$ is demonstrable in PA (and thus in HA), then $\sim A(\alpha)$ is demonstrable in Σ_I and this also in Σ_K. Gödel does not mention the disjunction property – at this point, he still made the mistake (as in the Yale lecture) of interpreting the intuitionistic disjunction like the classical one. The mistake is corrected on p. 89.1. In any case, also the disjunction property follows in a similar manner.

The main focus of the last lecture, from page 107 to page 117, is to prove that $\overline{\Sigma}$ invalidates the law of the excluded middle. The result is obtained by first defining a model M for $\overline{\Sigma}$ (and Σ) and then constructing an arithmetical statement $\varphi(x)$ such that $\sim(x)(\varphi(x) \vee \sim\varphi(x))$ is true in M. The lecture starts with an introduction to general recursive functions. Strangely, Gödel refers several times to discussion on these things "last time". However, since the last topic was the $\overline{\Sigma}$ translation and its soundness (p. 73–106), this cannot be the intended reference. Oddly enough, there is no gap in the pagination. Perhaps Gödel did not make extensive notes for the introductory lecture on recursive functions; in any case, page (107′) appears to be a summary of this lecture.

To interpret the system $\overline{\Sigma}$, Gödel uses what is essentially a model of Hereditarily Effective Operations (see Troelstra 1973, section 2.4.11). The domain consists of objects of a given type, so that the objects of the lowest level are simply integers. Objects of type $\tau(t_1 \dots t_k)$ are represented by the codes n for partial recursive functions $f_n(x_1 \dots x_k)$ that have arguments of appropriate types and that are extensional for extensionally equal arguments.[22] Gödel's $\varphi(x)$ then codes the statement "the recursive function number x is undefined for some argument," i.e. "x codes a non-total recursive function." Now, if $\varphi(x)$ could be decided for any x, then one could solve the halting problem in HA. From this it follows that it cannot be the case that $(x)(\varphi(x) \vee \sim\varphi(x))$.

After the seventeen-year silence

Gödel chose not to publish any of his 1933–1941 lectures on intuitionism. A graduate student Frederick W. Sawyer wrote to Gödel in 1st February, 1974, asking about "Kreisel's remark to the effect that you had incorporated the Dialectica interpretation into your lectures at Princeton as early as 1941" (Gödel 2003b, 210). In an undated draft for a reply, Gödel writes that he had "several reasons why I did not publish it then. One was that my interest shifted to other problems, another was that there was not too much interest in Hilbert's Program at that time" (Gödel 2003b, 210).

The mention in the IAS Bulletin of October 1941 that Gödel is planning to study the connections of the functional system and the Continuum Hypothesis hints at a reason for why he lost his interest. In *Arbeitsheft* 7 (p. 33, reverse direction), we find a list of objectives Gödel wanted to achieve – these lists Gödel called "Programme" – which includes the following:

> 3. Extension of the consistency proof to the case where quantifiers occur in the recursive definition (and thereby ramified type theory) and analysis.
>
> 4. Consistency proof for $\mathfrak{S}$ by means of higher ordinals and determination of what ordinals are definable in $\mathfrak{S}_n$.
>
> 5. Extension of $\mathfrak{S}$ to transfinite types and determination of the ordinal numbers definable there and a proof of consistency.
>
> ⋮

[22] I.e., where x_i is extensionally equal to x_i', $f(x_1, \dots x_i, \dots x_k) = f(x_1, \dots x_i', \dots x_k)$.

> 7. Extension of functions definable in $\mathfrak{S}$ to Brouwerian types and calculation of the Brouwerian type for combinations of functions (also for functions whose existence is proven in a certain way) (together with the first reconstruction of Gentzen's proof) and consistency of analysis.[23]

In the eleventh *Arbeitsheft* (030029) we find plenty of formal sketches in the functional notation, with two longhand titles "Wid. freiheit v. $\neg(p)(p \vee \neg p)$" and "Wid. freiheit Analysis." What all of this suggests is that whereas Gödel was genuinely concerned with the question of constructivity of intuitionistic logic, his ultimate aim seemed to be to prove the consistency of continuum hypothesis and, eventually, of analysis.

However, none of these investigations seem to have led to concrete results. There are no signs that Gödel continued developing the functional system after 1942. Kreisel states that Gödel quit working with his interpretation "after he learnt of recursive realizability that Kleene found soon afterwards" (Kreisel 1987, 104). Perhaps Gödel himself became doubtful, both of the strength of his method and its foundational justification; and perhaps this was partly the reason why he turned away from mathematics and towards philosophy soon after the Princeton lectures.

The functional interpretation became known as late as in July 1957, when Georg Kreisel gave a talk on the unpublished results of Gödel at Cornell University. The title of the presentation was "Gödel's interpretation of Heyting's Arithmetic," but a great part of it concerned the extension of the interpretation to analysis (see Feferman 1998, 220–223). Kreisel lectured in Amsterdam soon after on the same topic. In the latter paper (Kreisel 1959), the translation of analysis

[23] 3. Erweiterung des Widerspruchsfreiheitsbeweises auf den Fall, dass in den rekursiven Df Quantoren auftreten (und dadurch verzweigte Typentheorie) und Analysis

4. Widerspruchsfreiheitsbeweis für $\mathfrak{S}$ mittels höherer Ordinalzahlen und Feststellung, welche Ordinalzahlen in $\mathfrak{S}_n$ definierbar

5. Erweiterung von $\mathfrak{S}$ auf transfinite Typen und Feststellung, welche Ordinalzahlen dort definierbar und Widerspruchsfreiheitsbeweis

⋮

7. Erweiterung der in $\mathfrak{S}$ definierbaren Funktionen in Brouwersche Typen und Berechnung des Brouwerschen Typus für Kombinationen von Funktionen (auch für Funktionen deren Existenz in gewisser Weise bewiesen) (und zusammen mit 1. Rekonstruktion der Gentzen Beweis) und Widerspruchsfreiheitsbeweis Analysis

is based on arbitrary continuous functionals and recursively continuous functionals. Kreisel's scheme differs from Gödel's in that whereas Gödel's interpretation validates Markov's Principle, Kreisel's interpretation, like the intuitionistic proof explanation, does not.

Gödel's first published work on the topic, titled "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes," appeared in 1958 in the journal *Dialectica*. The viewpoint of 1958 is somewhat different from the Princeton lectures. As the title suggests, the functional system is introduced in the context of the extended Hilbert Programme, giving a constructive consistency proof for arithmetic. Here Gödel has dropped the extensional interpretation of equality; computability is thus not proven but presupposed. Extension to stronger systems is only mentioned in the very last paragraph: "It is clear that, starting from the same basic idea, one can also construct systems that are much stronger than [the functional system] **T**, for example by admitting transfinite types or the sort of inference that Brouwer used in proving 'the fan theorem'." (Gödel 1958, 251).

Clifford Spector was the one to extend the *Dialectica* interpretation by bar recursion, a definitional schema corresponding to the principle of bar induction,[24] arriving at the proof for the consistency of analysis which Gödel had sought in the early 1940s (Spector 1962). The article was published in July 1961, not long after Spector's untimely death; it was Kreisel who prepared the article for print. In the postscript to Spector's article, Gödel writes that whereas Spector had stayed at the IAS during the academic year 1960–61, "the discussions P. Bernays and I had with Spector [...] took place after the main result [...] had been established already" (Spector 1962, 27). However, Gödel says, Kreisel's role in Spector's work was greater. Apparently, Kreisel and Spector had originally planned to publish a joint article on the topic. Later developments of Spector's method include, among many others (see Avigad and Feferman 1998, section 6), Howard's (1968) and Luckhardt's (1973) works.

Around 1970, several variants of the interpretation emerged. The Diller-Nahm functional interpretation (Diller and Nahm 1974) avoids the problem with the axiom $A \to (A \& A)$ mentioned above so that characteristic functions

[24] The status of bar induction, which is a generalization of Brouwer's Bar Theorem to higher types, as intuitionistically acceptable is not entirely clear. As Feferman notes, unlike Gödel's interpretation, Spector's proof did not have the aspect of reducing intuitionistic mathematics to a more constructive system (Feferman 1998, 222-223). Instead, the constructivity of intuitionism is here assumed without question.

for atomic formulas are no longer needed. Shoenfield's translation (Shoenfield 1967) is a direct interpretation of Peano Arithmetic via the negative translation. Parsons' 1970 article should also be mentioned as perhaps the earliest application of the functional interpretation to subsystems of arithmetic (Parsons 1970).

In the 1990s, there was a renewal of interest in the applications of the *Dialectica* interpretation. Gödel's original idea was to secure the constructivity of intuitionistic logic by recovering the existence and the disjunction properties of intuitionistic logic in his functional interpretation. The idea of relative constructivity has been generalized and extended by Ulrich Kohlenbach, who has applied the interpretation to extract other kinds of computational content from non-constructive proofs in stronger systems (see Kohlenbach 2008).

Looking at the *Dialectica* article, and its many added footnotes and corrections in the 1972 version, it appears as though Gödel was never fully satisfied with his work.[25] Since Gödel first developed his system in detail in the Princeton and Yale lectures in 1941, the focus has shifted from philosophical questions to formal work and concrete applications, and then again to other sorts of philosophical questions in the context of the "extended Hilbert Program" of seeking maximally constructive reductions of non-constructive systems. The abundance of variations and applications born out of Gödel's functional interpretation – which can hardly be done justice to in such a restricted space[26] – shows how Gödel's quest for justification of the constructivity of intuitionistic logic led to a fruitful field of research. From this point of view, Gödel's functional interpretation was nothing less than a success which will hopefully lead to further discoveries in the years to come.

Maria Hämeen-Anttila

[25] See also the Gödel-Bernays correspondence in (Gödel 2003a).

[26] For a more complete historical picture, see (Feferman 1998) and Troelstra's comprehensive introduction to the *Dialectica* article in Gödel's Collected Works (Troelstra 1990). (Avigad and Feferman 1998) is an accessible introduction to the *Dialectica* interpretation in particular. (Kohlenbach 2008) contains plenty of remarks and references to earlier works on applications of the functional interpretation.

Princeton Lectures on Intuitionism

Notebook 1

~~1~~[1]

Programs[2] [[Arbeitsheft]] 7 p. 47 and [[Arbeitsheft]] 7 p̌. 32[3] (No 4–6)

1. My theorem on the length of proofs.
2. Non-constructibility of existential statements together with Gentzen and Church's computable functions.
3. Also my remark on the realization of intuitionistic systems by computable functions (in [[Excerptenheft]] 3, p. 147).

1. An example of a recursion scheme in number theory for which the recursion axioms are no longer provable (because of Gentzen's proof).
2. The same for a transfinite (but intuitionistic) recursion scheme for functions of ordinal numbers and the system of analysis.
3. Can one not prove elegantly the existence of a recursive "enumeration" of the recursive functions (without a construction)?

---

[1] The entire page has been cancelled.

[2] Programme A.H. 7 p. 47 und A.H. 7 p̌. 32 (No 4–6)

1. Mein Satz über die Länge von Beweise.
2. Nicht-Konstruierbarkeit der Existenzaussagen zusammen mit Gentzen und Churchs berechenbaren Funktionen.
3. Ebenso meine Anmerkung über die Realisierung der intuitionistischen Systeme durch berechenbare Funktionen (in Exc. H. 3 p. 147).

1. Beispiel für ein Rekursionsschema in der Zahlentheorie, für welches nicht mehr die Rekursionsaxiome beweisbar (aufgrund von Gentzens Beweis).
2. Dasselbe für ein transfinites (aber intuitionistisches) Rekurssionsschema für Funktionen von Ordinalzahlen und das System der Analysis.
3. Kann man nicht die Existenz einer rekursiven "Abzählung" der rekursiven Funktionen (ohne Konstruktion) elegant beweisen?

Beweis d. Ax. d. int. Logik p. 89.4

[3] This refers to page 32 in *Arbeitsheft* 7, reverse direction.

Proof of the axioms of intuitionistic logic p. 89.4

Improvements for these lectures[4]

1. p. 2 To mention that one can consider the non-constructive existence proofs as senseless but that intuitionistic logic has even independently of that a sense.

2. [[Improvement of]] the formulas of the syllogistic proof, which simplifies the proof of partial transposition. In general, many formulas involving absurdity become earlier formulas directly when $\neg p$ is replaced by $p \to W$.

3. The requirement of provability of every classical identity of the functional calculus: $\neg\neg A \to A$ for atomic formulas (or $\neg A \vee A$ for atomic formulas, which implies the same for unquantified expressions) and mention that impredicative procedures are excluded.

[4] Verbesserungen dieser Vorlesungen

1. p. 2 Erwähnen, dass man die nicht-konstruktiven Existenzbeweise als sinnlos ansehen kann, aber auch unabhänging davon die intuitionistische Logik einen Sinn hat.
2. Die Formeln des syllogistischen Beweises, was den Beweis der partiellen Transposition vereinfacht. Überhaupt, viele Formeln betreffs Absurdität werden direkt frühere Formeln, wenn $\neg p$ durch $p \to W$ ersetzt wird.
3. Voraussetzung des Satzes über die Beweisbarkeit jeder klassischen Identität des Funktionenkalküls: $\neg\neg A \to A$ für Atomformeln (oder $\neg A \vee A$ für Atomformeln, was dasselbe für unquantifizierte Ausdrücke zur Folge hat) und Erwähnung, dass imprädikative Verfahren ausgeschlossen sind.
4. Weglassen, dass ein Axiomsystem nur dann intuitionistisch sinnvoll [[ist]], wenn die Grundbegriffe entscheidbar sind.
5. Bei der Ableitung der vielen Formeln das Ziel vorher erwähnen, dass jede Wahrheitstabelle beweisbar ist.
6. Für den Fall, dass mehr Zeit
   A. Theorem über $\neg, \cdot$ wobei $[\neg(p.\neg q) \equiv \neg\neg(p \to q)$ und ebenso für $\vee]$.
   B. Deduc. Th. des Aussagenkalküls.
   C. Am Anfang mehr auf die int[[uitive]] Bedeutung der intuitionistischen Axiome eingehen und auf philosophische Fragen der Konstruktivität.

4. Leave out that an axiom system is intuitionistically meaningful only if the basic concepts are decidable.

5. In the derivation of the various formulas, mention first the goal that each truth table is provable.

6. For the case of more time

   A. Theorem about $\neg$, $\cdot$ in which $\neg(p \cdot \neg q) \equiv \neg\neg(p \to q)$ and the same for $\vee$.

   B. Deduction theorem of the propositional calculus.

   C. In the beginning, go more into the intuitive meaning of the intuitionistic axioms and into philosophical questions about constructivity.

I

1. ~~Have to make use of symbols hence begin by making a list:~~

2. I.

   A. ~~Variables with fixed domain of variation: x, y integers~~

   B. ~~Variables for functions of integers, etc.~~

II. Operations of the calculus of propositions.

$(\sim p), \neg p, (p \supset q), p \rightarrow q, p \cdot q, (p \,\&\, q), p \vee q, p \equiv q$

Quantifiers

Existential quantifier $(\exists x)A(x)$
Universal quantifier $(x)A(x)$

Symbol of identity $=$

3. Rules and axioms for these symbols have been set up in accordance with how these symbols are used in actual mathematics. A closer examination of these rules leads to the following surprising fact:[5]

[5] This sentence replaces another that reads:

2

very often these rules yield a proof for an existential proposition e.g. of the form $(\exists x)A(x)$ but this proof gives no method actually to find such an integer a.

4. Let's take the following example:

Call a number a a Goldbach number if $2a$ is the sum of two primes, and define a sequence a_n as follows:

$$a_0 = 1$$

$$a_{n+1} = a_n + 1 \quad \text{if } a_n \text{ is a Goldbach number}$$

$$a_{n+1} = a_n \qquad \text{if } a_n \text{ is not a Goldbach number}$$

Then the sequence $\frac{1}{a_n}$ evidently has a condensation point because contained in the interval $[0, 1]$ (it is even convergent but that does not interest us now). You can also prove that it has a rational condensation point, namely either 0 or one of the numbers $\frac{1}{n}$ according as to whether the Goldbach theorem is true or not. So we can prove that there exists a rational number which is a condensation point but the proof gives no way to find it (because in order to find it you would have to solve Goldbach's problem).

5. Simplifying this example as much as possible we

3

obtain the following. Let P be any at present undetermined proposition and define a property of integers $\varphi(x)$ as follows:

$$\varphi(x) \equiv (x = 0 \cdot P) \vee (x = 1 \cdot {\sim}P)$$

Then we can prove: There exists a number $(\exists x)\varphi(x)$, namely if P is true 0 is such a number, if P is false then 1, but the proof gives no way of finding it.

6. From such considerations, the problem arises how to axiomatize mathematics in such a manner that such undesirable things as non-constructive existence proofs can never happen i.e. such that the proof of any existential proposition yields a way to find the thing whose existence is asserted. Let's call a logic in which this is the case a constructive logic.

[But a closer examination of these rules gives rise to certain objections against the way in which these logical symbols are used in mathematics. The most obvious starting point of these objections is this that]

7. In this example and also in the foregoing, you see very clearly which axiom of classical logic is responsible for the non-constructive existence proof. It is the law of excluded middle for arbitrary propositions, even such as we are not able to decide since in both cases, in order to prove the existence of this number x, we have to make a distinction of two cases according to whether a proposition P, which we cannot decide, is true or false. So the law of excluded middle will be one of the principles left out in a constructive logic.

4

8. There is another way of arriving at a non-constructive existence proof, namely the following one. Assume I succeed to derive a contradiction from the assumption $(x){\sim}\varphi(x)$. Then one will conclude in classical mathematics that $(\exists x)\varphi(x)$, but such a contradiction obtained from $(x){\sim}\varphi(x)$ need not necessarily yield a way to construct a number x for which $\varphi(x)$ is true even if the contradiction from $(x){\sim}\varphi(x)$[6] is obtained in a perfectly constructive way i.e. without using the law of excluded middle and similar things. You can see this from both examples given before. In both cases you can derive the contradiction from the negative proposition perfectly constructively.

9. Now how do we conclude from this contradiction to this existential assertion? First we conclude from the contradiction obtained ${\sim}(x){\sim}\varphi(x)$, but $(x){\sim}\varphi(x)$ means

5

the same thing as ${\sim}(\exists x)\varphi(x)$. Hence we conclude ${\sim}{\sim}(\exists x)\varphi(x)$, but now from this we conclude $(\exists x)\varphi(x)$, and this apparently is the place where the non-constructive element comes in. So the second principle we shall have to leave out will be the law of double negation ${\sim}{\sim}p \rightarrow p$.

10. This whole argument is of course only heuristic and it may seem a little arbitrary why we decide to drop just these two logical rules and not others which were likewise necessary to obtain the non-constructive existence proofs mentioned before. In order to find out in a systematic way which rules are to

[6] A negation is missing in the original text.

6

⟦be⟧ kept & which are to be dropped we have to analyze the meaning of these primitive terms of logic. In particular, we have to find out which existential assertions are concealed in them and then we have to admit only such axioms and such rules of inference as allow to construct the existential assertions contained in them. Then we can expect to obtain a constructive logic in the sense defined before.

7

11. The attitude of the intuitionists themselves concerning the meaning of these logical notions is this, that they take them as primitive and therefore cannot give any justification for their axioms but evidence. I don't think that this attitude is necessary but I think that these notions *can* be defined in terms of much simpler and clearer ones, at least in their application to definite mathematical theories e.g. number theory or analysis. To give such a definition and a consequent proof of the intuitionistic axioms is the chief purpose of these lectures. Only this definition, by the way, yields a proof that intuitionistic logic really is constructive in the sense defined before which is by no means trivial. But before I can give this definition of the primitive terms of intuitionistic logic, I must first develop intuitionistic logic to a certain extent in the usual axiomatic way

8

where all these notions are taken as primitive. But even if we choose this course, I think it is desirable first to explain the meaning of the primitive notions in terms of everyday language in the same manner as Euclid begins his elements by an explanation of the primitive terms, although he never uses these explanations ⟦in⟧ the subsequent proofs. This procedure is necessary in order to see that the assumption of certain axioms and the rejection of others is not arbitrary but corresponds to certain intuitions, although these intuitions are perhaps of a more or less vague nature. What is perfectly clear in this axiomatic treatment of intuitionistic logic is only that all theorems follow from the axioms and rules of inference.[7]

14. By far the most important and interesting of these notions here is $p \rightarrow q$. Now to explain the meaning of a proposition in a constructive system means to

[7] Items 12 and 13 are missing.

state under which circumstances one is entitled to assert it. And the answer in this case is: If one is able to deduce q from the assumption p. But one has to be careful: the assumption p in

9

a constructive logic means the assumption that a proof for p is given, since truth in itself without proof makes no sense in a constructive logic. So $p \to q$ means: Given a proof for p one can construct a proof for q or in other words: One has a method to continue any given proof of p to a proof of q. It is quite essential that $\to$ is not interpreted as meaning q is deducible from the assumption that p is true because certain theorems of intuitionistic logic don't hold for it.

[The following theorem, e.g., is true in intuitionistic logic: $p \to (q \to p)$, i.e., .[8] But it is not true that from the truth of p it follows that p is deducible from any assumption q, because p might be true and not demonstrable, hence not deducible, say, from the assumption $0 = 0$. But if a proof for p is given, then of course I can deduce p from any assumption q, since I can deduce it even from no assumption.]

10

15. The next notion to be explained is negation and this is the notion which differs most of all in classical and intuitionistic logic. Therefore it is denoted by another name "absurdity" and by another symbol $\neg$ instead of $\sim$.

Now if one wants to find out the circumstances under which one is entitled to assert $\sim p$ in mathematics quite generally, one can hardly think of anything else but: if a contradiction can be derived from p.

So let us denote some absurd proposition, e.g. the proposition $0 = 1$ by W; then $\neg$ is defined by

$$\neg p =_{Df} (p \to W).$$

It will turn out later that it is arbitrary which

11

absurd proposition you take for W because we shall see that any absurd proposition is deducible from any other in intuitionistic logic. You see at once that

[8] Gödel has left a gap in the text here.

one cannot expect the law of double negation to hold for this negation because $\neg\neg p$ means $(p \rightarrow W) \rightarrow W$. So it means: one can prove that a contradiction cannot be derived from p i.e. it means freedom from contradiction of p (to be more exact, demonstrable freedom from contradiction) or undisprovability of p, hence something quite different from p itself.

16. The next notion here is “and” and this notion is really so simple that an explanation is hardly possible or necessary.

17. The next is “or” and the constructive meaning

12

of $p \vee q$ will evidently be: one has a procedure of which one knows that it must lead either to a proof of p or of q. So e.g. $p \vee \neg p$ cannot at the present time be asserted about Fermat’s last theorem because that would mean one has a procedure either to prove it or to derive a contradiction from it, but one can assert it for the statement $2^{(2^{10})} + 1$ is a prime number, although one may not be able actually to decide this question because it would take too long to carry out the necessary calculations.

18. $p \equiv q$ means by definition $p \rightarrow q \cdot q \rightarrow p$.

19. The existential quantifier $(\exists x)\varphi(x)$ will evidently mean: I have a method to find a number a and a proof for $\varphi(a)$; and $(x)\varphi(x)$ means I have a method to prove $\varphi(a)$ for any number a which is given to me.

20. So this is the intuitive meaning of the logical notions in intuitionistic logic, and you will perhaps agree

13

with me that these notions are pretty complicated for primitive terms and that also they don’t have the desirable degree of clearness because they involve the notion of a *procedure* and of a *proof* which are perhaps not so absolutely clear. ~~(Now my interpretation which I am going to give later does not pretend to clarify these properly logical notions in quite general sense that would be an impossible enterprise but only as applied to certain definite mathematical systems, e.g. number theory.)~~

21. But I don’t want to spend any more time about these questions of meaning but shall now set up the axioms for these notions and I leave it to you to verify

that they are evident i.e. that all constructions asserted in them can be carried out

14

for the meaning of these symbols I have explained.

22. Intuitionistic logic has first been axiomatized by Heyting in 1930 in the Handlungen der Preussischen Akademie der Wissenschaften, later in an improved form by Gentzen (Math. Zs 39).

I shall use here a system of axioms which is different from both but more closely related to Gentzen's. I confine myself at first to the notions in the first line (excluding quantifiers) i.e. to the calculus of propositions. There we have the following primitive terms.

1. A certain class $\mathfrak{P}$ of things called propositions and denoted by $A, B, \ldots$
2. Three binary operations $\rightarrow\ \vee\ \cdot$ which applied to propositions

15

yield again propositions.

3. A certain proposition W called the absurdity.
4. A certain subclass of $\mathfrak{P}$, the *asserted* propositions, or rather, the propositions which *can be asserted*.

   That A is *assertable* is denoted by $\vdash A$, but I shall not make much use of this symbol $\vdash$ but rather state in words that A can be asserted.

The axioms are as follows:[9]

1. Arbitrary propositions A, B

[9] The opposing page is numbered 15.1 and begins with the definition of negation and equivalence, cancelled but then repeated and indicated as belonging to the end of page 15, followed by a cancelled incomplete sentence. This is followed by a shorthand passage:

> For negation, one needs instead of 6. the two axioms
>
> 1. $(A \supset B \cdot \neg B) \supset \neg A$
> [2. $(B \cdot \neg B) \supset A$]
>
> The second becomes superfluous if one has $\neg\neg A \supset A$.

1. $A \rightarrow A$
2. $A \cdot B \rightarrow A$
3. $A \cdot B \rightarrow B$
4. $A \rightarrow A \vee B$
5. $B \rightarrow A \vee B$
6. $W \rightarrow A$

About the last axiom I shall speak later.

II Group (rules of inference)

1. $\dfrac{A \qquad A \rightarrow B}{B}$ Rule of implication

2. $\dfrac{A \rightarrow B \qquad B \rightarrow C}{A \rightarrow C}$ Rule of syllogism

3. $\dfrac{A \rightarrow B \qquad A \rightarrow C}{A \rightarrow B \cdot C}$ Rule of conjunction

As you see the second group of axioms differs from the first in so far that they state that certain propositions can be asserted if others can. Also the theorems fall into these two groups.

4. $\dfrac{A \rightarrow C \qquad B \rightarrow C}{A \vee B \rightarrow C}$ Rule of disjunction

5. $\dfrac{A \cdot B \rightarrow C}{A \rightarrow (B \rightarrow C)}$ Export

6. $\dfrac{A \rightarrow (B \rightarrow C)}{A \cdot B \rightarrow C}$ Import

That's all.

Absurdity is introduced by the definition $\neg A =_{Df} A \rightarrow W$ and Aequivalence by $A \equiv B =_{Df} (A \rightarrow B) \cdot (B \rightarrow A)$.

16

This system is very natural and in addition has a certain symmetry with respect to "or" and "and". But for my purposes it is better to have a slightly less symmetric system. Namely axioms 1 and 2 can be replaced by

$1'.\ A \rightarrow A \cdot A \qquad 2'.\ A \cdot B \rightarrow B \cdot A$

and these two axioms make rule 3 superfluous. And this will be the system on which the subsequent deductions are based.

Let's first deduce the old axioms from the new ones.

1. We have

$$\begin{array}{ll} A \rightarrow A \cdot A & (1') \\ A \cdot A \rightarrow A & (3) \\ \hline A \rightarrow A & \text{by syllogism} \end{array}$$

2. Similarly 2 from 3 and 2′.

3. Rule

$$\begin{array}{ll} A \rightarrow (A \rightarrow B) & \text{simplification} \\ \hline A \rightarrow B & \end{array}$$

From the assumption

$$\begin{array}{ll} A \cdot A \rightarrow B & \text{by import} \\ A \rightarrow A \cdot A & \text{by } (1') \\ \hline A \rightarrow B & \text{syllogism} \end{array}$$

4.

$$\begin{array}{ll} A \rightarrow (B \rightarrow C) & \text{commutation} \\ \hline B \rightarrow (A \rightarrow C) & \end{array}$$

$$\begin{array}{ll} A \cdot B \rightarrow C & \text{import} \\ B \cdot A \rightarrow C & \text{by axiom } 2' \text{ and syllogism} \\ \hline B \rightarrow (A \rightarrow C) & \text{export} \end{array}$$

5. Formula $B \rightarrow (C \rightarrow B \cdot C)$

$$\begin{array}{ll} B \cdot C \rightarrow B \cdot C & (1) \\ \hline B \rightarrow (C \rightarrow B \cdot C) & \text{export} \end{array}$$

6. Now finally Rule 3

$$\begin{array}{ll} A \rightarrow B & \\ A \rightarrow C & \\ \hline A \rightarrow (C \rightarrow B \cdot C) & \text{syllogism formula (5)} \\ C \rightarrow (A \rightarrow B \cdot C) & \text{commutativity} \\ A \rightarrow (A \rightarrow B \cdot C) & \text{syllogism} \\ \hline A \rightarrow B \cdot C & \end{array}$$

7. General Leibniz

$$\frac{\begin{array}{l}A \to B\\ C \to D\end{array}}{A \cdot C \to B \cdot D}$$

$$\begin{array}{ll}A \cdot C \to A & \\ [\![A \to B]\!] & \\ A \cdot C \to B & \text{syllogism}\\ A \cdot C \to C & \\ [\![C \to D]\!] & \\ A \cdot C \to D & \text{syllogism}\\ \hline A \cdot C \to B \cdot D & \text{conjunction}\end{array}$$

7′.

$$\frac{A \to B}{A \cdot C \to B \cdot C}$$

18

It is to be noted that the axioms *for "or"* were not used so far and generally. It can be proved that we never have to make use of these axioms for "or" for proving formulas not containing "or" so that the remaining axioms form a closed system in themselves which is exactly what is usually called *positive logic*, provided you leave out the W and the axioms concerning it.

I shall not carry through the proofs of all theorems which I need but only list them and give some indication of the proof.

8. Addition of premisses

$$\frac{A \to B}{A \cdot C \to B} \quad \text{syllogism, axiom 2}$$

9. $\cdot$ is commutative and associative, or more generally: If you have any two expressions A, B composed of only letters $A_1, \ldots A_n$ and the symbol $\cdot$ (where the same letter may occur in different places) and if only every letter in B occurs also in A then $A \to B$ can be asserted.

Proof by induction on the number of letters occurring in B. For one letter in B it follows by iterated application of 8. If B contains more than one

19

letter $B = B_1 \cdot B_2$ where B_i [[have]] fewer [[symbols than B]]

$A \to B_1, A \to B_2$ holds by induction hence

$A \to B_1 \cdot B_2$

This theorem together with rule of syllogism allows you to interchange different terms in a premiss $A_1 \cdot A_2 \ldots A_n \to B$ and to strike out terms which appear doubly (which is one of the axioms in Gentzen's system).

20.[10]

$$\begin{array}{ll} A \to [(A \to B) \to B] & \\ [A \cdot (A \to B)] \to B & \text{export} \\ \hline (A \to B) \to (A \to B) & \text{commutativity} \end{array}$$

20′.

$$\begin{array}{l} A \to B \\ A \to (B \to C) \\ \hline A \to C \end{array} \qquad \begin{array}{l} B \to (A \to C) \\ A \to (A \to C) \\ \hline A \to C \end{array}$$

21.

21′. Corresponding rule of inference

$$\begin{array}{ll} A \to (B \to A) & \\ A \cdot B \to A & \text{export} \end{array}$$

22. Multiplying an implication with $C \to$, $\to C$

$$\begin{array}{l} A \to B \\ \hline 1.\ (C \to A) \to (C \to B) \\ 2.\ (B \to C) \to (A \to C) \end{array}$$

$$\begin{array}{ll} (C \to A) \cdot C \to A & \\ (C \to A) \cdot C \to B & \text{syllogism} \\ \hline (C \to A) \to (C \to B) & \text{export} \end{array}$$

$$\begin{array}{ll} B \to [(B \to C) \to C] & \\ A \to [(B \to C) \to C] & \text{syllogism} \\ \hline (B \to C) \to (A \to C) & \text{commutativity} \end{array}$$

[10] There are no items 10 to 19.

19·1

Derivation of the formulas from their corresponding rules

1.

$$
\begin{array}{l}
P \supset (A \supset B) \\
P \supset (B \supset C) \\
A \cdot P \supset B \\
A \cdot P \supset B \cdot P \\
B \cdot P \supset C \\
A \cdot P \supset C \\
P \supset (A \supset C)
\end{array}
$$

so for $P = (A \supset B) \cdot (B \supset C)$

$$\underline{\underline{((A \supset B) \cdot (B \supset C)) \supset (A \supset C)}}$$

2.

$$
\begin{array}{l}
B \cdot A \supset B \\
B \supset (A \supset B)
\end{array}
$$

3.

$$
\begin{array}{ll}
P \supset (A \supset B) & \\
P \cdot A \supset B & \\
P \cdot A \cdot X \supset B \cdot X & \\
P \supset (A \cdot X \supset B \cdot X) & P = A \supset B \\
\underline{\underline{(A \supset B) \supset (A \cdot X \supset B \cdot X)}} &
\end{array}
$$

19·2

4. Proof [[for]]export

$$
\begin{array}{ll}
P \supset (A \cdot B \supset C) & \\
P \cdot A \cdot B \supset C & \\
P \cdot A \supset (B \supset C) & \\
P \supset (A \supset (B \supset C)) & P = A \cdot B \supset C
\end{array}
$$

5. Import

$$
\begin{array}{ll}
P \supset (A \supset (B \supset C)) & \\
P \cdot A \supset (B \supset C) & \\
P \cdot A \cdot B \supset C & \\
P \supset (A \cdot B \supset C) & P = A \supset (B \supset C)
\end{array}
$$

6.

$$\begin{array}{l} P \supset (A \supset C) \\ P \supset (A \supset B) \\ \hline P \cdot A \supset B \cdot C \\ P \supset (A \supset B \cdot C) \quad P = A \supset C \cdot A \supset B \end{array}$$

19·3

7.

$$\begin{array}{l} P \supset (A \supset C) \\ P \supset (B \supset C) \\ \hline A \supset (P \supset C) \\ B \supset (P \supset C) \\ A \vee B \supset (P \supset C) \\ P \supset (A \vee B \supset C) \quad P = (A \supset C) \cdot (B \supset C) \end{array}$$

8.

$$\begin{array}{l} P \supset (x)[A \supset F(x)] \quad P \text{ [[and]] } A \text{ are } x\text{-free} \\ P \supset (A \supset F(x)) \\ P \cdot A \supset F(x) \\ P \cdot A \supset (x)F(x) \\ P \supset (A \supset (x)F(x)) \quad P = (x)[A \supset F(x)] \end{array}$$

9.

$$\begin{array}{l} P \supset (A \supset (x)F(x)) \\ P \cdot A \supset (x)F(x) \\ P \cdot A \supset F(x) \\ P \supset (A \supset F(x)) \\ P \supset (x)[A \supset (x)F(x)] \end{array}$$

19·4

10.

$$\begin{array}{l} P \supset (x)[A(x) \supset B] \\ A(x) \supset (P \supset B) \\ (\exists x)A(x) \supset (P \supset B) \end{array}$$

ii.

$$P \supset ((\exists x)A(x) \supset B)$$
$$(\exists x)A(x) \supset (P \supset B)$$
$$A(x) \supset (P \supset B)$$
$$P \supset (A(x) \supset B)$$
$$P \supset ((x)A(x) \supset B)$$

19·5

Last time I set up a system of axioms for the intuitionistic calculus of proposition with the primitive terms $\rightarrow \ \cdot \ \vee \ W$ and the following axioms

$$A \rightarrow A \cdot A \qquad A \rightarrow A \cdot B \qquad A \cdot B \rightarrow B \cdot A$$
$$A \rightarrow A \vee B \qquad B \rightarrow A \vee B \qquad W \rightarrow A$$

and the following rules of inference

$$\frac{A \quad A \rightarrow B}{B} \qquad \frac{A \rightarrow B \quad B \rightarrow C}{A \rightarrow C} \qquad \frac{A \cdot B \rightarrow C}{A \rightarrow (B \rightarrow C)} \qquad \frac{A \rightarrow C \quad B \rightarrow C}{A \vee B \rightarrow C}$$

Implication — Syllogism — ↕ Export / Import — Disjunction

If $\mathfrak{A}$ is an expression composed of the primitive terms $\rightarrow \ \cdot \ \vee \ W$ and of letters $A, B, \ldots$ denoting arbitrary propositions and if it can be proved from these axioms that $\mathfrak{A}$ can be asserted for any propositions $A, B, \ldots$ then $\mathfrak{A}$ is called an identity of the calculus of propositions; e.g. I proved last time $A \rightarrow ((A \rightarrow B) \rightarrow B)$ is an identity. But most of the theorems I proved last time were of a different nature, namely similar to the second group of axioms i.e. derived rules of inference. Moreover, the main

19·6

interest lies in the identities and the rules of inference are only means of obtaining them. Now the identities have the int⟦eresting⟧ property that they form themselves a model for the axioms. To be more exact, if you interpret the primitive terms in the following way

1. Proposition means expression composed of letters and these symbols $\rightarrow$ $\cdot$ $\vee$.

2. To apply the operation of $\rightarrow$ to two expressions A, B means to form the expression $A \rightarrow B$ and similarly for the other operations.

3. The class of assertable propositions are the identities.

Then it is easily seen that all axioms of both groups are satisfied and it is this particular model which one usually has in view if one speaks of the calculus of propositions. This also leads to a more direct definition of identities, namely: = anything obtained from expressions of these six forms where A, B now denote arbitrary expressions by a finite number of their rules of ⟦inference⟧.

Now let us continue the deduction from the axioms. So far I have only proved theorems about $\rightarrow$. Now let's begin with *absurdity*.

20

Theorem about $\vee$: commutativity, associativity[11]

23.

$$\frac{\begin{array}{l} A \rightarrow B \\ C \rightarrow D \end{array}}{A \vee C \rightarrow B \vee D} \qquad \frac{\begin{array}{l} A \cdot B \rightarrow C \\ A' \cdot B \rightarrow C \end{array}}{(A \vee A') \cdot B \rightarrow C} \quad \text{by export and import}$$

23.1

$$\begin{array}{ll} \neg A \vee B \rightarrow (A \rightarrow B) & \text{but not vice versa} \\ \neg A \rightarrow (A \rightarrow B) & \text{proved before} \\ B \rightarrow (A \rightarrow C) & \end{array}$$

23.1

$$(\neg A \vee A) \rightarrow (\neg\neg A \rightarrow A) \quad \text{same proof}$$

23.2

$$A \vee B \rightarrow \neg(\neg A \cdot \neg B)$$

24. Next come the theorems about $\neg$.

[11] Item 23, two items both labelled 23.1, and item 23.2 have been cancelled.

Axiom 6 cannot be proved but almost:

$W \to \neg A$ since $W \to (A \to W)$

Therefore this axiom is used pretty rarely in the subsequent development. I shall always state it explicitly if a theorem depends on it. The essential results I have in view are independent of this axiom 6 and also the axioms about or. But nevertheless I think this axiom is perfectly justified in intuitionistic logic. It holds for the interpretation I am going to give but also for the meaning of the symbols which I explained in the preceding lecture, since $W \to A$ means ⟦that⟧ one has a procedure to construct a proof of A if a proof for W is given. But every procedure will do this because it can never happen that a proof for W is given.

24.

$\neg A \to (A \to B)$	this depends on axiom 6
$(A \to W) \to (A \to B)$	
$W \to B$	multiplication with $A \to$

21[12]

25.

$$\frac{A \to B}{\neg B \to \neg A}$$

means

$(B \to W) \to (A \to W)$ multiplication $\to W$

26.

$$\frac{A \to \neg B}{B \to \neg A}$$ transposition 2. kind i.e. $\to \neg$ commutativity

[12] Facing right page contains the following derivation:

[$(A \to B) \to (\neg B \to \neg A)$ formula of syllogism]
$(A \to B) \cdot (B \to C) \cdot A \to B$
(″ ″ $B \to C$
(″ ″ $A \to C$
Application $\to (A \to C)$
$\neg A \to \neg(A \cdot B)$

$$\frac{A \cdot B \to W}{\neg(A \cdot B)} \quad \text{equivalent to both by export and import}$$

Hypothesis: because $A \to (B \to W)$ consequence $B \to (A \to W)$.

27.

$$\frac{C \to \neg\neg C}{\neg C \to \neg C} \quad \text{from (26)} \qquad\qquad C \to [(C \to W) \to W]$$

28.

$$\neg\neg\neg C \to \neg C \quad \text{transposition}$$

29. *Partial transposition*

$$\frac{C \cdot A \to B}{C \cdot \neg B \to \neg A} \qquad\qquad \frac{C \cdot A \to \neg B}{C \cdot B \to \neg A}$$

For a proof:

$$(C \cdot B \to W) \cdot A \to W$$

$$\begin{array}{ll} B \to [(B \to W) \to W] & \\ C \cdot A \to [(B \to W) \to W] & \text{syllogism} \\ C \cdot \neg B \to \neg A & \text{import} \end{array}$$

$$C \cdot A \cdot B \to W$$

$$\begin{array}{ll} C \cdot B \cdot A \to W & \text{import} \\ C \cdot B \to \neg A & \text{export} \end{array}$$

30.

$$\frac{\begin{array}{l} A \to B \\ A \to \neg B \end{array}}{\neg A}$$

$$\frac{\begin{array}{l} A \to B \\ A \to (B \to W) \end{array}}{A \to W}$$

30*

$$\frac{\begin{array}{l} A \cdot \neg A \to B \\ A \cdot (A \to W) \to W \end{array}}{W \to B}$$ syllogism

31.

$$A \cdot \neg B \to \neg(A \to B)$$

$A \cdot (A \to B) \to B$

$A \cdot \neg B \to \neg(A \to B)$ partial transposition

32.

$\neg\neg(A \to B) \to (\neg\neg A \to \neg\neg B)$ distrib. for $\neg\neg$

$\neg\neg(A \to B) \cdot \neg B \to \neg A$ partial transposition 31

$\neg\neg A \to \neg\neg B$ 〃 〃

$\neg\neg(A \to B) \to (\neg\neg A \to \neg\neg B)$ export

Inverse

$\neg A \to (A \to B)$

$\neg(A \to B) \to \neg\neg A$

$(\neg\neg A \to \neg\neg B) \cdot \neg(A \to B) \to \neg\neg(A \cdot (\neg\neg A \to \neg\neg B))$

$(\neg\neg A \to \neg\neg B) \cdot \neg(A \to B) \to \neg\neg B$

$\neg\neg B \to \neg\neg(A \to B)$

$(\neg\neg A \to \neg\neg B) \cdot \neg(A \to B) \to$ contradiction

33.

$\neg\neg(A \cdot B) \to \neg\neg A \cdot \neg\neg B$

$A \cdot B \to A$

$\neg\neg A \cdot B \to \neg\neg A$ etc.

Inverse

$A \to (B \to A \cdot B)$

$\neg\neg A \to \neg\neg(B \to A \cdot B)$

$\neg\neg(B \to A \cdot B) \to (\neg\neg B \to \neg\neg(A \cdot B))$

$\neg\neg A \to (\neg\neg B \to \neg\neg(A \cdot B))$ syllogism

23

34. Finally we need the following formula for "or"

$$\neg A \cdot \neg B \rightarrow \neg(A \vee B)$$

but that is

$$(A \rightarrow W) \cdot (B \rightarrow W) \rightarrow [(A \vee B) \rightarrow W]$$

that is, one of our rules of inference but expressed as a formula. Now it can be shown quite generally that for any demonstrated rule of inference also the corresponding formula can be demonstrated but I shall not prove this case ⟦which is⟧ quite simple.

Sufficient ⟦to show that⟧

$A \vee B \rightarrow [(A \rightarrow W)(B \rightarrow W) \rightarrow W]$ by commutativity

but $A \rightarrow [$ 〃 〃 〃 $]$ can be ass⟦erted⟧

since $[A \cdot (A \rightarrow W)] \rightarrow W$ can, hence $A \cdot (A \rightarrow B) \cdot (B \rightarrow W) \rightarrow W$ by addition of premiss, hence the theorem ⟦follows⟧ by export. But in the same way $B \rightarrow [\quad]$ can be proved hence $A \vee B$ by the rule of disjunction.

1.* $(\neg A \vee A) \rightarrow (\neg\neg A \rightarrow A)$

1.

$A \rightarrow (\neg\neg A \rightarrow A)$
$\neg\neg\neg A \rightarrow (\neg\neg A \rightarrow A)$
$\neg A \rightarrow \neg\neg\neg A$
$\neg A \rightarrow (\neg\neg A \rightarrow A)$
$(\neg A \vee A) \rightarrow (\neg\neg A \rightarrow A)$ depends on axiom 6

2. $A \vee B \rightarrow \neg(\neg A \cdot \neg B)$

$\neg A \cdot \neg B \rightarrow \neg A$
$\neg\neg A \rightarrow \neg(\neg A \cdot \neg B)$ but $A \rightarrow \neg\neg A$
$A \vee B \rightarrow \neg(\neg A \cdot \neg B)$

3.

$$\frac{\begin{array}{l} A \to C \\ \neg A \to C \end{array}}{\neg\neg C} \qquad \text{namely} \qquad \frac{\begin{array}{l} \neg C \to \neg A \\ \neg C \to \neg\neg A \end{array}}{\neg\neg C}$$

35. Now what is the relationship of this calculus of propositions to the classical? There we have the following results:

1. $\neg\neg A \to A$ is not an identity, proved by Heyting. This was shown in the paper of Heyting I quoted last time.
2. If we add the axiom $\neg\neg A \to A$ then we obtain the classical calculus, and
3. If A is an identity of the classical calculus $\neg\neg A$ is an identity of the intuitionistic calculus.

24

I shall give a brief indication of a proof for the second of these theorems. In classical logic the meaning of the logical operations can be explained by the truth tables e.g.

$A \cdot B$

$_A{}^B$	T	F
T	T	F
F	F	F

$A \to B$

$_A{}^B$	T	F
T	T	F
F	T	T

etc., and a formula is true in the classical calculus if it gives the truth value T whatever truth values you may assign for the letters contained in it. Now the truth tables can in a sense themselves be expressed by formulas, e.g. the second by the following four formulas:

$$A \cdot B \to (A \to B)$$
$$A \cdot \neg B \to \neg(A \to B)$$
$$\neg A \cdot B \to (A \to B)$$
$$\neg A \cdot \neg B \to (A \to B)$$

and now we have the remarkable fact that all formulas corresponding to the truth tables of any

25

of the logical operations $\neg$, $\rightarrow$, $\cdot$, $\vee$, are identities also in intuitionistic logic. E.g. we had $B \rightarrow (A \rightarrow B)$, hence by addition of premiss $A \cdot B \rightarrow (A \rightarrow B)$. In order to see that you can prove all of these formulas and the analogous ones for the other notions. You have only to work up the formulas which we have proved already and sometimes add a premiss. But from this fact it follows by complete induction that also for a composite expression A which contains any number of letters, say $A_1, \ldots, A_n$, you can prove its truth table or rather the formulas corresponding to its truth table in intuitionistic logic. I.e. let U be any conjunction of these letters or negations of these letters. Let's call such a conjunction a primitive conjunction (here you have all the primitive conjunctions of two letters A_1, A_2). For n letters there are exactly 2^n primitive conjunctions. Then I say if U is any primitive conjunction of the letters $A_1, \ldots, A_n$ contained in the expression A then either $U \rightarrow A$ or $U \rightarrow \neg A$ is an identity

26

in intuitionistic logic (according as to whether this or that is true in classical logic). The theorem is of course true if U is a primitive conjunction of perhaps more letters than those contained in A and in this form we prove it by induction on the number of letters of which A is composed. If A is a single letter it is trivial. If it contains more than one letter it must be of the form $B \circ C$ where $\circ$ is some of the logical operations ($\vee \cdot \rightarrow$) and where B, C contain fewer letters. Hence for B, C we have already $\underline{U \rightarrow B}$ or $U \rightarrow \neg B$ and $U \rightarrow C$ or $\underline{U \rightarrow \neg C}$ is an identity, hence ⟦there are⟧ four cases. Assume e.g. the underlined, then

$$
\begin{aligned}
&U \rightarrow B \cdot \neg C \quad \text{but} \\
&B \cdot \neg C \rightarrow B \circ C \\
&\qquad\quad \rightarrow \neg(B \circ C)
\end{aligned}
$$

Hence by the rule of syllogism $U \rightarrow B \circ C$ or $U \rightarrow \neg(B \circ C)$ is an identity which was to be proved and similarly in the other three cases are dealt with. From this argument it follows that if A is an identity of the classical calculus, then $U \rightarrow A$ is an intuitionistic identity for any primitive conjunction U of the letters contained in A

27

and we have now to deduce from this that then A itself is an identity and this is done by diminishing successively the number of letters contained in U. Let's denote an arbitrary primitive conjunction[13] of the letters $A_2, \ldots, A_n$ by U'. Then $A_1 \cdot U'$ and $\neg A_1 \cdot U'$ is a primitive conjunction of $A_1, \ldots, A_n$ hence

$$A_1 \cdot U' \rightarrow A, \ \neg A_1 \cdot U' \rightarrow A$$

demonstrable by export;

$$A_1 \rightarrow (U' \rightarrow A); \neg A_1 \rightarrow (U' \rightarrow A)$$

and we want $U' \rightarrow A$.[14]

So the rule of inference we need is this:

$$\frac{\begin{array}{l} P \rightarrow Q \\ \neg P \rightarrow Q \end{array}}{Q}$$

From the intuitionistic axioms alone it follows only that $\neg\neg Q$. Therefore here is the place (and the only ⟦place⟧) where we apply the additional axiom of double negation. So we can cancel successively the letters A_i from U until we finally obtain the formula A and this concludes the proof that the additional axiom $\neg\neg A \rightarrow A$ gives the whole classical logic. It is to be noted that this remains true if you leave out $W \rightarrow A$ because this is a consequence of $\neg\neg A \rightarrow A$ since $W \rightarrow \neg A$ we proved before, hence $W \rightarrow \neg\neg A$ for any A hence $W \rightarrow A$ by syllogism.

28′[15]

36. With this I am concluding this treatment of the calculus of propositions and am beginning with the theory of the quantifiers which is usually termed calculus of predicates. In this theory the primitive objects are no longer propositions

---

[13] Gödel has mistakenly written "primitive disjunction" instead of "primitive conjunction" twice in this paragraph.

[14] Gödel has forgotten to add the index 1 to the A in the antecedent several times.

[15] The five successive pages 28′, 29′, 30′, 31′, and 32′ have been ripped off the notebook. It appears that Gödel wrote them anew, pages 28–30 below, and inserted the old versions next to the new ones.

because it makes no sense to apply a quantifier to a proposition. In order that it make sense to apply (x) is something this[16]

29′

something must depend on the variable x hence be a propositional function rather than a proposition [and the result of applying the quantifiers to any A will then in general again be a propositional function because it may contain several variables. Propositions are a special case of propositional functions, namely those with 0 variables.] That we have to do with propositional functions instead of propositions gives the notion of assertion a different meaning. That a propositional function is asserted means that all propositions obtained by substitution for the variable arbitrary objects of their respective domains are asserted.

So the primitive notions of the new system will be

1. Class of propositional functions – but ⟦these are⟧ not the only primitive objects

2. Variables

3. Terms⟦, that is,⟧ a certain superclass of the variables in the applications (i.e. the models of this system)

These primitive objects are symbols and combinations of symbols. In particular the variables are usually single letters and the terms are single letters or composite expressions which

30′

denote objects of the theory under consideration, e.g. 2, but also $2+3$ ⟦are⟧ terms and also $x + y$. Terms may contain variables.

The primitive operations are

1. $\rightarrow$, $\vee$, $\cdot$ yielding propositional functions if applied to propositional functions.

[16] This page ends with a cancelled passage with items numbered from 1 to 3, almost identical with the list on p. 28 below.

2. The two operations of quantification $(x)A$, $(\exists x)A$ which applied to a propositional function and a variable give a propositional function.

3. Substitution denoted by $A\binom{x}{t}$, an operation which applied to a propositional function A, a variable x and a term t yields a propositional function A. It means: The result of substituting t for the variable x [for the free variables not those bound by quantifiers].

Next we have the notion that a propositional function is independent of the variable x, defined by $A\binom{x}{t} = A$ for any t [which means in the application that A does not contain the variable x].

31′

Now as to the axioms we have at first certain axioms concerning substitution namely

1. $A\binom{x}{x} = A$ for any variable x 1.′ $W\binom{x}{t} = W$[17]

2. Substitution is distributive with respect to $\vee \ \rightarrow \ \cdot$

   $$(A \rightarrow B)\binom{x}{t} = A\binom{x}{t} \rightarrow B\binom{x}{t}$$

3. With respect to quantifiers it behaves like this

   $$[(x)A]\binom{x}{t} = (x)A$$

   $$[(x)A]\binom{y}{t} = (x)[A\binom{y}{t}] \quad \text{for } y \neq x \quad \text{similarly for } \exists$$

From this it follows by definition of independence that $(x)A$ and $(\exists x)A$ is independent of x.

Now let us call atomic propositional function any such propositional function which ⟦is⟧ neither a disjunction nor a conjunction nor an implication nor a quantification of any other proposition. Then it is reasonable to assume as an axiom that any propositional function is obtained by a finite number of applications of the logical operations from atomic ones. Sometimes ⟦we also have⟧ uniqueness.

[17] Here Gödel has mistakenly written $A\binom{x}{x} = x$ and $W\binom{x}{t} =$.

28

36. With this I am concluding the treatment of the calculus of propositions and beginning with the theory of quantification (calculus of functions). That could be done in the same abstract way as for the calculus of propositions introducing quantification as a new operator (or rather two operators), where however the primitive objects would now have to be called propositional functions instead of propositions because in order to apply a quantifier the expressions to which you apply it must contain a variable x. But it ⟦is⟧ better for my purposes to confine attention to a partial model of this abstract system. In this model the propositions are combinations of symbols, more particular combinations of symbols of the following kinds.

1. First letters $x, y, z, \ldots$ (*called variables for* individuals).
2. Another kind of letters α, β denoting constant individuals.
3. A third kind of letters K, R, S denoting relations between individuals, monadic $K(x)$, dyadic $R(x, y)$, etc. (where to each letter is assigned a certain definite number of arguments).
4. A fourth kind of letters f, g, h denoting functions whose values and arguments are individuals again, $f(x), g(x, y)$, etc., again with any number of arguments.

29

37. Expressions containing only the first, second, and fourth kind of symbols are called *terms*. To be more exact, i.e. $x, y, \ldots\ \alpha, \beta, \ldots$ are terms and if $t_1, \ldots, t_n$ are terms and f is a function letter variable with n arguments then $f(t_1, \ldots, t_n)$ is again a term.

38. Now if $t_1, \ldots, t_n$ are terms and R is a relation letter with n arguments, then $R(t_1, \ldots, t_n)$ is called an *atomic propositional function or atomic formula*.

39. A *propositional function* or *formula* in general is defined thus:

1. Every atomic propositional function is a propositional function and W is one.

2. If A, B are propositional functions then $A \vee B$, $A \rightarrow B$, $A \cdot B$, $(x)A$, $(\exists x)B$ is again one, where (x) is any arbitrary individual variable (which may or may not occur in A and which may or may not be bound in A).

It is to be noted that the letters and formulas written on the blackboard are never themselves the formulas about which we speak but they denote these formulas. E.g. the variable x is not a definite variable of the formalism under consideration but a variable running over all variables of the form under consideration, and $\rightarrow$ is not the symbol of implication of the formalism but it denotes the operation of writing two formulas beside each other with a symbol of implication in between them. Particularly interesting from this standpoint is this operation of application. It denotes itself in a sense.

30

40. A variable in an expression to which a quantifier refers is called bound, otherwise ⟦it is called⟧ free. A variable may be bound in one place and free in another in the same *propositional function* e.g. $K(\underline{x}) \vee (x)R(x, y)$.

41. If t is a term and A a propositional function then by $A(\frac{x}{t})$ is meant the result of substituting x by t in all places where it is free. [Each quantifier occurring in a formula has a certain scope e.g. here and the binary logical operations have two scopes: Domination, inside]

42. Next we have the notion of assertion, i.e. the asserted propositional functions form a certain subclass of the propositional functions characterized by the axioms. The meaning of assertion is now slightly different because now the asserted expressions are propositional functions, i.e. contain free variables.

30·1[18]

Last time I defined what I understand by a *propositional function* and what ⟦I understand by⟧ *terms*, namely certain combinations of symbols, to be more

[18] Pages 30·1, 30·2, 30′, and 31′, as well as pages 33, 33·2, and 34·3 are ripped off pages inserted between the pages 30–31, 31–33·1, and 34–35. It is not entirely clear in which order the pages should be arranged. The summary pages 30·1 and 30·2 seem to follow the page 30. There seem to be two alternative continuations to p. 31: Page 34 is the original page 32 with the new page number heavily written over the old one; the page numbers 35 and 36 are also later additions drawn over previous page numbering. We give the original version, pages 34–36, first, followed by the new version on pages 32–33 and the addition pages 33·1, 33·2, and 34·3.

exact of four kinds of symbols: *variables for individuals*, *constant individuals, relations, functions*. I shall use the word expression for we also defined certain operations on expressions, namely $A \cdot B, A \vee B, A \rightarrow B$. These three operations produce a new propositional function out of two given ones, namely the one obtained by $(x)A$ ⟦and⟧$(\exists x)A =$ binary operations producing a propositional function out of a variable and a propositional function where this operation is performed by writing.

[⟦I⟧ wish to remark that it is more conv⟦enient⟧ to define this operation in a little different manner, namely: changing first the variable x whenever it occurs in A into another symbol, say by underlining it, and then write the quantifier in front. That comes to this: that we use another kind of symbols for bound variables (namely underlined letters). Now if we confine the word term only to expressions containing no underlined variables

30·2

one can drop the cumbersome restriction in the first two axioms of quantification we had last time. Let us adopt this not for the sequel.]

We further had a ternary operation substitution $A\binom{x}{t}$ producing a propositional function out of a propositional function, a variable and a term and we had the notion "An expression A does not depend on x" which meant: x does not occur free in A or in other words that $A\binom{x}{t} = A$ for any t. I also need the operation of simultaneous substitution for several variables $A\binom{x\ y}{t_1\, t_2} = A\binom{x}{t_1}\binom{y}{t_2}$ if y occurs in t_1.

A *proposition* is to be considered as a special kind of propositional function namely as one without free variables. These five operations and also $\neg \ \overrightarrow{\leftarrow}$ I call the logical operations. Next I define the class of asserted propositional functions or identities of the intuitionistic calculus by the following axioms.

31

Assertability of a propositional function means that the proposition can be asserted for arbitrary constants put in place of the free variables. Of course this remark is concerned only with the int⟦ended⟧ meaning of the formula and its application.

43. Now as axioms we have

1. All axioms and rules of the calculus of propositions formulated for the propositional functions, e.g. for any propositional function A, $A \rightarrow A \cdot A$ can be asserted or if A, B, C are any propositional functions and ⟦if⟧ $A \rightarrow B$, $B \rightarrow C$ can be asserted then ⟦so can⟧ $A \rightarrow C$.

2. The following axioms for the quantifiers

I group

1. For any A and any variable x and any term t: $[(x)A] \rightarrow A\binom{x}{t}$

2. $A\binom{x}{t} \rightarrow (\exists x)A$

which I call formal axioms as opposed to rules of inference to be more ex⟦act⟧.

II group

3. $\dfrac{A \rightarrow B}{A \rightarrow (x)B}$ A independent of x

4. $\dfrac{A \rightarrow B}{(\exists x)A \rightarrow B}$ B independent of x

These are the axioms of intuitionistic logic as usually stated. There are however others usually not stated explicitly but satisfied in every application and which therefore should be added.

34

42. Now let us deduce some theorems. Since we have assumed all axioms and rules of the calculus of propositions we can prove all theorems proved before replacing the term proposition by propositional function.

43. Theorem

$$(x)A \rightarrow A$$
$$A \rightarrow (\exists x)A$$

can be asserted for any A putting $x = t$ in axioms 1, 2.

44.

$$\frac{A \to B}{(x)A \to (x)B}$$

$$\begin{array}{ll}
(x)A \to A & \text{axiom} \\
(x)A \to B & \text{syllogism} \\
(x)A \to (x)B & \text{since } x \text{ not free in } A
\end{array}$$

45.

$$\begin{array}{l}
A \to B \\
(\exists x)A \to (\exists x)B
\end{array}$$

Proof ⟦is⟧ the same ⟦as in 44.⟧

46.

$$(\exists x)A \to \neg(x)\neg A$$

But not inverse ⟦as in⟧ in classical logic⟦'s⟧ definition ⟦of the quantifiers⟧.

$$\begin{array}{ll}
(x)\neg A \to \neg A & \\
A \to \neg(x)\neg A & \text{transposition 1} \\
(\exists x)A \to \neg(x)\neg A & \text{rule } \exists
\end{array}$$

46.1

$$(x)\neg A \to \neg(\exists x)A$$

Also inverse but not nec ⟦essarily⟧

Transp⟦ose⟧ to 46

46.1

$$\begin{array}{ll}
\neg\neg(x)A \to (x)\neg\neg A & \text{commutativity of } (x) \text{ with } \neg\neg, \\
 & \text{inverse ⟦does⟧ not ⟦hold⟧}
\end{array}$$

$$\begin{array}{ll}
(x)A \to A & \\
\neg\neg(x)A \to \neg\neg A & \text{transposition applied ⟦twice⟧} \\
\neg\neg(x)A \to (x)\neg\neg A &
\end{array}$$

35

47. As I remarked before if we add $\neg\neg A \to A$ for every propositional function A we obtain the ordinary calculus exactly as it was the case with the calculus of propositions.

48. But the second theorem about the intuitionistic calculus of propositions, namely the theorem by *Glivenko* that if A is an identity of the ordinary calculus then $\neg\neg A$ is one of the intuitionistic, becomes false now.

E.g. classically for every propositional function A, $(x)(A \vee \neg A)$ is an identity but intuitionistically not even $\neg\neg(x)(A \vee \neg A)$ is an identity for every propositional function⟦e.g.⟧ $A = (y)R(y, x)$ (I shall prove that later), i.e. we can add without contradiction the axiom $\neg(x)(A \vee \neg A)$ for certain propositional functions. We cannot add $\neg(A \vee \neg A)$ for any propositional function or proposition because by Glivenko's theorem $\neg\neg(A \vee \neg A)$ can be asserted for every propositional function. I.e. in intuitionistic logic, the law of excluded middle cannot be negated without contradiction for any

36

single proposition but its simultaneous assertion for a set of propositions can be negated without contradiction. Brouwer claims even more, namely that this is a theorem of intuit⟦ionism⟧ for certain propositional functions.

49. Other classic theorems of this type (where not even $\neg\neg$ is intuitionistically demonstrable):

$$(x)[\neg\neg A \rightarrow A]$$
$$(x)\neg\neg A \rightarrow \neg\neg(x)A$$

i.e. it might happen ⟦that both⟧ $(x)\neg\neg A$, $\neg(x)A$. $\neg\neg(x)A \rightarrow (x)\neg\neg A$ can be proved, hence $\neg\neg(x)A$ ⟦is⟧ stronger.

50. But instead of this Glivenko theorem we have here two other theorems connecting classical and intuitionistic logic, namely[19]

~~51. The purpose of the calculus of functions is of course to be applied to mathematical theories e.g. number theory~~

⟦Let⟧ k ⟦be⟧ a class of natural numbers[20]

$$(k)[\overbrace{(x)k(x)}^{\varphi(k)} \vee (\exists x)\sim k(x)]$$

[19] This passage continues on page 37.

[20] This shorthand passage at the bottom of p. 36 does not seem to belong anywhere in particular.

$$(k)[\varphi(k) \vee \neg\varphi(k)]$$

But ⟦there exists⟧ already a relation R such that

$$\neg(k)[\varphi(k) \vee \neg\varphi(k)]$$

where

$$\varphi(k) = (x)R(k, x)$$

32

An application of this logic consists in this: that the letters K, R for relations and f, g for functions now denote certain definite constant functions or relations. The primitive functions or relations of this theory e.g. $x = y, x > y, x + y, x \cdot y$ and the letters α, β for constant individuals denote certain definite objects of the theory e.g. π. Furthermore we have certain axioms specific for the theory under consideration (e.g. $x + y = y + x$) and perhaps rules of inference specific for the theory e.g. complete induction

$$\frac{F(0)\ ,\ F(x) \rightarrow F(x+1)}{F(x)}$$

and a proposition can be asserted in this theory if it follows from these specific axioms and rules of the theory together with the logical axioms and rules.

52. And now if this theory is at all intuitionistically meaningful we have to assume that these primitive

33

relations R and functions f are constructive, i.e. one must have a procedure which allows to decide for any given individuals $a_1, \ldots, a_n$ whether the relation $R(a_1, \ldots, a_n)$ subsists and a procedure which allows to calculate the value of any primitive functions $f(a_1, \ldots, a_n)$ for given arguments. And this has the consequence that for atomic expressions the whole classical calculus of propositions will hold. To be more exact, let us call our expression unquantified if it contains no quantifiers i.e. is built up of atomic propositions by means of the operations of calculus of propositions alone. Then of course also for all unquantified expressions you can decide their truth or falsehood for any argument i.e. $\neg A \vee A$ assertable for unquantified expressions. But this implies $\neg\neg A \rightarrow A$.

53. Therefore if we think only of this application, we can add to the intuitionistic calculus of functions the following axiom:

$\neg\neg A \to A$ can be asserted for any unquantified expression A.

By our previous theorem this has the consequence that

33·1

every identity of the classical calculus can be asserted for unquantified expressions. Therefore the final definition of identity of the intuitionistic calculus of functions is this: The identities are the smallest class of propositional functions which can be taken as the class of assertable propositions in accordance with these axioms (including this one) which means the same thing as: An identity is a propositional function obtained from formal axioms by a finite number of applications of the rules of inference. And the identities of the ordinary classical calculus are defined in an analogous way assuming this axiom for all propositional functions A (not only atomic ones).

In the application it happens very often that we have several kinds (or *types* of individuals) [e.g. points and straight lines in geometry] and correspondingly several kinds of individual variables running over these different types. Practically nothing is changed by that.

33·2

The only things that are changed are the following:

1. The definition of term $f(a_1, \ldots, a_n)$ is now a term only if $a_1, \ldots, a_n$ a⟦re⟧ terms of certain specified types (determined by f) and this whole expression is then a term of a definite type determined by f.

2. In the formal axioms of quantification we have now the restriction that t must be a term of the same type as x.

Everything else including all theorems I am going to prove remain literally the same.

If we want to build up the calculus of functions abstractly we would have the following primitive terms: propositional function, term, var⟦iable⟧ ($\subseteq$ ⟦term⟧), ass⟦ertable⟧ propositional function, W, the five logical operations, the operation of substitution $A\binom{x}{t}$ and we would have to assume in addition to these axioms several axioms about substitution, namely distributivity resp. commutativity with respect to logical operations $(A \to B)\binom{x}{t} = A\binom{x}{t} \to B\binom{x}{t}$ and $[(x)A]\binom{x}{t} = (x)A$ which expresses the fact that[21]

34·3

An atomic expression would be defined as one which is neither and for certain propositions we would need the axiom that any propositional function can be obtained ...

It is easily seen that all these axioms are satisfied in the model described if identities are taken as assertable propositional functions.

37

54. *Theorem.* If you confine yourself to such expressions as contain only the logical operations $\to$, $\cdot$, (x) (hence also $\neg$) (but do not contain $\vee$, $\exists$) then classical and intuitionistic logic become identical, i.e. every classical identity not containing $\vee$ and $\exists$ is an intuitionistic identity. Now these notions in classical logic are sufficient to define the others, since

$$(\exists) \equiv \neg(x)\neg \quad p \vee q \equiv \neg(\neg p \cdot \neg q)$$

55. Proof ⟦is⟧ very easy: we prove that $\neg\neg A \to A$ can be asserted for any formula A not containing $\vee$, $\exists$. Proof by induction on the number of logical symbols in A:

1. If $= 0$ then either A is atomic or $= W$. In the first case theorem is true by the axioms, in the second: $\neg\neg W \to W$ or $((W \to W) \to W) \to W$ since $W \to W$ is assertable and generally, if A ⟦is⟧ assertable and B ⟦is⟧ arbitrary then $(A \to B) \to B$ because

$A \to [(A \to B) \to B]$	proved
Theorem	syllogism

2. If > 0, A must have either one of the following forms:

1. $A = B \to C$
2. $A = B \cdot C$

[21] The backside of page 33·2 is numbered 34·3; there are no pages 34·1 and 34·2. This is probably Gödel's mistake and the following page should have been numbered 33·3.

3. $A = (x)B$ for some variable x

where B, C contain fewer logical symbols ⟦than A⟧

38

and by induction we know $\neg\neg B \rightarrow B, \neg\neg C \rightarrow C$ can be asserted. I have to show that the following expressions also be assertable:

$$\neg\neg(B \rightarrow C) \rightarrow (B \rightarrow C)$$
$$\neg\neg(B \cdot C) \rightarrow B \cdot C$$
$$\neg\neg(x)B \rightarrow (x)B$$

But before we proved the following distributivities for $\neg\neg$:

1.	$\neg\neg(B \rightarrow C) \rightarrow (\neg\neg B \rightarrow \neg\neg C)$	
	$B \rightarrow \neg\neg B$	double negation
	$(\neg\neg B \rightarrow \neg\neg C) \rightarrow (B \rightarrow \neg\neg C)$	multiplication $\rightarrow \neg\neg C$
	$\neg\neg(B \rightarrow C) \rightarrow (B \rightarrow \neg\neg C)$	
	$\neg\neg C \rightarrow C$	inductive assumption
	$(B \rightarrow \neg\neg C) \rightarrow (B \rightarrow C)$	multiplication $B \rightarrow$
	$\neg\neg(B \rightarrow C) \rightarrow (B \rightarrow C)$	syllogism
2.	$\neg\neg(B \cdot C) \rightarrow (\neg\neg B \cdot \neg\neg C)$	
	$\neg\neg B \cdot \neg\neg C \rightarrow B \cdot C$	Leibniz
	$\neg\neg(B \cdot C) \rightarrow B \cdot C$	syllogism
3.	[(1) $\neg\neg B \rightarrow B$	assumption]
	$(x)B \rightarrow B$	
	(2) $\neg\neg(x)B \rightarrow \neg\neg B$	transposition 2. kind
	$\neg\neg(x)B \rightarrow B$	syllogism (1)(2)
	$\neg\neg(x)B \rightarrow (x)B$	rule of universal quantification

These distributivities ⟦are⟧ not true for $\vee$: $\neg\neg(A \vee B) \rightarrow \neg\neg A \vee \neg\neg B$ ⟦does not hold⟧.

So if we understand by “propositional function” only a propositional function not containing $\vee$ $\exists$ then ⟦we⟧ can assert 1.) of course all former axioms (excluding those containing $\vee$ $\exists$), 2.) $\neg\neg A \rightarrow A$[22] for every propositional function. But this is exactly a system of axioms for the classic system of the calculus of functions (for expressions not containing $\vee, \exists$).

[22] Gödel has mistakenly written $A \rightarrow \neg\neg A$.

This theorem is very surprising and it becomes still more so if we consider what axioms were really used in its proof. 1.) We have not used axioms for $\vee$ and $\exists$; 2.) we have not used $W \rightarrow A$ (as you can easily check). I.e. we have used

1. Positive logic:

   $A \rightarrow A \cdot A$
   $A \cdot B \rightarrow A$
   $A \cdot B \rightarrow B \cdot A$
   The rules of export and import, syllogism & implication.

2. The axiom and rule for the universal quantifier.

So these few and apparently constructive axioms suffice to deduce ⟦in a⟧ sense the whole classical non-constructive logic.

39.1[23]

Last time I proved a theorem about the relationship between ordinary and intuitionistic calculus of functions under the assumption that the atomic expressions are decidable for any given argument. And this theorem was the following: Every identity of the classical calculus not containing $\vee$ or $\exists$ is an identity of the intuitionistic calculus. And furthermore we had the following corollary: If you have two theories, one intuitionistic and one classical, with the same specific axioms and if furthermore

1. The primitive relations of the theories are decidable and the primitive functions calculable;

2. For every axiom containing $\exists x,\ p \vee q$ also the corresponding theorem with $\neg(x)\neg,\ \neg(\neg p \cdot \neg q)$ holds in the intuitionistic theory. Then

Every *theorem* of the classical theory which doesn't contain $\exists$ ⟦or⟧ $\vee$ holds also in the intuitionistic theory. If the theorem under consideration contains defined terms then of course the requirement is that in their definitions no $\vee$ and $\exists$ occurs.

Examples of theories which satisfy the hypothesis are e.g. number theory or algebra

[23] The pages 39.1-39.3 were apparently ripped off from the first notebook and stacked between pages 63^{iv} and 64 of the second one. They appear to be an alternative continuation for p. 38.

39.2

(in the sense of the theory of algebraic equations) or also metamathematics. Theories which do not satisfy the hypothesis are analysis or set theory. They don't satisfy the hypothesis because the specific axioms are not the same in classical and intuitionistic analysis ⟦or set theory⟧. Classical analysis and set theory admit impredicative definitions which means that ⟦they⟧ assume certain existential axioms about real numbers and sets, which are rejected by the intuitionists.

But the interesting fact is that the difference between classical and intuitionistic analysis doesn't arise from the rejection of the law of excluded middle and non-constructive existential assertions, but from the impredicative definitions which were rejected already by the half-intuitionists (e.g. Poincaré and Borel). [So it turns out that the restrictions which Brouwer puts on classical methods of proof don't go beyond those of the half-intuitionists as far as the formalism is concerned.]

39.3

As to the question whether the intuitionistic ⟦theories⟧ actually always satisfy this requirement that the primitive functions must be calculable and the primitive relations decidable, I am afraid the answer must be no since Heyting uses the set-theoretical ε-relation as a primitive and a set in the intuitionistic sense (which is called species) may contain an arbitrary series of quantifiers in its definition. But this does not disprove my statement that you always can analyse so far as to obtain decidable relations, because if you have a definite species then it is defined by some propositional functions, and I think it turns out that the atomic expressions in this proposition are decidable if you analyse far enough. And this makes it possible to define the non-decidable ε-relation in terms of the decidable one (so that you have to take only the latter as primitive).

By the way, from my theorem it would follow that if you confine yourself to species in whose definition no $\vee$ ⟦or⟧ $\exists$ occurs, you can assert $\neg\neg(x\,\varepsilon\,y) \rightarrow (x\,\varepsilon\,y)$.

39

56. This then has the consequence that also for certain theories containing specific axioms it is true that if a theorem doesn't contain the existential quantifier

and $\vee$, then it holds intuitionistically whenever it can be proved in the corresponding classical theory where by "corresponding" I mean a theory with the same specific axioms. This evidently holds under the following assumptions about the theory conc⟦erned⟧: If the specific axioms contain $\vee$ and $\exists$ then proposition obtained by replacing $p \vee q$ by $\neg(\neg p \cdot \neg q)$ ⟦and⟧ $(\exists x)$ by $\neg(x)\neg$ must hold in the intuitionistic theory and this ⟦is⟧ practically always satisfied. Counterexample:[24] an axiom $\neg(x)(\exists y)R(x,y)$, $\quad \neg(x)\neg(y)\neg R(x,y)$ doesn't follow, $(\exists x)(y)\neg R(x,y)$ doesn't follow either.

57. This gives then of course an intuitionistic proof for the freedom from contradiction for the classical theory under consideration, e.g. number theory.

But such a consistency proof is of no great value as long as no satisf⟦ying⟧ i.e. really constructive meaning for the primitive symbols of intuitionistic logic is given.

58. Now I wish to mention a second theorem connecting classical and intuitionistic logics which holds under the same assumptions. In order to formulate it I have to define a stronger (i.e. more constructive) kind of negation than absurdity. ~~Take a proposition of the form $(x)A$. Then~~

~~40~~[25]

the statement $(\exists x)\neg A$ evidently is also a kind of negation of A but stronger than $\neg(x)A$ because

$$(\exists x)\neg A \to \neg(x)A$$

can be asserted but not vice versa [$\neg\neg$ of the inverse means $(x)\neg\neg A \to \neg\neg(x)A$].

59. Now take a more complicated statement $(x)(\exists y)A$. How negation to be defined here:

$$(\exists x)(y)\neg A \text{ and this is } \neq \neg(x)(\exists y)A$$

and in general you see the constr⟦uction⟧ is to be defined by shifting the sign of absurdity as far inside as possible. By classical logic you can always shift a negation over quantifiers:

---

[24] This addition at the bottom of the page is written in shorthand.

[25] The first page numbered 40 is cancelled. On the bottom of the cancelled page there is a shorthand addition: is the positive form really the strongest? ⟦Ist die positive Form wirklich die stärkste?⟧ An uncancelled page 40 and a shorthand list of "maxims" begin the second notebook. They're followed by a fully cancelled page 41 and then a new page 41. We give the cancelled pages 40–41 first.

$$\neg(x) \ : \ (\exists x)\neg \qquad \neg(\exists x) \ : \ (x)\neg$$

But in classical logic you can shift the negation also over the other logical symbols $\neg \ \vee \ \cdot$, namely:

$\neg(p \vee q) \equiv (\neg p) \cdot (\neg q)$

$\neg(p \cdot q) \equiv \neg p \vee \neg q$

$\neg(p \rightarrow q) \equiv p \cdot \neg q$

~~41~~

But iterated application of this procedure you can always accomplish that symbols of negation apply only to atomic formulas.

The uniquely determined formula obtained from a formula A in this manner I denote by $\overline{A}$. Then $\overline{A} \ \equiv \ A$ in classical logic and in intuitionistic logic we have $\overline{A} \ \rightarrow \ A$ can be asserted (but not vice versa). So $\overline{A}$ is the intuitionistically strongest formula which is classically equivalent to A under these rules of shifting the negation. I don't prove this.

60. Now constructive negation of course defined by $\sim A \ = \ \overline{\neg A}$ and now the second theorem connects classical and intuitionistic logic reads like this: if A is classically provable then $\neg\sim A$ is intuitionistically demonstrable. So not $\neg\neg A$ but $\neg\sim A$.

Notebook 2

[40′]

Maxims Lectures[26]

0. Don't write down every word but *only the framework* (do not get distracted by the upcoming ideas).

1. Before writing further, *read through what was done the day before* and check the program for what needs to be done, then write 5 more pages (*perhaps after planning roughly beforehand*).

2. *Control all hours and breaks.*

40

Let us call an expression positive if no symbol of $\rightarrow$ or $\neg$ occurs outside of a quantifier or to be more exact: An expression is called positive if it is obtained from unquantified expressions by sole application of the operations (x), $(\exists x)$, $\vee$, $\cdot$. In particular any normal form is a positive formula. In classical logic there exists for every expression a uniquely determined positive expression which is equivalent to it. It is obtained by first replacing $p \rightarrow q$ by $\neg p \vee q$ and the shifting the negation as far to the inside as possible using

41

De Morgan form and the formulas:

$$\neg(x) \equiv (\exists x)\neg \quad \neg(\exists x) \equiv (x)\neg$$

Let's denote the positive expressions obtained in this way from A by $\overline{A}$. E.g.

26

Max. Vorl.

0. Nicht jedes Wort schreiben, sondern nur das Gerüst (nicht durch das Eintragen von Einfällen abhalten lassen).

1. Vor Beginn des Weiterschreibens das durchlesen, was am Tag vorher gemacht und Nachsehen des Programms, was nun zu machen ist und dann 5 Seiten weiterarbeiten (eventuell nachdem vorher ungefähr überlegt).

2. Alle Stunde kontrollieren und Ruhepause.

$$A = (x)(\exists y)R(x,y) \rightarrow \neg(x)K(x)$$
$$\overline{A} = (\exists x)(y)\neg R(x,y) \vee (\exists x)\neg K(x)$$

Then $\overline{A} \equiv A$ ⟦holds⟧ classically but not intuitionistically. $\overline{A}$ is in a sense the most constructive statement equivalent with A under these rules (of shifting the negation) i.e. $\overline{A} \rightarrow A$ can be asserted intuitionistically for any propositional function. And now I define constructive negation of A by $\sim A = \overline{\neg A}$ e.g. $\sim(x)K(x)$ would be $(\exists x)\neg K(x) \not\equiv \neg(x)K(x)$. (Applied to positive expressions constructive negation has a simple meaning, namely dualising i.e. replacing (x) by $(\exists x)$ and vice versa, and $\cdot$ by $\vee$ and

42

negating the unquantified expressions. Hence it is clear that ⟦the constructive⟧ negation of a positive expression is again positive.)

And now one can prove this: If A is classically demonstrable then $\neg\sim A$ can be asserted in intuitionistic logic (you see ⟦that⟧ $\neg\sim A$ is a weakening of $\neg\neg A$ since $\sim A$ is stronger than $\neg A$.)

Proof: Auxiliary theorem. Let A be a positive expression and let us denote by A' the expression obtained from A by replacing $(\exists x)$ by $\neg(x)\neg$ and $p \vee q$ by $\neg(\neg p \cdot \neg q)$ and leaving it unchanged otherwise. Then $A \rightarrow A'$, A' defined by

$$\begin{array}{lcll} A' & = & A & \text{[[for } A \text{ atomic]]} \\ ((\exists x)A)' & = & \neg(x)\neg(A') & \\ ((x)A)' & = & (x)A' & \\ (A \cdot B)' & = & A' \cdot B' & \\ (A \vee B)' & = & \neg(\neg A' \cdot \neg B') & \end{array}$$

[42′][27]

Let[28] A_p be the positive expression that belongs to A. Then $A_p \rightarrow A$ holds. Let N_p be the positive expression that belongs to $\neg A$. So then $N_p \rightarrow \neg A$ holds.

[27] This passage on the bottom of the page right to p. 41 is written in shorthand German.

[28] Sei A_p der zu A gehöriger positiver Ausdruck, so gilt $A_p \rightarrow A$.

Sei N_p zu $\neg A$ gehöriger positiver Ausdruck, also $N_p \rightarrow \neg A$.

Dann gilt: von den beiden Ausdrücken $\neg A_p$ $\neg N_p$ ist mindestens einer wahr (in der später gegebenen Realisierung).

A bedeutet: jede reelle Zahl ist konstruierbar, also gilt $\neg A_p$. Könnte man beweisen, dass nicht $\neg N_p$ gilt, so wäre gezeigt, dass A falsch ist $\quad N_p \rightarrow \neg A_p \quad A_p \rightarrow \neg N_p$.

Then it holds: At least one of the expressions $\neg A_p, \neg N_p$ is true (in the realization given later).

A means: each real number is constructible so $\neg A_p$ holds. If one could prove that $\neg N_p$ does not hold, then it would be shown that A is false:

$$N_p \rightarrow \neg A_p \quad A_p \rightarrow \neg N_p$$

43

But now the theorem to be proved follows also im⟦mediately⟧ by recursion on the number of logical symbols occurring in A.

1. $A \rightarrow A'$ ⟦is⟧ assertable for matrices.

2. Assume $A \rightarrow A'$ ⟦is⟧ true. ⟦Then this is⟧ true also for $(\exists x)A$.

⟦Therefore⟧ $(\exists x)A \rightarrow [(\exists x)A]'$ ⟦and we get⟧ from ⟦the inductive⟧ assumption $(\exists x)A \rightarrow (\exists x)A'$. Hence $(\exists x)A \rightarrow \neg(x)\neg A'$ because $(\exists x) \quad \rightarrow \quad \neg(x)\neg$.

But $[(\exists x)A]' = \neg(x)\neg A'$. Hence the theorem ⟦follows⟧ and in the same way ⟦this can be⟧ proved for other three cases using certain formulas we proved earlier.

44

But now from this lemma the theorem to be proved follows immediately.

Assume A is classically demonstrable. Then also $\neg(\overline{\neg A})'$ is classically demonstrable because $\overline{}$ and $'$ are a both operations which give equivalent expressions in classical logic. But now this expression contains no $\vee$ and $\exists$ because they have been eliminated according to the definition of $'$. Hence $\neg(\overline{\neg A})'$ is also intuitionistically demonstrable by the previous theorem.

But $\neg(\overline{\neg A})' \rightarrow \neg(\overline{\neg A})$ since $\overline{\neg A} \rightarrow (\overline{\neg A})'$ by lemma since $(\overline{\neg A})'$ is a positive expression and we have only to apply transposition. But $\neg(\overline{\neg A})'$ is exactly $\neg(\sim A)$ by the definition of $\sim$.

This theorem also holds for any theory with

45

any additional specific axioms (under the same assumption we had before). We need this assumption because we applied the former theorem and all the other steps of the proof go through without any assumption about the the theory.

45.1

With this I am concluding the axiomatic treatment of intuitionistic logic. The results obtained have been pretty much surprising in so far as they show that in a sense the whole classical logic is contained in the intuitionistic logic. Of course it is contained only formally i.e. the same formulas can be proved but the meaning of these formulas is completely different (e.g. $\neg(x)\varphi(x)$ ⟦and⟧ $\sim(x)\varphi(x)$). But this difference of meaning makes the result still more surprising since this means that the non-constructive classical logic has a constructive interpretation. And this makes one doubtful whether intuitionistic logic really is constructive or if not perhaps some non-constructive elements are hidden in the axioms, which is quite possible regarding the great complicatedness in the primitive terms. I hope the interpretation which I am going to give will help to decide this question in favor of intuitionistic logic at least in the case when this logic is applied in theories with decidable primitive notions as e.g. number theory or algebra. Namely, it turns out that in this case the primitive terms of intuitionistic logic can be defined in terms of a system which is constructive in a more precise and stronger sense; namely this system it will satisfy the following requirements:

1. The operations of the calculus of propositions are applied only to decidable statements in which case there is no question as to their meaning and in which case classical logic doubtlessly holds. So in particular, $\neg$, $\rightarrow$, etc. is never applied to propositions containing quantifiers because the quantifiers destroy the decidability, but only to *unquantified* expressions.

46

2. No existential quantifiers at all occur i.e. mere existential propositions cannot at all be pronounced but only the underlying constructions can be pronounced.

3. Of course the primitive functions will be calculable and the primitive relations decidable.

So every proposition in such a system looks like this: It is an unquantified expression $M(x, y, \ldots a, b, \ldots)$ containing certain variables $x, y, \ldots$ and certain constants $a, b, \ldots$ and its assertion means this: If any objects x, y falling under the range of the respective variables are given and if I calculate the function applied to them in this expression and then decide the truth or falsehood of the atomic propositions in this expression, then $M(x, y, \ldots a, b, \ldots)$ turns out to be true. Recursive number theory, e.g., is such a system.

47

Existential quantification can be introduced as mere abbreviation, if one wishes to, by the following rules: If t is any constant term (i.e., one which contains no variables but only constant functions and individuals) and if some expression $A(t)$ (containing t) can be asserted, then $(\exists x)A(x)$ can be asserted, and this will be the *only* rule for existential quantification which we allow. In particular, existential assumptions never occur as premisses in any inference (nothing can be concluded from them).

Hence such a system is trivially constructive in the sense defined in my first lecture: If $(\exists x)A(x)$ can be proved then it can only have been obtained by an application of the rule just stated and hence the last but one formula of the proof $A(x)$ gives the construction of such an x. So existential quantification is a mere abbreviation in such a system.

48

And now in a system Σ of this kind I explain the meaning of the logical operations as applied to a proposition of this same system Σ, and the result will again be a proposition of Σ, and since ⟦Σ⟧ comprises the whole recursive number theory this gives, in particular, a definition of the logical terms as applied in number theory.

The individuals of this system Σ are divided into an infinite number of different types and I have first to make some preliminary cons⟦iderations⟧ about these types.

The lowest type consists of the non-negative integers. I denote this type by I. The other types are defined by the following recursive stipulation.

If t_1, t_2 are any types defined already then $t_1 \tau t_2$ is the type of functions whose argument is of type t_2 and whose value is of type t_1 e.g.

$I \tau I$ = type of functions of integers whose values are again integers.

$(I \tau I)\tau(I \tau I)$ = type of functions whose arguments are functions of integers and whose values are likewise functions of integers.

49

I don't introduce types for functions with several arguments but am treating functions with several arguments in Church's way i.e. $f(x,y)$ is replaced by a function with one argument g by $(g(x))(y) = f(y)$.

The types are divided into levels in the following way:

1. $I \,\varepsilon\, L_0$ and only $I \,\varepsilon\, L_0$.

2. $t_1 \tau t_2 \,\varepsilon\, L_{k+1}$ then and only then if both $t_1, t_2 \,\varepsilon\, L_0 + \cdots + L_k$ and at least one of the two types t_1, t_2 belong to L_k.

The levels are mutually exclusive, $L_k \cdot L_s = 0$ ⟦when⟧ $k \neq s$ and every level except the zero'th and the first comprise more than one type.

To each type t (except I) belongs a certain value type $V(t)$[29] namely

$$V(t) = t_1$$

if $t = t_1 \tau t_2$ and a certain argument type

$$Arg(t) = t_2.$$

50

Evidently the value type and argument type always belong to a lower level than T itself.

Now the undefined symbols of the system Σ are as follows:

1. For each type we have an infinity of variables belonging to this type (denoted by Latin letters $x, y, \ldots a, b, \ldots F, G, \ldots$) and

2. An infinity of constants denoted by Greek letters (φ, χ, Φ, etc.).

We assume that the letters belonging to different types be distinguished in some way e.g. by superscripts marking the type. In addition to Greek letters the symbol 0 denotes a constant of type I.

The symbols introduced so far are nothing else but what I called formerly the individual variables and individual constants. ⟦Addition: ν (= successor)⟧

[29] Here Gödel has written $W(T)$ but thereafter uses $V(t)$ instead.

3. One primitive binary relation $=$ (identity).

4. One primitive function "application," namely application of a function to its argument. No specific symbol for it

51

is introduced but application [[is]] denoted in the usual way by brackets [[where]] $F(a)$ means "the result of applying the function F to the object a."

5. We have the symbols of the calculus of propositions $\sim, \cdot, \vee, \supset, \equiv$ which are the classical notions since they are applied only to decidable propositions, namely unquantified expressions. Therefore I denote them by other symbols. The corresponding intuitionistic [[symbols]] I shall denote by $\neg, \&$, v, $\rightarrow, \leftrightarrows$. They don't occur in the system Σ.

It is clear how *the expressions of this system* are to be built up.

1. I define what *a term of type t* is.

i. Every constant or variable of type t is a term of type t.

ii. If a is a term of type t_2 and b is a term of type t_1, $b(a)$ is a term of type $t_1 \tau t_2$.

I wish to remark without proof that this way of writing, i.e., putting only the argument in brackets and not the function, also if the function is itself a composite expression, is sufficient to avoid ambiguities (no further brackets necessary).

A term is called "constant" if it contains no variables, i.e., only Greek letters.

52

2. If A, B[30] are any terms of type I, then $A = B$ is an atomic expression or *prime formula*. It is quite essential that identity is only applied to terms of type I, i.e., integers, because in order to satisfy the requirement of constructivity I

[30] Here the terms A, B have originally been written in lower case, then corrected in formulas to upper case. In a few places, Gödel has not written over the lowercase letters, but as the distinction does not appear to carry any meaning, we have changed them all to upper case.

enumerated [[last]] time, we must have decidable primitive relations. But $f = g$ is not in general decidable. E.g. for functions of integers f, g because this means $(x)(f(x) = g(x))$, and even if f and g are calculable for any argument, you may not be able to decide whether or not for *all* integers $f(x) = g(x)$.

It is to be noted that the symbol of equality as used here is a metamathematical operation which produces expressions out of other expressions (similarly as $\supset$, $\vee$, etc). If used in this sense I put a dot above it. So $A \doteq B$ means: the expression obtained from the expressions A, B by joining them by a sign of equality. So this dotted equality is an operation (performed on expressions) whereas undotted = is a relation between expressions (namely the identity relation). E.g. we can state $A\binom{x}{t} = A$ if x [[is]] not contained in A. So this was the definition of an atomic expression.

3. A *propositional function* or *expression* in the system Σ is anything obtainable from prime formulas by the operations of the calculus of propositions only.

$52'$

Last time I described a certain system of types where the lowest type I is formed by the integers and every other type consists of the function with one argument whose argument and value are of certain given other types. Then I began to describe a formal system that I called Σ. The primitive symbols of this system were the following:

1. For each type an infinity of variables belonging to this type (denoted by Latin letters $x, y, F, \ldots$).
2. For each type an infinity of constants belonging to this type (denoted by Greek letters, one constant [[denoted]] by the symbol 0).
3. The symbol of $=$ (which plays the role of the only primitive relation of this system).
4. Brackets denoting the operation of application a which is the only primitive function of this system.
5. The operations of the calculus of propositions $\sim, \cdot, \vee, \supset, \equiv$. They are applied only to unquantified expressions where they coincide with the classical notions. Therefore I denote them in a different manner.

52″

These are all the primitive symbols of the system Σ (so in particular we have no quantifiers).

I have defined already what a term of the system Σ is or rather what a term of a given type is; namely

1. Variables and constants of type t are terms of type t.

2. If B is a term of type t_2 and A a term of type $t_1 \tau t_2$, then $A(B)$ is a term of type t_1.

The next thing to do is to define what an *expression or propositional function* of Σ is.

53

If A is a term of a type $\neq I$ and B_1 a term whose type is the argument type of A, then $A(B_1)$ is again a term, whose type however belongs to a lower level than A because it is the value type of A. If it is not yet the type I, one can iterate this procedure and form $A(B_1)(B_2)$, where B_2 is an (appropriate) argument for $A(B_1)$. After a finite number of steps the expression must become of type I because the level decreases with every step. So for every term A there exists a series of terms $B_1, \dots, B_n$ such that if I write this series in brackets behind A I obtain a term of type I. I call such a series (which is uniquely determined as to the type of its members) a *complete argument series* for A. If A, B are two terms of the same type but $\neq I$, then $A \doteq B$ has no meaning so far, but I define it now to denote the expression:

$$A(\mathfrak{x}) \doteq B(\mathfrak{x})$$

where $\mathfrak{x}$ is a complete argument series consisting of variables different from each other and those in A, B.

54

Now to the axioms and rules of Σ. There are three groups:

1. *Logical axioms.* Every expression obtained by taking an identity of the classical calculus of propositions and substituting arbitrary expressions in place of

the propositional variables is a formal axiom. The expressions substituted may of course contain variables since assertion of a propositional function means assertion of every proposition obtained.

II. *Mathematical axioms.* They concern the constants which we denote by Greek letters.

1. I choose the letter ν to denote the successor function for integers (a function of type $I\tau I$) and we have the axioms

   $$\left\{\begin{array}{c} \nu(x) \doteq \nu(y) \supset x \doteq y \\ \sim(\nu(x) \doteq 0) \end{array}\right\} \quad \text{Peano}$$

   Thus ν gives a unique notation for the integers $0, \nu(0), \nu(\nu(0)), \ldots$. The terms of this sequence I call *numerals*.

2. The other Greek letters will denote functions which can be defined in terms of ν either explicitly or by recursion.

A. We admit the following schemes of explicit definition: If φ is the function to be defined, then the definition looks like this:

   $$\varphi(\mathfrak{x}) \doteq A$$

   where $\mathfrak{x}$ is a complete

55

argument series for φ consisting of variables all different from each other and where A is an arbitrary term of type I which contains no variables except at most those of the series $\mathfrak{x}$ and in addition to those variables contains only previously defined constants.

B. Now a recursive definition of a function φ: Here we have to suppose that the argument type of φ is I, the value type is arbitrary, and the definition looks like this

   $$\begin{array}{ll} \varphi(0)(\mathfrak{x}) \doteq A & \mathfrak{x} \\ \varphi(\nu(x))(\mathfrak{x}) \doteq B & \mathfrak{x}, x \end{array}$$

   where $\mathfrak{x}$ is a complete argument series for $\varphi(0)$ consisting of variables different from each other and from x, and A, B are terms containing no other variables but the first $\mathfrak{x}$, the second $\mathfrak{x}$, x, and A contains only previously defined constants, and B in addition to previously defined constants the letter φ but only in the combination $\varphi(x)$ (by $x, \mathfrak{x}$).

56

I wish to remark that it is not necessary to assume this scheme of definition in all its generality, but you can confine yourself to introducing by recursion the function $P(n, f)$ which means f^n in the usual ⟦sense⟧ (raising to the n^{th} power) and then all functions definable recursively by this scheme can be defined explicitly in terms of these constants P. Also in the scheme of explicit definition one could confine oneself to certain special cases. But I don't need this reduction at present.

If I say that these explicit and these recursive definitions are the axioms, I mean more exactly the following:

I enumerate in some way all constant functions definable by these schemes, where functions defined in a different way are to be considered as different functions (even if they are extensionally the same function), and then you can associate with each such constant a definite symbol (say a Greek letter

⟦57⟧

with a certain superscript). Then all these infinitely many defining equalities [written in the corresponding Greek letters with subscripts] are to be axioms of the system.

So the following metamathematical theorem will be true given any term A containing no other variables besides those of the series $\mathfrak{x}$. Then you can determine a constant φ such that $\varphi(\mathfrak{x}) \doteq A$ can be proved in Σ. This holds also if A is not of the type zero owing to the definition of $\doteq$, so this case (namely it means $\varphi(\mathfrak{x})(\mathfrak{y}) \doteq A(\mathfrak{y})$ for a certain sequence of variables $\mathfrak{y}$, but this falls under the scheme assumed). Same remark applies to recursive definition. Of course the same function, e.g. $+$, may have infinitely many letters denoting it.

III. Last group of axioms concerns equality:

1. $x \doteq x$ for some variable x
2. $x \doteq y \supset (A\binom{z}{x} \equiv A\binom{z}{y})$ for any expression A
   implies $\supset T\binom{z}{x} \doteq T\binom{z}{y}$

This is a formal axiom for any expression A and any variables x, y, z.

Rules of inference

I. *Logical* $A \quad A \supset B \quad / \quad B$

Substitution $A \quad / \quad A\binom{x}{T}$

⟦where⟧ T ⟦is⟧ any term of same type as x containing perhaps variables.

Together with the first group of axioms, this implies that all rules of inference of the classical calculus of propositions can be derived, because always $A \supset B$ is an identity ⟦when $B = A\binom{x}{T}$⟧.

II. *Mathematical* If A is any expression and x any variable, then

$$A\binom{x}{0} \quad A \supset A\binom{x}{\nu(x)} \quad / \quad A$$

There is a stronger principle of induction which however can be derived from this which reads like this: If $x_1, \ldots, x_n$ are any variables $\neq x$ and $T_1, \ldots, T_n$ any terms of the same types, respectively, then from

$$A\binom{x}{0} \quad A\binom{x_1, \ldots, x_n}{T_1, \ldots, T_n} \supset A\binom{x}{\nu(x)}$$

you can conclude A. The terms ⟦are⟧ quite arbitrary.

I am not interested now in this derivation but shall assume this stronger rule as an axiom and wish only to give the intuitionistic reason for the correctness of this inference. Let's write A in the form $B(x, x_1, \ldots, x_n)$ marking the variables occurring in it. Then the second premiss means by transposition $\sim B(\nu(x), x_1, \ldots, x_n) \supset \sim B(x, T_1, \ldots, T_n)$ i.e. a counterexample for the proposition B to be proved allows you to derive another one with a smaller x, so after a finite number of steps, you obtain one with $x = 0$ which is excluded by the first premiss. It is clear that this inference is intuitionistically correct since the terms T_I are calculable.

58

III. The next rule no 3 concerns identity.

Let S, T ⟦be⟧ any terms of equal type and $\mathfrak{x}$ a complete argument series consisting of variables different from each other and the variables in S, T and let E be any expression containing a certain variable z of the same type as S and T. Then from

$$S(\mathfrak{x}) \doteq T(\mathfrak{x})$$

you conclude

$$E\binom{z}{S} \equiv E\binom{z}{T}$$

You see this is a certain *principle of extensionality*. If two functions S, T are extensionally the same then anything assertable about one also holds for the other. I need this principle in a little stronger form: Namely let P be any expression not containing any variable of $\mathfrak{x}$ but perhaps variables of S and T. Then [[addition: necessary because the definition by cases, see p. 91]]

$$\frac{P \supset S(\mathfrak{x}) \doteq T(\mathfrak{x})}{P \supset E\binom{z}{S} \equiv E\binom{z}{T}}$$

This implies

1. The above by $P = 0 \doteq 0$
2. $P \supset \mathfrak{x}\binom{z}{S} = \mathfrak{x}\binom{z}{T}$ $\quad$ $\mathfrak{x}$ a term of type I
3. $E\binom{z}{S} \equiv E\binom{z}{T}$

We cannot express this rule of inference by a formula in Σ because the premiss would have to contain bound variables (it says that for every German [[letter]] $\mathfrak{x}$ this holds but in Σ we cannot apply $\supset$ to formulas involving quantifiers).

59′

First I have to show that our scheme of definition includes definition by cases. The theorem which expresses this fact concerns the system Σ alone. All expressions and terms are supposed to be expressions respectively terms of Σ. It is the following:

Let T, S be two terms of type I containing no variables except those of the finite sequence $\mathfrak{x}$ and A a propositional function containing no other variables than $\mathfrak{x}$. Then you can find a constant ϱ for which the following two propositions are demonstrable in Σ:

$$A \supset \varrho(\mathfrak{x}) \doteq T$$

$$\sim A \supset \varrho(\mathfrak{x}) \doteq S$$

Proof: Namely at first you can find for any expression A a constant α such that

$$(\alpha(\mathfrak{x}) \doteq 1) \equiv A$$

$$(\alpha(\mathfrak{x}) \doteq 0) \equiv \sim A$$

is demonstrable in Σ.

Proof. 1. Define a constant function $\iota(xy) = {0 \atop 1} \equiv {\sim(x=y) \atop x=y}$ and functions representing the truth tables with 1 or T and 0 or F, e.g. a function $\beta(0) = 1, \beta(1) = 0, \beta', \beta$ etc., and now putting ι instead of $=$ and these β, β' instead of the logical operations, you obtain a term T for which the above is demonstrable, hence ⟦you obtain⟧ also a constant.

But by means of this α, ϱ can be defined as follows:

$$\varrho(\mathfrak{x}) = \alpha(\mathfrak{x}) \cdot T + \beta(\alpha(\mathfrak{x})) \cdot S$$

where $+ \cdot$ are ordinary addition and multiplication which can be defined recursively.

59

Now let us ask whether this system satisfies the requirements of strong constructivity laid down in the last but one lecture. The first two, namely that the operations of the calculus of propositions are applied only to expressions without quantifiers and that we have no existential quantifier are satisfied.[31] Now to the third proposition which says that the primitive relations are decidable and the primitive functions calculable. Now the only primitive relation = is evidently decidable for any two primitive objects to which it is applied since for any two numbers, you can decide whether they are of equal length or not. Hence it remains to be shown that all functions denoted by Greek letters are calculable. Now it is clear what it means that a function φ of

60

the first level is calculable. It means that for any number k, you can can find a number l such that $\varphi(k) = l$ is demonstrable. But what does it mean for functions of higher levels to be calculable? The natural definition which suggests itself is the following one:

A function F of any level is said to be calculable if for any complete argument series $\mathfrak{a}$ consisting of given calculable functions (and perhaps numerals), you can find a number k for which $F(\mathfrak{a}) = k$ is demonstrable. This definition

[31] There is an incomplete sentence written above this passage which seems to read "no quantif. and gen. ass. expr.".

of calculability is not circular because it presupposes the notion of calculability only for the argument types of F, i.e., for functions of a lower level. So it is an inductive definition of calculability (the induction going by numerical levels) ⟦which⟧ applies to system Σ. This definition means that we call a Greek letter calculable if for any complete series of arguments consisting of calculable Greek letters there exists a numeral, etc.

Now the function ν evidently is calculable in this sense and furthermore it can be proved

61

that in the two schemes of definition, we have only calculable functions in the definiens. The function defined will likewise be calculable and therefore all constant functions of the system Σ denoted by Greek letters are calculable. I don't want to give this proof in more detail because it is of no great value for our purpose for the following reason. If you analyze this proof it turns out that it makes use of the logical axioms also for expressions containing quantifiers and since it is exactly these axioms which we want to deduce from the system Σ.

~~So our attitude must be this that the axioms of Σ (in part⟦icular⟧ the schemes of definition) must be admitted as constructive without proof and it is shown that~~

62

~~the axioms of intuitionistic logics can be deduced from them with suitable definitions. This so it seems to me is a program~~[32]

63^i

Last time I set up the axioms and rules of inference of a certain formal system Σ in terms of which I want to interpret intuitionistic logic. Three groups of axioms:

1. All axioms of the classical calculus of propositions (because there are no quantifiers)

[32] Page 63 has been ripped off the notebook and inserted between pages 62 and 63^i; however, the pages 63^i to 63^{iv} seem to rather belong here, as the text on p. 63 continues on p. 64 and not page 63^i. Page 62 contains, a few lines below the cancelled passage, two new paragraphs which we have moved below p. 63^{iv}.

2. Two of Peano's axioms for the successor function ν and the defining axioms for all other constant functions.

3. The two axioms of the equality sign.

Three groups of rules of inference:

1. Rule of implication and rule of substitution.

2. Rule of complete induction and

3. Rule of extensionality which I formulated in the following manner: Let S and T be any terms of the same type and $\mathfrak{x}$ a complete argument sequence for S and T consisting of variables ⟦of appropriate types⟧ and let A be an expression not containing any variables of $\mathfrak{x}$ but perhaps variables of S and T.

Then from

$$A \supset S(\mathfrak{x}) \doteq T(\mathfrak{x})$$

we can infer

$$A \supset E(\tfrac{z}{S}) = E(\tfrac{z}{T})$$

where E is any expression.

The intuitionistic meaning of adding this hypothesis A is the following. Let $x_1, \ldots, x_n$ be all the variables contained in A. Then these variables may occur also in S and T

63[ii]

but not in $\mathfrak{x}$. The conclusion means that this holds for any constant $a_1, \ldots, a_n$ put in place of the $x_1, \ldots, x_n$ but this is clear because if a_i are such that they make $A(a_1, \ldots, a_n)$ false this implication holds. But if the a_i make A true, then substitute a_i in the premiss. Then $S_a(\mathfrak{x}) = T_a(\mathfrak{x})$ for every $\mathfrak{x}$, hence the conclusion by the ordinary principle of extensionality.

The following two rules are immediate consequences

$$A \supset P(\tfrac{z}{S}) = P(\tfrac{z}{T})$$

Furthermore by taking $A = (0 \doteq 0)$ we get the rules of extensionality in their usual form and furthermore also the following rule

$$\frac{S(\mathfrak{x}) = T(\mathfrak{x}), \quad \varphi(\tfrac{z}{S})}{\varphi(\tfrac{z}{T})}$$

63^{iii}

Now to the question whether this system satisfies the requirements of constructivity laid down in the last but one lecture it comes to this: We have to show that every atomic expression containing no variable is decidable i.e. $T_1 \doteq T_2$ (where T_1, T_2 are of type I is decidable). For this purpose it is sufficient to show that for any constant terms of type I, there exists a numeral n such that $T \doteq n$ is demonstrable. Now it is not difficult to prove that making use of intuitionistic logic, but this proof is of no particular value for us, because we want to reduce intuitionistic logic to the system Σ. However, it seems to be possible to give another proof which makes use of transfinite induction up to certain ordinal (probably up to the first ε-number would be sufficient).

63^{iv}

Of course if you choose this course then the question arises in which manner to justify the inductive inference up to a certain ordinal number and one may perhaps be of the opinion that the axioms of Σ are simpler as a basis than this transfinite induction by which we want to justify them. Whatever the opinion to this question may be, in any case, it can be shown that intuitionistic logic, if applied in number theory (and also if applied in this whole system Σ) can be reduced to this system Σ. In order to accomplish this reduction, I must first introduce existential quantifiers in the manner described in the last but one lecture.

There[33] exists however another proof. Namely it is possible, instead of making use of the logical operators applied to quantified expressions, to use the calculus of the ordinal numbers (to be more exact of the ordinal numbers $< \varepsilon_0$) + ⟦and⟧ . .

I shall speak about this proof later on. The idea is the following: In order to show that every function is calculable it is sufficient to show that every constant term of type I can be transformed into a numeral by replacing in it

[33] This passage has been written on p. 62 below the cancelled incomplete passage.

63

successively all defined symbols by their definiens, and in order to show that this process of replacing comes to an end after a finite number of steps, you can associate an ordinal $< \varepsilon_0$ with each term and then show that this ordinal is diminished by every replacement.

I am first introducing *two new kinds of variables:*

existential denoted by $\overline{x}, \overline{y}$,

universal denoted by $\underline{x}, \underline{y}$.

I assume that to each of the former variables x corresponds exactly one existential $\overline{x}$ and $\underline{x}$ (hence infinitely many to each type).

The formal system obtained by the introduction of these variables I call $\overline{\Sigma}$ i.e. a *term* of $\overline{\Sigma}$ is defined by:

1. Every $x, \overline{x}, \underline{x}$ and every Greek letter α is a term.
2. If T, S are terms, then $T(S)$ ⟦is a term⟧ if S, T have

64

the appropriate types.

Propositional functions are obtained of the terms in exactly the same way as before.

I call expressions actually containing the new variables of second kind, and expressions not containing the new variables, i.e. belonging to Σ, of first kind. Expressions of the second kind are to be considered as propositional functions depending only on the free variables.

The possibility of denoting quantification in this manner (namely by two kinds of new variables $\underline{x}, \overline{x}$ without any specific symbols like $\exists$) is of course due to the fact that we want to admit only propositions of this special kind $(\exists x_1 \ldots x_m)(y_1 \ldots y_n)$ where all existential quantifiers precede all universal ones. So a propositional function $A(x, \overline{y}, \overline{z}, \underline{u}, \underline{v})$ containing besides free variables x also the new variables $\overline{y}, \overline{z}$ and $\underline{u}, \underline{v}$ means in ordinary notation this:

$$(\exists\, \overline{y}\, \overline{z})(\underline{u}\, \underline{v}) A(x, \overline{y}, \overline{z}, \underline{u}, \underline{v})$$

and that this $\underline{A}$ is asserted means in usual notation

$$(x)(\exists\, \overline{y}\, \overline{z})(\underline{u}\, \underline{v})A(\underline{x}, \overline{y}, \overline{z}, \underline{u}, \underline{v})$$

So far I have defined what meaningful expressions of $\overline{\Sigma}$ are; now what are the axioms:

Axioms

1. All former axioms and rules are assumed but only for propositional functions of the first kind (hence the same wording).

2. We have the rule of the existential quantification. It will be little more complicated than I explained in the informal exp⟦osition⟧ because the existential variable may be dominated by a universal variable x. In order to infer such a statement, the proposition from which we infer must have the following form: $A(x, t_1(x), t_2(x), u, v)$ i.e., the terms which are replaced by existential variables may contain the variable $\underline{x}$ but not the variables $\underline{u}$, $\underline{v}$, so that we obtain the following rule:

   Let A be an expression of the first kind containing the terms $T_1, \ldots, T_m$ and let $x_1, \ldots, x_n$ be any variables of A not occurring in these terms T_i. Then we can infer from

   65

   the expression obtained from A by replacing the terms $T_1, \ldots, T_m$ by existential variables $\overline{f}_1, \ldots, \overline{f}_m$ of the same types as T_i and different from each other, and by replacing x_i by $\underline{x}_i$. Then if A can be asserted so can B.

   ~~This rule is a little more complicated than the one I mentioned in the informal expl⟦anation⟧ in so far as the terms T_i which are replaced by existential variables $\overline{F_i}$ need not be constant terms but may contain variables in accordance with the fact that in the assertion the existential variables are in general dominated by universal variables. But in order to infer this proposition, the terms t_1, t_2 which you replace by existential variables must not depend on these variables u, v. Therefore the restriction that only such variables as do not occur in the terms t_1, t_2 may be turned into underlined variables.~~

This is the system $\overline{\Sigma}$. It is immediate that $\overline{\Sigma}$ is constructive[34] since rule 2 is

[34] Gödel has written "that Σ is constructive" where he certainly means $\overline{\Sigma}$.

the only one involving propositions of the second kind and it involves such propositions only in the conclusion. It is evident that a proposition of the second kind $A(x_1, \ldots, x_n, \overline{f}_1, \ldots, \overline{f}_m, \underline{y}_1, \ldots, \underline{y}_r)$ can be proved in $\overline{\Sigma}$ then and only then

66

if, for some terms $T_1, \ldots, T_m$ not contained in the variables $y_1, \ldots, y_r$, the expression $A(x_1, \ldots, x_n, T_1, \ldots, T_m, y_1, \ldots, y_r)$ can be proved in Σ. And this condition again is equivalent with the following: if

$$A(x_1, \ldots, x_n, \alpha_1(x_1, \ldots, x_n), \ldots, \alpha_m(x_1, \ldots, x_n), y_1, \ldots, y_r)$$

is demonstrable in Σ for some constants $\alpha_1, \ldots, \alpha_m$ because there exist constants in Σ satisfying the defining equalities

$$\alpha_1(x_1, \ldots, x_n) \doteq T_1$$
$$\vdots$$
$$\alpha_m(x_1, \ldots, x_n) \doteq T_m$$

To be more exact, if T_i should contain some variable different from $x_1, \ldots, x_n$, we form first terms T_i' by replacing these superfluous variables by arbitrary constants and then these are correct definitions with T_i' instead of T_i.

For $n = 0$ we obtain the following special case: $A(\overline{a_1}, \ldots, \overline{a_m}, \underline{y_1}, \ldots, \underline{y_r})$ is demonstrable in $\overline{\Sigma}$ if and only if there are constants $\alpha_1, \ldots, \alpha_n$ such that $A(\alpha_1, \ldots, \alpha_m, y_1, \ldots, y_r)$ is demonstrable in Σ.

Propositional functions differing only in the letters used

67

for the universal and existential variables I call congruent. Evidently congruent expressions are equivalent as to demonstrability (if one is demonstrable in $\overline{\Sigma}$ the other one is so too).

Furthermore, also expressions differing only: 1.) in the letters used for the existential variables and 2.) ~~of the arguments of some existential variables are equivalent as to demonstrability. I mean that the arguments are exactly the same terms in both expressions but their arrangement is different in both expressions, which of course may imply that the type of these existential variables is different in these expressions (I don't give the proof)~~

The operations of the calculus of propositions $\sim, \vee, \cdot, \supset$ can of course be applied also to expressions of the second kind in the sense of writing them beside each other, but the meaning is completely different from the usual one, e.g. it is possible that $A \supset B$ and A can be proved but B is false. But still some analogies subsist, e.g. if A, B then $A \cdot B$ (if all bound variables are different in A, B) or if $A \supset W$ can be asserted, then $\sim A$ and vice versa.

In the sequel, I shall have to consider very often not single expressions but sets of expressions (more exactly *finite ordered sets or sequences of expressions*). I shall denote them by German letters $\mathfrak{A}, \mathfrak{B}$, in particular sequences of variables by small

68

letters $\mathfrak{a}, \mathfrak{f}, \mathfrak{x}, \mathfrak{y}, \bar{\mathfrak{x}}, \bar{\mathfrak{y}}, \underline{\mathfrak{x}}, \underline{\mathfrak{y}}$ and sequences of constants by Greek letters α, β. The case of a sequence with one member and with 0 members $\wedge$ ⟦is⟧not excluded (one member = expression). I have to make use of several operations on these sequences.

1. By $\mathfrak{x}\mathfrak{y}$ or $\mathfrak{x}, \mathfrak{y}$ I denote the sequence obtained by writing the sequence $\mathfrak{y}$ behind the sequence $\mathfrak{x}$ ($\wedge\mathfrak{x} = \mathfrak{x}\wedge = \mathfrak{x}$).

2. By $\mathfrak{x}; \mathfrak{y}$ I denote the sequence obtained by writing $\mathfrak{y}$ behind $\mathfrak{x}$ but leaving out members which already occur in $\mathfrak{y}$.

3. If $\mathfrak{T}$ is a sequence of any n terms $\mathfrak{T} = (t_1, \ldots, t_n)$ and $\mathfrak{S}$ likewise of any m terms $\mathfrak{S} = (s_1, \ldots, s_m)$, I denote by $\mathfrak{T}(\mathfrak{S})$ the following sequence with n members:

First member $\quad t_1(s_1)(s_2)\ldots(s_m)$

$$\vdots$$

$$t_n(s_1)(s_2)\ldots(s_m)$$

under the assumption that the types of the t_i and the s_i

69

are such that these expressions here are again meaningful terms.

In particular, by this definition, $\wedge(\mathfrak{S}) = \wedge, \mathfrak{T}(\wedge) = \mathfrak{T}$.

3′. Evidently we have $\mathfrak{T}(\mathfrak{S}_1\mathfrak{S}_2) = \mathfrak{T}(\mathfrak{S}_1)(\mathfrak{S}_2)$. Sometimes ⟦it is⟧ more perspicuous to write $\mathfrak{S}$ as an index $\mathfrak{T}_{\mathfrak{S}} = \mathfrak{T}(\mathfrak{S})$.

I am also introducing types for these series of expressions. The type of the series $a_1, \ldots, a_n$ is to be the series of $t_1, \ldots, t_n$ of the types of $a_1, \ldots, a_n$. So new types = series of old types. And if $t\tau s$ are two of the new types, I mean by $t\tau s$ the type of the series $\mathfrak{T}$ of terms such that $\mathfrak{T}(\mathfrak{S})$ is of type t whenever $\mathfrak{S}$ is of type s or if $s = \{s_1, \ldots, s_n\}$[35] $\quad t = \{t_1, \ldots, t_n\}$ then

$$\begin{aligned} t\tau s = {} & (t_1\tau s_1)\tau s_2 \ldots \tau s_n \\ & \quad \vdots \\ & (t_n\tau s_1)\tau s_2 \ldots \tau s_n \end{aligned}$$

4. $\mathfrak{T}$ and $\mathfrak{S}$ are sequences of terms of the same type (i.e. i^{th} member of $\mathfrak{S}$ same type as i^{th} member of $\mathfrak{T}$).

I denote by:

$$\mathfrak{T} \doteq \mathfrak{S}$$

the following system of equalities

$$t_1 \doteq s_1, \ldots, t_n \doteq s_n \qquad \text{e.g. } \wedge \equiv \wedge = \wedge$$

and I say that $\mathfrak{T} \doteq \mathfrak{S}$ is demonstrable in Σ if all single equalities are demonstrable.

Now let A be an expression containing a certain variable x (perhaps in several places). The remainder of the expression A obtained by striking out this variable x wherever it occurs I call *matrix*. I shall use the letters M, N, K ⟦to denote matrices⟧. So by this definition, a matrix would be a sequence of symbols with vacant spaces, which becomes a propositional function if the vacant spaces are filled by a term of an appropriate type.

70

However, it is more convenient to fill the vacant space by a new kind of symbol ξ (different from all symbols introduced so far) and call the expression thus obtained a matrix. In an analogous manner an expression obtained by striking out n different variables $x_1, \ldots, x_n$ of arbitrary types and replacing them by the symbols $\xi_1, \ldots, \xi_n$ is called a matrix with n arguments. It is to be noted that the

[35] Gödel has originally written here $s = \{s_1, \ldots, s_k\}$ but defines s as $s_1, \ldots s_n$ thereafter.

symbols $\xi_1, \ldots, \xi_n$ are not specified as to type. They are just tokens to identify the vacant spaces.

If M is a matrix with n arguments and $\mathfrak{T}$ a sequence of n terms $t_1, \ldots, t_n$ of appropriate types, I denote by $M[\mathfrak{T}]$ the expression obtained by putting these n terms in place of the empty spaces of

71

this matrix (i.e., in place of the ξ_i), i.e., $M[t_1, \ldots, t_n] = M \begin{bmatrix} \xi_1,\ldots,\xi_n \\ t_1,\ldots,t_n \end{bmatrix}$.

Now let A be any expression and let $\mathfrak{a}$ be the sequence of all free variables occurring in A arranged say in the order of their occurrence and in the same manner $\bar{\mathfrak{f}}$ and $\underline{\mathfrak{x}}$ underlined the sequences of all existential respectively universal variables in A. Then you can find a uniquely determined matrix M such that $A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$.

This is the standard representation of expressions which I shall use. The aforementioned necessary and sufficient condition that A can be asserted in $\overline{\Sigma}$ can now be stated as follows: If for a series ϱ of constants of appropriate type, $M[\mathfrak{a}, \varrho(\mathfrak{a}), \mathfrak{x}]$ can be asserted in Σ.

The *advantage of* this notation involving series of variables is that many things can be

72

formulated literally in the same way as if $\mathfrak{a}, \mathfrak{f}, \mathfrak{x}$ were single variables:

E.g. *the rule of equality*: If $\mathfrak{T}$ and $\mathfrak{S}$ are series of terms of the same types as $\mathfrak{a}$ and if $\mathfrak{T} \doteq \mathfrak{S}$ can be asserted in Σ then $M[\mathfrak{T}, \mathfrak{f}, \mathfrak{x}] \equiv M[\mathfrak{S}, \mathfrak{f}, \mathfrak{x}]$ ⟦can be⟧ asserted in Σ, etc.

Also the metatheorems corresponding to the rules of definition can be pronounced for finite sequences of terms: e.g.

1. If $\mathfrak{x}$ is a series of variables in Σ and $\mathfrak{T}$ a series of n terms containing no other variables besides those of $\mathfrak{x}$, then you can find a series ϱ of constants such that $\varrho(\mathfrak{x}) \doteq \mathfrak{T}$ is demonstrable in Σ where $\varrho(\mathfrak{x})$ need not be of type I.

73

Now I can begin to define the meaning of the logical operations $\cdot, \vee, \rightarrow, (\ ), (\exists)$ as applied to expressions of $\overline{\Sigma}$ in such a manner that if $A, B \,\varepsilon\, \overline{\Sigma}$ then $A \rightarrow B$ is again $\varepsilon\, \overline{\Sigma}$ and likewise for the others. The expressions $A \rightarrow B$ etc. are defined only up to congruences and these operations $\rightarrow$ etc. will be inv⟦ariant⟧ with respect to congruences i.e. the notation of the

74

bound variables plays no role. Therefore I can assume that the bound variables in A are all different from the bound variables in B and that also all bound variables are different from all free variables. So let be

$$A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$

$$B = N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$$

where the variables $\mathfrak{f}, \mathfrak{x}$ are different from $\mathfrak{g}, \mathfrak{y}$ and $\mathfrak{a}, \mathfrak{b}$ different from $\mathfrak{f}, \mathfrak{x}, \mathfrak{g}, \mathfrak{y}$, but of course $\mathfrak{a}, \mathfrak{b}$ may have common variables or even completely coincide. Of course some or all of these series of variables may be empty. The free variables $\mathfrak{a}, \mathfrak{b}$ are to be considered as parameters, i.e. if we define $A \rightarrow B$ we define the meaning of the following expression (using the customary notation) $(\exists \mathfrak{f})(\mathfrak{x})M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \rightarrow (\exists \mathfrak{g})(\mathfrak{y})N[\mathfrak{b}, \mathfrak{g}, \mathfrak{y}]$ which is a propositional function depending on the variables $\mathfrak{a}, \mathfrak{b}$. So we have to do with two expressions with a prefix of this particular form [where all existential quantifiers precede all universal ones and the problem is to express this again by a propositional function of the same particular form, but that is very easy].

75

First

$$A \,\&\, B \text{ simply} = A \cdot B$$

$$A \text{ v } B \quad '' \quad = A \vee B$$

Let us see that this corresponds to the intuitionistic meaning:

$A \,\&\, B$ thus defined is $M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \cdot N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$ and written with the usual symbolism that means

$$(\exists \mathfrak{f})(\exists \mathfrak{g})(\mathfrak{x})(\mathfrak{y})(M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \cdot N[\mathfrak{b}, \mathfrak{g}, \mathfrak{y}])$$

and actually this last expression is equivalent to the conjunction of the two first by the rules of shifting the quantifiers. The same holds for v.

75^i

Last time I extended the formal system Σ to another system $\overline{\Sigma}$ by introducing existential variables $\overline{x}$ and universal variables $\underline{y}$ which I shall call variables of the second kind as opposed to the former ones, and we introduced one rule of inference concerning propositions containing these variables. The problem which we want to solve now is this: we want to define these binary operations $A \vee B$, $A \,\&\, B$, $A \rightarrow B$ which applied to expressions A, B of $\overline{\Sigma}$ give again expressions of $\overline{\Sigma}$, and two binary operations $(x)A$, $(\exists x)A$ which applied to an expression of $\overline{\Sigma}$ and a variable of the first kind (a free variable) yield expressions of $\overline{\Sigma}$, and then we want to prove that the axioms and rules of inference are satisfied for this interpretation. [Where by an asserted proposition we have to understand of course one asserted owing to the axioms and rules of $\overline{\Sigma}$.]

Last time I defined already conjunction and disjunction by the stipulation

75^{ii}

$$A \,\&\, B = A \cdot B$$
$$A \,\mathrm{v}\, B = A \vee B$$

where these two operations on the right side mean writing the two expressions beside each other and joining them by a symbol of conjunction, resp. disjunction.

Next[36] I have to define implication: So let $A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$, $B = N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$. We have to define $A \rightarrow B$. Let us consider first the special case where $\mathfrak{x}, \mathfrak{y}, \mathfrak{f}, \mathfrak{g}$ consist each of only one variable i.e. in the usual notation, $A = (\exists f)(x)\Phi(f, x)$, $B = (\exists g)(y)\Psi(g, y)$,[37] and the problem [[is]] to transform this implication

$$(\exists f)(x)\Phi(f, x) \rightarrow (\exists g)(y)\Psi(g, y)$$

into an expression of the same form i.e. where all existential quantifiers precede all universal ones.

76

That is not possible by simply shifting the quantifiers. But we can use the following heuristic argument. This expression means: If there exists an f satisfying

[36] This passage appears on p. 75 separated by a thick line. It seems to fit best below p. 75^{ii}.

[37] Gödel has here written only $A =$, $B =$

a certain condition then there exists a g satisfying another condition. In a constructive logic that will mean: We have a procedure p which allows us to obtain such a g if such an f is given i.e. this implication means:

$$(\exists p)(f)[(x)\Phi(f,x) \rightarrow (y)\Psi(p(f),y)]$$

and now here the operation of implication is applied to an expression of a simpler type (since no longer existential quantifiers occur). But what can it mean in a constructive logic that if for all x something is true then for all y something else is true? The simplest meaning which suggests itself is this: Given

77

a counterexample for the second assertion, one can construct a counterexample for the first, i.e. the expression in square brackets will be equivalent to

$$(\exists r)(y)[\sim\Psi(p(f),y) \rightarrow \sim\Phi(f,r(y))]$$

hence the whole expression to

$$(\exists p)(f)(\exists r)(y)[\Phi(f,r(y)) \rightarrow \Psi(p(f),y)]$$

Now here implication is applied only to expressions containing no quantifiers. Therefore we can replace it by $\supset$. But here we have too many changes between $\exists$ and ( ), but $(f)(\exists r)$ simply means that there exists a function q which associates such a function r with each a. I.e. this expression is equivalent to:

$$(\exists p)(\exists q)(f\,g)[\Phi(f,q(f)(y)) \supset \Psi(p(f),y)]$$

and this is again an expression of $\overline{\Sigma}$. So this is the definition of implication for expressions with only one existential and one universal variable.

This definition could also be arrived at as follows. Let's bring this implication to a normal form. That can be done in different manners since the order of quantifiers is not uniquely determined. Let's do it in this manner that as far as possible the existential quantifiers come after the universal ones. (This is in a sense the weakest normal form because $(\exists x)(y)\ldots \supset (y)(\exists x)\ldots$ but not vice versa.) We obtain

$$\begin{aligned} &(f)(\exists g)(y)(\exists x)[\Phi(f,x) \supset \Psi(g,y)] \\ \equiv\ &(\exists\, p\, q)(f\, g)\{\Phi[(f,q(f,y)] \supset \Psi[p(f),y]\} \end{aligned}$$

You see we have here two new variables p, q

$$\left\{\begin{array}{lll} p & \text{of type} & g\tau f \\ q & '' & (x\tau f)\tau y \\ & \text{or} & x\tau(f, y) \end{array}\right.$$

I am using the type symbol τ also for arguments which are not themselves type symbols but terms of types under consideration.

In the case where we have instead of single variables f, g, x, y, series of such variables, we shall have functions p, q of several variables and several functions instead of two. To be more exact:

78

If A, B are two expressions written down here then[38]

$$A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$
$$B = N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$$

Let $\mathfrak{p}$ be a series of variables of type $\mathfrak{y}\tau\mathfrak{f}$

$p_i(\mathfrak{f})$ is meaningful and of the same type as g_i, i.e., $\boxed{p_i \text{ is of type } y_i\tau\mathfrak{f}}$

and let $\mathfrak{q}$ be a series of variables of type $\mathfrak{x}\tau(\mathfrak{f}, \mathfrak{y})$

$q_i(\mathfrak{f}, \mathfrak{y})$ is meaningful and of the same type as x_i, i.e.,

$$\boxed{q_i \text{ is of type } x_i\tau(\mathfrak{f}, \mathfrak{y})}$$

and assume in addition that these variables $\mathfrak{p}, \mathfrak{q}$ are different from each other and from the variables occurring already in A, B. Then

$$A \rightarrow B =_{Df} M[\mathfrak{a}, \underline{\mathfrak{f}}, \bar{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{y}})] \supset N[\mathfrak{b}, \bar{\mathfrak{p}}(\underline{\mathfrak{f}}), \underline{\mathfrak{y}}]$$

In case of $\mathfrak{f}, \mathfrak{g}, \mathfrak{x}, \mathfrak{y}$ consisting each of one variable, this is exactly the former expression. In the general case or[39] if you prefer $A \rightarrow B = A\binom{\bar{\mathfrak{f}}\ \ \underline{\mathfrak{x}}}{\underline{\mathfrak{f}}\ \ \bar{\mathfrak{q}}(\underline{\mathfrak{f}}\,\underline{\mathfrak{y}})} \supset B\binom{\bar{\mathfrak{g}}}{\bar{\mathfrak{p}}(\underline{\mathfrak{f}})}$ where $\mathfrak{f}, \mathfrak{g}, \mathfrak{x}, \mathfrak{y}$ are the series of all existential respectively universal variables occurring in A respectively in B ⟦and⟧ any $\mathfrak{q}, \mathfrak{p}$ new variables of appropriate type.

This definition of $A \rightarrow B$ comprises of course also the case where A or B or both contain no existential variables or no universal variables. In this case

[38] The sequence of variables denoted here by $\mathfrak{p}$ seems to originally have been a Sütterlin $\mathscr{h}$ later corrected into a letter that does not match any Sütterlin letter. Because in the proof of soundness of the intuitionistic axiom, both this letter and the letter h occur in the same formulas, we have interpreted the nondescript letter as p.

[39] The rest of the sentence has later been cancelled.

one or several of the series $\mathfrak{f}, \mathfrak{g}, \mathfrak{x}, \mathfrak{y}$ will be empty. E.g., this will happen for the predicate of absurdity defined by $\neg A = A \rightarrow (0 = 1)$. Here B contains no variables, hence $\mathfrak{g}, \mathfrak{y}$ are empty, hence $\underline{\mathfrak{f}}, \underline{\mathfrak{y}} = \underline{\mathfrak{f}}$ and we obtain

$$\neg A = M[\mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{q}}(\underline{\mathfrak{f}})] \supset (0 = 1)$$

which is equipollent in $\overline{\Sigma}$ to

$$\sim M[\mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{q}}(\underline{\mathfrak{f}})]$$

I call two expressions equipollent if the proof of one allows you to construct the proof of the other and vice versa.

78′

In order to obtain $A \rightarrow B$ out of two given expressions A, B of $\overline{\Sigma}$ what we have to do is this:

1. Form this implication $A \supset B$.

2. We replace the existential variables of A by the corresponding universal variables.

3. We associate with each universal variable x_i of A a new existential variable q_i of this type and with each existential variable g_i of B a new existential variable p_i of this type.

4. We replace each universal variable x_i of A by a term whose first symbol is the corresponding existential variable q_i and whose argument series consists of all variables $\underline{\mathfrak{f}}, \underline{\mathfrak{y}}$, and finally

5. We replace each existential variable of B by a term whose first symbol is the corresponding existential variable q_i and whose argument series consists of the variables of $\underline{\mathfrak{f}}$.

In order that this expression $A \rightarrow B$ be uniquely determined up to the not⟦ation⟧ of bound variables it is necessary that we have a definite arrangement of the variables $\mathfrak{f}, \mathfrak{y}$. Owing to the definition of M,

79

the proof that these two expressions are equipollent follows immediately from the criterion which reduces demonstrability in $\overline{\Sigma}$ to demonstrability in Σ since in Σ, we have $A \supset 0 = 1$ equipollent to $\sim A$ because we have the whole calculus of propositions. So in usual notation and in case of $\mathfrak{f}, \mathfrak{x}$ consisting of one variable,

$$\neg(\exists f)(x)\varphi(f, x)$$

means

$$(\exists p)(f){\sim}\varphi(f, p(f))$$

If in particular A has only universal variables (no existential) i.e. if also $f = \wedge$, then we have no arguments here. But this means in essence not $(\exists q){\sim}M[a, q]$, i.e. e.g. $\neg(x)\varphi(x)$, where $\varphi(x)$ is an unquantified expression, is by this definition $= (\exists x){\sim}\varphi(x)$, i.e., the same as constructive negation. This has the consequence that intuitionistic logic is constructive in even a stronger sense that defined in my first lecture. But this equivalence holds of course only for unquantified expressions.

If A, B *are expressions of the* first kind i.e. if $\bar{\mathfrak{f}} = \underline{\mathfrak{x}} = \underline{\mathfrak{y}} = \overline{\mathfrak{g}} = \wedge$[40] then of course also $\overline{\mathfrak{p}}, \overline{\mathfrak{q}} = \wedge$, hence $A \rightarrow B = A \supset B$.

The same is true already *if only* A is of the first kind because $\underline{\mathfrak{f}} = \wedge$ and $\overline{\mathfrak{p}}(\underline{\mathfrak{f}}) = \overline{\mathfrak{p}}$, but then the type of $\overline{\mathfrak{p}} =$ type of $\overline{\mathfrak{g}}$. Hence B and this second member here differ only in the notation of the bound variable which is irrelevant, and the first member does not differ at all from A.

Now I have to define the meaning of quantification:

1. Let A be an expression of $\overline{\Sigma}$. Then

$$(\exists x)A =_{Df} A(\tfrac{x}{\overline{x}})$$

or if $\overline{x}$ contained already $= A(\frac{x}{y})$ (up to congruences), hence if A does not contain x, $(\exists x)A = A$.

80

2. $(x)A$ is defined as follows: Let A be:

$$A = M[x, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{z}}]$$

[40] Gödel has mistakenly written $\bar{\mathfrak{f}} = \bar{\mathfrak{x}} = \overline{\mathfrak{y}} = \mathfrak{g} = \wedge$.

assuming x actually occurs as a free variable in A, $x \sim\varepsilon\, \mathfrak{a}$.

$$(x)A = M[\underline{x}, \mathfrak{a}, \overline{\mathfrak{g}}(\underline{x}), \underline{\mathfrak{z}}]$$

where $\overline{\mathfrak{g}}$ is a series of variables of type $\overline{\mathfrak{f}}\tau x$.

So in order to obtain $(x)A$ from A what we have to do is this:

1. Underlining the variable x wherever it occurs.
2. Associate with each existential variable $\overline{f}_i$ of A an existential variable $\overline{g}_i$ of type $f_i \tau x$.
3. Replace each existential variable of A by a term whose first symbol is the corresponding existential variable $\overline{g}_i$ and whose argument series consists of only one member, namely x.

In the case where x does not occur in A we don't have $(x)A = A$ because $\underline{x}$ appears here also in this case, but ⟦the expressions are⟧ of course equipollent.

As an exemplification of this definition, let us consider the case where $\mathfrak{f}$ and $\mathfrak{z}$ consist of exactly one variable, $A = (\exists f)(z)\varphi(x, f, z)$. Then $(x)A$[41] will mean in a constructive logic that one has a function g which allows to compute for any given x an f satisfying the condition i.e. $(\exists g)(x\ z)\varphi(x, g(x), z)$. But this is exactly the above expression written in the usual notation.

It is easily seen that always: A and $(x)A$ are equipollent in $\overline{\Sigma}$. For: A demonstrable in $\overline{\Sigma}$ means for some constant ϱ:

$$M[x, \mathfrak{a}, \varrho(x, \mathfrak{a}), \mathfrak{z}] \text{ in } \Sigma$$

$(x)A$ in $\overline{\Sigma}$ means

$$M[x, \mathfrak{a}, \sigma(\mathfrak{a})(x), \mathfrak{z}] \text{ in } \Sigma$$

but this is almost the same (except for the order

81

of arguments of the constants ϱ and σ), hence in terms of such a σ you can define such a ϱ and vice versa.

The difficulty arises that the operations of $(x)A$, $(\exists x)A$ were defined already in general considerations about intuitionistic logic. Namely, they meant

[41] Here Gödel has originally written $(x)(\exists f)(z)A$.

simply putting the symbol of quantification in front, and now we have an entirely different and much more complicated operation of quantification, namely going over from this expression to that and from this to that. Therefore I shall write Π, Σ for this quantification. So Π, Σ are related to $(\,)$, $\exists$ in the same manner as $\rightarrow$ to $\supset$.

The next thing I have to do is to show that the axioms of intuitionistic logic are satisfied i.e. the following interpretation is a model of the abstract system of intuitionistic logic.

1. Propositional function = propositional function of $\overline{\Sigma}$.
2. Variable = variable of Σ (variable of the first kind).
3. Terms = terms of Σ (containing only variables of the first kind).
4. Asserted propositional function = propositional function asserted in $\overline{\Sigma}$ (owing to axioms of $\overline{\Sigma}$).
5. The logical operations have the meaning just defined.
6. Substitution $A\binom{x}{t}$ means replacement of x by t.

But in this model, not only the intuitionistic axioms are satisfied, but also of course the same axioms of Σ and furthermore also the following principle of complete induction: If $A \varepsilon \overline{\Sigma}$, then if $A\binom{x}{0}$ and $A \supset A\binom{x}{\nu(x)}$ ⟦can be⟧ asserted in $\overline{\Sigma}$, then A ⟦can⟧ also ⟦be asserted⟧ in $\overline{\Sigma}$.

So one can prove that the logical operations (as I defined them) yield a model for intuitionistic logic including the principle of complete induction.

You can look at this state of affairs also from another viewpoint. Namely, let us denote by Σ_I the formal system obtained from Σ by introducing quantifiers and the logical operations in the usual way and assuming the usual axioms of intuitionistic logic, including the rule of complete induction for arbitrary expressions (involving

82

as many quantifiers as you like).

So Σ_I is what I called before an application of intuitionistic logic where the specific axioms are

1. The axioms of Σ and

2. The rule of complete induction for arbitrary expressions.

Then it is possible to associate with each expression A of Σ_I an expression A' of $\overline{\Sigma}$ by the following inductive definition:

1. $A' = A$ for atomic formulas.

2. $(A \subset B)' = A' \rightarrow B'$

   $(A \vee B)' \ = A'$ v B' etc.

Similarly for all-operation

$$[(x)A]' = \Pi x(A').$$

By this recursive definition, evidently to each expression of Σ_I

83

⟦corresponds⟧ exactly one of $\overline{\Sigma}$. But owing to the fact that the axioms of intuitionistic logic are satisfied for this interpretation and also the rule of complete induction, it follows that if A ⟦is⟧ demonstrable in Σ_I then ⟦so is⟧ A' in $\overline{\Sigma}$. And also the inverse is true: if A' is demonstrable in $\overline{\Sigma}$ then A is demonstrable in Σ_I, trivial because every correct proof of $\overline{\Sigma}$ is itself a correct proof in Σ_I.

84

Now this mapping of Σ_I on $\overline{\Sigma}$ owing to the operation A' yields the desired proof of constructivity of intuitionistic logic. For assume $(\exists x)(A)$ ⟦is⟧ demonstrable in Σ_I (where this expression is supposed to contain no more free variables). Then $(\Sigma x)(A')$ ⟦is⟧ demonstrable in $\overline{\Sigma}$. Hence by the definition of Σ, $A'(\frac{x}{\overline{x}})$ ⟦is⟧ demonstrable in $\overline{\Sigma}$, but then $A'\binom{x}{\alpha}$ ⟦is⟧ demonstrable in Σ for some constant term α. But then $A'\binom{x}{\alpha}$ ⟦is⟧ demonstrable also in Σ_I.

So you have: If $(\exists x)A$ is demonstrable in Σ_I then there *exists a constant term* α *such that* $A\binom{x}{\alpha}$ *is demonstrable in* Σ_I.

85

Owing to the fact that if $A(x)$ is an atomic expression

$$\neg(\Pi x)A(x) = (\Sigma x){\sim}A(x),$$

we conclude that even the proof of the absurdity of $(x)A(x)$ in Σ_I gives a means to find the counterexample if only the propositional function A under consideration is atomic (the same thing could be proved for unquantified, i.e., decidable ⟦formulas⟧). But now this expression involves no $\exists$, $\vee$. Therefore if it is classically demonstrable, it is also intuitionistically demonstrable. So if we denote by Σ_K the system obtained from Σ by adding the rules of classical logic, then we have that even a proof in Σ_K of ${\sim}(x)A(x)$ gives you a means to construct the counterexample. This is a result obtained also by Gentzen at the occasion of his consistency proof for number theory.

86

Classical number theory is a portion of Σ_K (and intuitionistic number theory a portion of Σ_I), namely it is in a sense of the lowest level of Σ_K (resp. Σ_I). Therefore owing to the mapping A', also every number-theoretic theorem A will have a corresponding proposition A' in $\overline{\Sigma}$ which is demonstrable in $\overline{\Sigma}$. But it is to be noted that A' will in general by no means be number-theoretic if A is number-theoretic. Because by the applications of the logical operations (in particular $\rightarrow$), a heightening of the level takes place. Namely $A \rightarrow B$ contains variables whose arguments and values are variables of A, B, hence these variables are of a higher level. So if you have sufficiently many $\rightarrow$ signs in a number-theoretic proposition A, the corresponding proposition A' may have an arbitrarily high level. It can be shown that this heightening of the levels is necessary in this sense that is is impossible to interpret number theory in the subsystem, say $\overline{\Sigma}_n$ of $\overline{\Sigma}$, which contains only the propositions and axioms of $\overline{\Sigma}$

87

up to the level n (i.e. involving variables only up to the level n). Namely, this is impossible because each system $\overline{\Sigma}_n$ can be proved consistent within number theory.

Finally, I wish to remark that this whole scheme of defining the logical notions has a certain relation to what Russell intended in the §9 of the *Principia Mathematica*. Namely, it is chiefly the question of defining the meaning of the logical operations for expressions involving quantifiers provided that this meaning for unquantified expressions is given. Russell tried to accomplish this by means of the rules of shifting the quantifiers, e.g.

88

$p \vee (x)\varphi(x) =_{Df} (x)[p \vee \varphi(x)]$. It is easily seen that this scheme cannot go through if you wish to admit in your proofs only intuitionistic logic. For the following reason: This scheme of definition can give as the meaning of a proposition A involving quantifiers only a normal form of A. But there are identities of intuitionistic logic no normal form of which is an identity of intuitionistic logic.

Trivial:

$\neg(x)(\exists y)[F(x) \cdot \neg F(y)]$
but $(\exists x)(y)[\neg(F(x) \cdot \neg F(y))]$
$\neg F(x) \vee F(y)$
$(y)[\neg(F(a) \vee F(y))]$

More complicated:

$$(u\, z)[F(u) \vee \neg F(z) \supset R(u, z)] \supset \neg(x)(\exists y)\neg R(u, z)$$

88′

The structure of this expression A' is in general pretty intricate. There is however one case where A' is a well-known expression whose equivalence with A is very often used in formal logic, namely if A is a positive expression (i.e., no $\supset$ or $\neg$ occurs outside of any quantifier).

In this case A' is obtained as follows (let's assume all different bound variables of A are denoted by different expressions) and let us call a bound variable dominated by another if the corresponding quantifier dominates. Now in this case A' is obtained from A by

1. Dropping all quantifiers.
2. Underlining all universal variables.
3. Replacing each existential variable x by term of the form $\overline{f}(\underline{u}_1)\ldots(\underline{u}_m)$ where $u_1, \ldots, u_m$ are all universal variable dominating x and where the letters $\overline{f}$ are different for different existential variables x.

This transformation is used e.g. for proving Skolem-Lowenheim's theorem. Herbrand called the functions f *"fonctions d'indice."*

⟦88″⟧

The proof that this expression constructed by means of the fonction d'index (call it A'') really is the same as A' is immediate by induction on the number of logical symbols in A.

1. If zero then $A' = A$ ⟦and⟧ $A'' = A$, otherwise

$$A = B \cdot C, B \vee C, (x)B, (\exists x)B$$

and by induction $B' = B''$, $C' = C''$. But also e.g.

$$A' = (B \cdot C)' = B' \,\&\, C' = B' \cdot C' = B'' \cdot C'' = (B \cdot C)'' = A''$$

and ⟦this is⟧ as simple in the case of "or".

$$((\exists x)A)' = A'(\tfrac{x}{x})$$

$$((\exists x)A)'' = A''(\tfrac{x}{x})$$

$$((x)A)' = (\Pi x)A' = (\Pi x)A''$$

But now $(\Pi x)A''$ is obtained from A'' by turning x into $\overline{x}$ and replacing each existential variable f (i.e. each index function) by a term $\overline{h}(\underline{x})$ where $\overline{h}$ is a new existential variable. But in exactly the same manner $[(x)A]''$ is obtained from A''.

89·1

Introduction 1

In the two preceding lectures I defined the logical operations v, &, $\rightarrow$, Σ, Π in such a manner that applied to expressions of the formal system $\overline{\Sigma}$, they yield again expressions of $\overline{\Sigma}$. I am sorry I made a mistake in one of the definitions, namely of "or". I defined A v $B = A \vee B$ provided that the bound variables in A and B are different from each other. Now this is a very natural and simple notion of disjunction and it is intuitionistically admissible, but it is not the notion in Brouwer's and Heyting's logic for the following reason. Brouwer's disjunction has this property that if $A \vee B$ is demonstrable then always either A or B is demonstrable but that is not the case with this notion.

E.g. $A = (\exists y)(x)\varphi(y, x) \quad B = (\exists z)(u)\psi(z, u)$

$$A \text{ v } B = (\exists\, y\, z)(x\, u)[\varphi(y, x) \vee \psi(z, u)]$$

Now if you have two objects y, z for which you can prove $(x\, u)[\varphi(y, x) \vee \psi(z, u)]$, then you know by classical logic that either $(x)\varphi(y, x)$ is true or $(u)\psi(z, u)$ is true, but it may very well happen that you are unable to decide which of the two

89·2

Introduction II

possibilities is realized although you can prove this formula. And this has also the consequence that not all axioms of intuitionistic logic can be proved for this notion of "or", namely the following rule of inference cannot:

⟦From⟧ $A \to C$ $B \to C$ ⟦conclude⟧ $A \text{ v } B \to C$.

For this reason, we have to choose a more complicated definition of "or". Namely, let $\overline{u}$ be an existential variable not occurring in A and B. Then

$$A \text{ v } B \;_{def}\!= (A \cdot \overline{u} = 0) \vee (B \cdot \overline{u} = 1)$$

(provided that bound variables ⟦are different in A and B⟧).

It is easily seen that now it is true that if you can prove A v B, you can prove either A or B, because a proof for A v B in $\overline{\Sigma}$ allows you to find a number $\overline{u}$ satisfying this condition. This number can only be either 0 or 1 and accordingly as to whether $\overline{u} = 0$ or $= 1$ you can prove A or you can prove B. And for this notion all intuitionistic axioms can be proved, but it would perhaps be worth while to investigate what axioms this other simpler notion of "or" satisfies.

89·3

Introduction III

Before going on I wish to remind you shortly of the definition of $A \to B$:

Let $\overline{\mathfrak{f}}$ be the series of all universal variables of A in the order of their occurrence ⟦and⟧ likewise $\underline{\mathfrak{x}}$ the series of all existential variables of A in the order of their occurrence. ⟦Let⟧ $\overline{\mathfrak{g}}$, $\underline{\mathfrak{y}}$ ⟦denote⟧ the same thing for the expression B. Then we associated with each universal variable of A an existential variable (of a certain type) and we called the series of all associated variables $\overline{\mathfrak{q}}$ and in the same manner to each existential variable of the second expression, we associated a new existential variable of a certain type (let's call its series $\overline{\mathfrak{p}}$) and then

$$A \rightarrow B = A \supset B \begin{pmatrix} \bar{\mathfrak{f}}\ \underline{\mathfrak{x}} & \underline{\mathfrak{y}} \\ \underline{\mathfrak{f}}\ \overline{\mathfrak{p}}(\underline{\mathfrak{f}},\underline{\mathfrak{y}}) & \overline{\mathfrak{q}}(\underline{\mathfrak{f}}) \end{pmatrix}$$

The types of these series of variables $\overline{\mathfrak{p}}$ and $\overline{\mathfrak{q}}$ are of course determined by the condition that this expression be meaningful. In addition, all variables ⟦should be⟧ different from each other and ⟦text ends⟧

89·4

Next let us see that $W \rightarrow B$, cnf. ⟦p. 93⟧ bottom.[42]

Now let us prove axiom $A \rightarrow A \,\&\, A$. For this purpose, I have to make some preparations, cnf p. 59.[43]

The assumption that T, S be of type I is evidently superfluous, since if they are of a higher type, $\varrho(\mathfrak{x}) \doteq T$ is by definition the same formula as $\varrho(\mathfrak{x})(\mathfrak{y}) \doteq T(\mathfrak{y})$ for an appropriate series $\mathfrak{y}$ of variables, where now $T(\mathfrak{y})$ is of type I. Therefore we can apply the former theorem to $\begin{cases} T(\mathfrak{y}) \\ S(\mathfrak{y}) \end{cases}$. Furthermore the theorem can be generalized for sets of equalities instead of one expression. I.e. we must only define $A \supset \mathfrak{M}$ where $\mathfrak{M}$ is a finite sequence of expressions, say $M_1, M_2, \ldots, M_k$ and A an expression. In this case $A \supset \mathfrak{M}$ is by definition the sequence $A \supset M_1, \ldots, A \supset M_k$. Then we have the following theorem: If $\mathfrak{S}, \mathfrak{T}$ are two sequences

89·5

of terms of the same type containing no other variables but those of the series $\mathfrak{x}$ and if A is an expression containing no other variables but those of $\mathfrak{x}$, then there exists a sequence of constants ϱ such that

$$\left.\begin{array}{r} A \supset \varrho(\mathfrak{x}) \doteq \mathfrak{S} \\ \sim A \supset \varrho(\mathfrak{x}) \doteq \mathfrak{T} \end{array}\right\} \text{is demonstrable in } \Sigma$$

Of course also the rules of extensionality can be generalized for the case where the premiss is of this form (i.e. consists of a set of implications) simply by applying the former rule of extensionality k times if we have k equalities here. I.e. from $A \supset \mathfrak{S} \doteq \mathfrak{T}$ (where $\mathfrak{S}\,\mathfrak{T}$ are sequences of terms $S_1, \ldots, S_k, T_1, \ldots, T_k$)

---

[42] The loose pages 89–106 were originally numbered 1–17; Gödel has later erased the old page numbers and written new numbers on top. The references on these pages still use the old pagination; below, we have replaced them by the new page numbers.

[43] Page 59 of the lecture notes does not relate to this in any way.

you can infer $A\binom{\mathfrak{x}}{\mathfrak{S}} \equiv B\binom{\mathfrak{x}}{\mathfrak{T}}$ where $\mathfrak{x}$ is a sequence of variables of the same type as $\mathfrak{S}$ and $\mathfrak{T}$.

And now I can give the proof of this axiom here, cnf. ⟦addition to p. 89⟧ under the line.[44]

⟦89·6⟧

1. Mention[45] somewhere that a system can have different types of basic objects.

2. As constants ⟦use⟧ Greek letters with indices and special mathematical symbols.

3. Consider $\cdot \sim$.

4. numerals?

5. Superfluous.

[44] This refers to the section $1'$ at the bottom of "Addition to 89".

[45]

1. Irgendwo erwähnen, dass ein System verschiedene Arten von Grunddingen haben kann.

2. Als Konstanten griechische Buchstaben mit Index und spezielle mathematische Zeichen.

3. $\cdot \sim$ betrachten

4. numerals?

5. Überschüssig

Proof of the soundness of the intuitionistic axioms

⟦89·7⟧[46]

Note![47] $\doteq$ denotes logical identity (and it will only be used ⟦to mean⟧

$$f = f' \cdot y = y' \supset f(y) = f'(y'))$$

What this means for a procedure in a certain sense is that the definition of the procedure must be the same (not just the extension).

In the proof practically nothing else is used except for the rule of definition $f(x_1, \ldots, x_k) = A(x_1, \ldots, x_k)$ for arbitrary meaningful expressions A and the rules $f(x_1, \ldots, x_k) = A(x_1, \ldots, x_k)$ and $f(x_1, \ldots, x_l)(x_{l+1}, \ldots, x_k) = A(x_1, \ldots, x_k)$ [that such an f exists must be shown (uniqueness unnecessary)]. The corresponding more complicated rules are *logically* provable from these by iteration.

In the induction axiom (p. 101) a countable iteration of these rules occurs. In principle, one needs functions of several variable only to prove the rule of definition (in Church's interpretation). *However, one needs them* for this purpose, and also for the inductive definition.

89

But now let us begin with the proof of the intuitionistic axioms. Let us confine our attention first to those axioms and rules which don't contain $\vee, \exists$ (which are sufficient). They are the following:

---

[46] This note, written in shorthand on two slips of paper, can be found inside the envelope that contains the missing pages 89–106. The first slip of paper also contains Gödel's instructions for arranging the missing pages.

[47] Zu beachten! $\doteq$ bedeutet *logische Identität* (und es wird nur verwendet $f = f' \cdot y = y' \supset f(y) = f'(y')$), was bei Verfahren in gewissem Sinne bedeutet, dass die *Definition des Verfahrens* dieselbe sein muss (nicht bloss die Extension). In den Beweis wird praktisch nichts verwendet als die Definitionsregel $f(x_1, \ldots, x_k) = A(x_1, \ldots, x_k)$ für beliebige sinnvolle Ausdrücke A, und die Regel $f(x_1, \ldots, x_k) = A(x_1, \ldots, x_k)$ und die Regel $f(x_1, \ldots, x_l)(x_{l+1}, \ldots, x_k) = A(x_1, \ldots, x_k)$ [dass es ein solches f gibt, muss gezeigt werden (Eindeutigkeit überflüssig)]. Die entsprechenden komplizierten Regeln sind durch Iteration aus dieser logisch beweisbar.

Beim Induktionsaxiom (p. 101) findet eine abzählbare Iteration dieser Regeln statt. Man braucht im Prinzip Funktionen mehrerer Variablen nur, um die Definitionsregel (in Churchscher Interpretation) zu beweisen. Aber man braucht sie für diesen Zweck. Ebenso braucht man sie für induktive Definition.

Axioms

$A \to A\&A$ p. 89 bottom

$A\&B \to B$ 89'

$A\&B \to B\&A$ 92 bottom

$W \to B$ 93 bottom

$(x)A \to A\binom{x}{t}$ 100

(The last one can be simplified, see p. 100)

Rules

$$\frac{\begin{array}{l}A\\ A \to B\end{array}}{B}\quad 95$$

$$\frac{\begin{array}{l}A \to B\\ B \to C\end{array}}{A \to C}\quad 96$$

$$\frac{(A\&B) \to C}{A \to (B \to C)}\ \updownarrow\quad 97$$

$$\frac{A \to B}{A \to (x)B}\quad \text{98 bottom}$$

Complete Induction 101 bottom

[[The axioms]] *1, 2, 3* [[are]] *substitutable*[48] by the rule of inference

$$\frac{\begin{array}{l}A \to B\\ A \to C\end{array}}{A \to B\&C}$$

and

$$A\&B \to A,\ A\&B \to B.$$

Important! Moreover, commutativity is in any case superfluous when both [[formulas]] of the previous lines hold.

The proofs are all quite simple in principle. They are straightforward consequences of the definitions but mostly the necessary formal apparatus is pretty heavy. Therefore I think it will be sufficient to carry through the proofs in some instances which will suffice to make the method clear.

Let's take as an example for an axiom $A\&B \to B$ and for a rule

[48] *1, 2, 3 ersetzbar* durch die Schlussregel

$$\frac{\begin{array}{l}A \to B\\ A \to C\end{array}}{A \to B\&C}$$

und $A\&B \to A$, $A\&B \to B$.

Wichtig! Ferner ist Com. auf jeden Fall überflüssig, wenn die beiden der vorhergehenden Zeilen [[gelten]].

$$\frac{(A\&B)\rightarrow C}{A\rightarrow(B\rightarrow C)}$$

So let us begin with this axiom: $A\&B \to B$. Addition at the end of these notes (next page).

So what we have to prove is this:

1. If A is an arbitrary formula of $\overline{\Sigma}$ then $A \longrightarrow (A\&A)$ which is a formula of $\overline{\Sigma}$ can be proved in $\overline{\Sigma}$. Similarly for the other axioms of group I.

2. If A, B are formulas of $\overline{\Sigma}$ and if A, $A \longrightarrow B$ are demonstrable in $\overline{\Sigma}$, then B can also be proved in $\overline{\Sigma}$.

Addition to p. 89

So let

$$A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$
$$B = N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$$

hence $A\&B = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \cdot N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$,

hence $A\&B \longrightarrow B =$ (*first associate series of variables*)

$$M[\mathfrak{a}, \underline{\mathfrak{f}}, \bar{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}, \underline{\mathfrak{y}})] \cdot N[\mathfrak{b}, \underline{\mathfrak{g}}, \bar{\mathfrak{q}}'(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}, \underline{\mathfrak{y}})] \supset N[\mathfrak{b}, \bar{\mathfrak{p}}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}), \underline{\mathfrak{y}})]$$

That this can be proved in $\overline{\Sigma}$ means that for some constants ϱ, ϱ', σ this can be proved in Σ:

$$M[\mathfrak{a}, \mathfrak{f}, \varrho_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{g}, \mathfrak{y})] \cdot N[\mathfrak{b}, \mathfrak{g}, \varrho'_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{g}, \mathfrak{y})] \supset N[\mathfrak{b}, \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{g}), \mathfrak{y}]$$

But there are constants ϱ, ϱ', σ for which

$$\begin{cases} \varrho'(\mathfrak{a};\mathfrak{b})(\mathfrak{f}, \mathfrak{g}, \mathfrak{y}) \doteq \mathfrak{y} \\ \sigma(\mathfrak{a};\mathfrak{b})(\mathfrak{f}, \mathfrak{g}) \doteq \mathfrak{g} \end{cases} \quad \text{can be proved in } \Sigma.$$

But now the expression obtained from this by substitution can be proved

$$M[\mathfrak{a}, \mathfrak{f}, \varrho(\mathfrak{a};\mathfrak{b})(\mathfrak{f}, \mathfrak{g}, \mathfrak{y})] \cdot N[\mathfrak{b}, \mathfrak{g}, \mathfrak{y}] \supset N[\mathfrak{b}, \mathfrak{g}, \mathfrak{y}]$$

1.′ Next comes the axiom $A \longrightarrow A\&A$. So assume $A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$. Then $A\&A$ will be

90

by definition of & the expression $M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \cdot M[\mathfrak{a}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$ where $\mathfrak{g}, \mathfrak{y}$ are variables all different from $\mathfrak{f}, \mathfrak{x}$ but of the same types.

Now in order to obtain $A \to (A \& A)$ I have to write first formally

$$M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \supset M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \cdot M[\mathfrak{a}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$$

Now I have to replace in the first member the existential variable by universal ones:

$$M[\mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{x}}, \underline{\mathfrak{y}})] \supset \begin{cases} M[\mathfrak{a}, \overline{\mathfrak{p}}_1(\underline{\mathfrak{f}}), \underline{\mathfrak{x}}] \\ M[\mathfrak{a}, \overline{\mathfrak{p}}_2(\underline{\mathfrak{f}}), \underline{\mathfrak{y}}] \end{cases}$$

Next I have to associate with each existential variable of the second term a new variable (of appropriate type). Existential variables of the second member are $\bar{\mathfrak{f}}, \bar{\mathfrak{g}}$ associated with $\overline{\mathfrak{p}_1}, \overline{\mathfrak{p}_2}$. Then associate with each universal variable of first member a new existential variable, call it series $\overline{\mathfrak{q}}$ and then I have to replace ⟦the first sequence $\underline{\mathfrak{x}}$ with $\overline{\mathfrak{q}}$⟧ etc.

91

Now in order to show that this can be proved in $\overline{\Sigma}$ we have (by the previously proved criterion) to construct series of constants $\varrho, \sigma_1, \sigma_2$ (corresponding to the existential variables $\overline{\mathfrak{q}}, \overline{\mathfrak{p}_1}, \overline{\mathfrak{p}_2}$) such that

$$(\Phi) \quad M[\mathfrak{a}, \mathfrak{f}, \varrho(\mathfrak{a})(\mathfrak{f}, \mathfrak{x}, \mathfrak{y})] \supset M[\mathfrak{a}, \sigma_1(\mathfrak{a})(\mathfrak{f}), \mathfrak{x}] \cdot M[\mathfrak{a}, \sigma_2(\mathfrak{a})(\mathfrak{f}), \mathfrak{y}]$$

can be proved in Σ.

But the constants satisfying the following defined equalities will do that:

$$\begin{cases} \sigma_1(\mathfrak{a})(\mathfrak{f}) \doteq \mathfrak{f} \\ \sigma_2(\mathfrak{a})(\mathfrak{f}) \doteq \mathfrak{f} \\ \varrho(\mathfrak{a})(\mathfrak{f}, \mathfrak{x}, \mathfrak{y}) \doteq \mathfrak{x} & \text{if } {\sim} M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \\ \varrho(\mathfrak{a})(\mathfrak{f}, \mathfrak{x}, \mathfrak{y}) \doteq \mathfrak{y} & \text{if } M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \end{cases}$$

The constants, or rather series of constants, for which this first series of equalities can be asserted

92

in Σ exist by a previous lemma[49] since you can easily verify the variable conditions (the variables occurring as arguments are different from each other and on the right side there are no other variables) and A contains no other variables but at most those occurring ⟦in $\mathfrak{a}, \mathfrak{f}, \mathfrak{x}$⟧.

2. For these constants the implication to be proved can be demonstrated in Σ. Namely

1.) if $\sim M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}]$ then by the rules of extensionality the first member becomes equivalent to $M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}]$ hence false, hence the implication is true.

2.) If however $M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}]$ then the implication becomes equivalent to $M[\mathfrak{a}, \mathfrak{f}, \mathfrak{y}] \supset M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \cdot M[\mathfrak{a}, \mathfrak{f}, \mathfrak{y}]$.

But $M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}]$ implies this owing to the formula $A \supset (B \supset A \cdot B)$.

$\sim M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset (\Phi \equiv [M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset \ldots])$ demonstrable, hence
$\sim M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset \Phi$ demonstrable (in Σ) because
$\sim M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset (M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset \ldots)$ demonstrable (in Σ)
$M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset (\Phi \equiv \{M[\mathfrak{a}, \mathfrak{f}, \mathfrak{y}] \supset M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \cdot M[\mathfrak{a}, \mathfrak{f}, \mathfrak{y}]\})$ hence
$M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset \Phi$ demonstrable (in Σ) because
$M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \supset \{M[\mathfrak{a}, \mathfrak{f}, \mathfrak{y}] \supset M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}] \cdot M[\mathfrak{a}, \mathfrak{f}, \mathfrak{y}]\}$

II. The next axiom I want to prove is

$$A \& B \rightarrow B \& A$$

So let $A = M[\mathfrak{a}, \mathfrak{f}, \mathfrak{x}]$ ⟦and⟧ $B = N[\mathfrak{b}, \mathfrak{g}, \mathfrak{y}]$.

93

$A \& B \rightarrow B \& A$ becomes the following expression[50]

$$M[\mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{q}_1}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}, \underline{\mathfrak{y}}, \underline{\mathfrak{x}})] \cdot N[\mathfrak{b}, \underline{\mathfrak{g}}, \overline{\mathfrak{q}_2}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}, \underline{\mathfrak{y}}, \underline{\mathfrak{x}})] \supset N[\mathfrak{b}, \overline{\mathfrak{p}_1}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}), \underline{\mathfrak{y}}] \cdot M[\mathfrak{a}, \overline{\mathfrak{p}_2}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}), \underline{\mathfrak{x}}]$$

Now if you put $\varrho_i(\mathfrak{a}; \mathfrak{b})$ for $\mathfrak{q}_i$ and $\sigma_i(\mathfrak{a}; \mathfrak{b})$ for $\mathfrak{p}_i$ and define $\sigma_1, \sigma_2, \varrho_1, \varrho_2$ by

[49] See p. 66 of the lecture notes.

[50] Here Gödel has erroneously written the consequent of the implication as $N[\mathfrak{a}, \overline{\mathfrak{p}_1}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}), \underline{\mathfrak{y}}] \cdot M[\mathfrak{b}, \overline{\mathfrak{p}_2}(\underline{\mathfrak{f}}, \underline{\mathfrak{g}}), \underline{\mathfrak{x}}]$.

$$\begin{cases} \sigma_1(\mathfrak{a};\mathfrak{b})(\mathfrak{f},\mathfrak{g}) \doteq \mathfrak{g} \\ \sigma_2(\mathfrak{a};\mathfrak{b})(\mathfrak{f},\mathfrak{g}) \doteq \mathfrak{f} \\ \varrho_1(\mathfrak{a};\mathfrak{b})(\mathfrak{f},\mathfrak{g},\mathfrak{y},\mathfrak{x}) \doteq \mathfrak{x} \\ \varrho_2(\mathfrak{a};\mathfrak{b})(\mathfrak{f},\mathfrak{g},\mathfrak{y},\mathfrak{x}) \doteq \mathfrak{y} \end{cases}$$

Then this implication takes on the form $R \cdot S \supset S \cdot R$ (since $A\&B = M[\mathfrak{a},\bar{\mathfrak{f}},\underline{\mathfrak{x}}].N[\mathfrak{b},\bar{\mathfrak{g}},\underline{\mathfrak{y}}]$ [[and]] $B\&A = N[\mathfrak{b},\bar{\mathfrak{g}},\underline{\mathfrak{y}}] \cdot M[\mathfrak{a},\bar{\mathfrak{f}},\underline{\mathfrak{x}}]$).

III. $W \rightarrow B$. So let $B = N[\mathfrak{b},\bar{\mathfrak{g}},\underline{\mathfrak{y}}]$. But here the implication is of the first kind therefore $W \rightarrow B = W \supset B$ (as rem[[arked]] before).

94

Hence $W \rightarrow A$ is the expression $0 = 1 \supset N[\mathfrak{a},\bar{\mathfrak{g}},\underline{\mathfrak{y}}]$.

That can be asserted in $\overline{\Sigma}$ if and only if $0 = 1 \supset N[\mathfrak{a},\sigma(\mathfrak{a}),\mathfrak{y}]$[51] can be asserted in Σ for appropriate constants σ [[addition: completely arbitrary]]. But $0 = 1 \supset \mathfrak{A}$ for any expression $\mathfrak{A}$; hence [[the proof is]] finished.

These are all axioms concerning the notions $\rightarrow$ $\&$ only.

Now let us check the rules of inference concerning [[implication]]. [[Addition: rule of export and import p. [[97]] bottom.]] These are

$$\frac{A \qquad A \rightarrow B}{B} \qquad\qquad \frac{A \rightarrow B \qquad B \rightarrow C}{A \rightarrow C} \qquad\qquad \frac{A\&B \rightarrow C}{A \rightarrow (B \rightarrow C)}$$

and vice versa

Let's assume in all these cases that $A = M[\mathfrak{a},\bar{\mathfrak{f}},\underline{\mathfrak{x}}]$, $B = N[\mathfrak{b},\bar{\mathfrak{g}},\underline{\mathfrak{y}}]$ [[and]] $C = K[\mathfrak{c},\bar{\mathfrak{h}},\underline{\mathfrak{z}}]$, where $\mathfrak{a}$, $\mathfrak{f}$, $\mathfrak{x}$ etc. are all the existential and universal variables

95

occurring in the expressions respectively.

Then

$$A \rightarrow B = M[\mathfrak{a},\underline{\mathfrak{f}},\bar{\mathfrak{q}}(\underline{\mathfrak{f}},\underline{\mathfrak{y}})] \supset N[\mathfrak{b},\bar{\mathfrak{p}}(\underline{\mathfrak{f}}),\underline{\mathfrak{y}}]$$

$$B \rightarrow C = N[\mathfrak{b},\underline{\mathfrak{g}},\bar{\mathfrak{q}}'(\underline{\mathfrak{g}},\underline{\mathfrak{z}})] \supset K[\mathfrak{c},\bar{\mathfrak{p}}'(\underline{\mathfrak{g}}),\underline{\mathfrak{z}}]$$

$$A \rightarrow C = M[\mathfrak{a},\underline{\mathfrak{f}},\bar{\mathfrak{q}}''(\underline{\mathfrak{f}},\underline{\mathfrak{z}})] \supset K[\mathfrak{c},\bar{\mathfrak{p}}''(\underline{\mathfrak{f}}),\underline{\mathfrak{z}}]$$

[51] Gödel has originally written $M[\mathfrak{a},\sigma(\mathfrak{a}),\mathfrak{y}]$.

That these formulas can be asserted in Σ means that

$$(1) \quad M[\mathfrak{a}, \mathfrak{f}, \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{y})] \supset N[\mathfrak{b}, \pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}), \mathfrak{y}]$$

$$(2) \quad N[\mathfrak{b}, \mathfrak{g}, \sigma'_{\mathfrak{b};\mathfrak{c}}(\mathfrak{g}, \mathfrak{z})] \supset K[\mathfrak{c}, \pi'_{\mathfrak{b};\mathfrak{c}}(\mathfrak{g}), \mathfrak{z}]$$

can be asserted in Σ for certain constants ⟦$\pi, \pi', \sigma,\ \sigma'$⟧.

IV. Now assume A ⟦is⟧ assertable in $\overline{\Sigma}$[52] ⟦and⟧ $A \to B$ ⟦is⟧ assertable in $\overline{\Sigma}$. ⟦Thus⟧ $M[\mathfrak{a}, \varrho_{\mathfrak{a}}, \mathfrak{x}]$ ⟦is⟧ assertable in Σ and $M[\mathfrak{a}, \mathfrak{f}, \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{y})] \supset N[\mathfrak{b}, \pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}), \mathfrak{y}]$ ⟦is⟧ assertable in Σ for some constants ϱ, σ, π.

Then by rule of substitution: $\begin{cases} \varrho_{\mathfrak{a}} & \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f},\mathfrak{y}) \\ \mathfrak{f} & \mathfrak{x} \end{cases}$

$$(1) : M[\mathfrak{a}, \varrho_{\mathfrak{a}}, \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{y})] \text{ in } \Sigma$$

$$(2) : M[\mathfrak{a}, \varrho_{\mathfrak{a}}, \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{y})] \supset N[\mathfrak{b}, \underbrace{\pi_{\mathfrak{a};\mathfrak{b}}(\varrho_{\mathfrak{a}})}_{=_{Df}\varrho'(b)}, \mathfrak{y}] \text{ in } \Sigma$$

⟦are⟧ assertable in Σ,

96

hence also $N(\mathfrak{b}, \bar{\mathfrak{f}}, \underline{\mathfrak{y}})$ ⟦is assertable⟧ by rule of existential quantifier [since the variable $\mathfrak{y}$ doesn't occur in the terms to be replaced by existential variable they can be changed into universal variables].

V. Now assume $A \to B, B \to C$ ⟦are⟧ assertable in $\overline{\Sigma}$ i.e. these ⟦are assertable⟧ in Σ for certain constants $\pi, \pi', \sigma,\ \sigma'$.

Now substitute $\begin{cases} \sigma'_{b;c}(\mathfrak{g},\mathfrak{z}) \\ \mathfrak{y} \end{cases}$ in (1) and then $\begin{cases} \pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}) \\ \mathfrak{g} \end{cases}$ in (2).[53] Then the middle term N becomes identical in the first and second premiss hence by the rule of syllogism

$$M[\mathfrak{a}, \mathfrak{f}, \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \sigma'_{\mathfrak{b};\mathfrak{c}}(\pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}), \mathfrak{z}))] \supset K[\mathfrak{c}, \pi'_{\mathfrak{b};\mathfrak{c}}(\pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f})), \mathfrak{z}]$$

Now, we can find constants μ, ξ satisfying

97

the equalities $\mu_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f}) \doteq \pi'_{\mathfrak{b};\mathfrak{c}}(\pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}))$ and $\xi_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f}, \mathfrak{z}) \doteq \sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \sigma'_{\mathfrak{b};\mathfrak{c}}(\pi_{\mathfrak{a},\mathfrak{b}}(\mathfrak{f}), \mathfrak{z}))$.

[52] Gödel has here written, "assertable in Σ", but this is clearly a mistake.

[53] Gödel has written "in (1) ⟦and⟧ (2)" for the second substitution, although the substitution only applies to (2).

Hence by rule of existential quantification, putting $\genfrac{}{}{0pt}{}{\overline{\mathfrak{q}}}{\xi_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}}\quad\genfrac{}{}{0pt}{}{\overline{\mathfrak{p}}}{\mu_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}}$ and changing $\mathfrak{z}$ (which doesn't occur) in $\underline{\mathfrak{z}}$ and $\mathfrak{f}$ in $\underline{\mathfrak{f}}$, we obtain exactly the formula to be proved.

Actually, the $\mathfrak{a},\ \mathfrak{b},\ \mathfrak{c}$ should *not* be attached to $\mathfrak{q},\ \overline{\mathfrak{q}}' \dots$ etc., but first the $\mathfrak{q},\ \overline{\mathfrak{q}}'$ etc. need to be replaced by *constants*.[54]

VI. Now as to rule of export and import all you have to do is to write down the corresponding expressions in Σ and then you will find that they are, up to trivial transformation of the calculus of propositions, exactly the same expressions for the premiss and conclusion, so you have both the rule of export and of import at once without any calculation.

98

Assume $A = M[\mathfrak{a},\overline{\mathfrak{f}},\underline{\mathfrak{x}}],\ B = N[\mathfrak{b},\overline{\mathfrak{g}},\underline{\mathfrak{y}}],\ C = K[\mathfrak{c},\overline{\mathfrak{h}},\underline{\mathfrak{z}}]$

$$A\&B = M[\mathfrak{a},\overline{\mathfrak{f}},\underline{\mathfrak{x}}].N[\mathfrak{b},\overline{\mathfrak{g}},\underline{\mathfrak{y}}]$$
$$C = K[\mathfrak{c},\overline{\mathfrak{h}},\underline{\mathfrak{z}}]$$

(I) $(A\&B)\rightarrow C = \{M[\mathfrak{a},\underline{\mathfrak{f}},\overline{\mathfrak{q}}_{\mathfrak{a}}(\underline{\mathfrak{f}},\underline{\mathfrak{g}},\underline{\mathfrak{z}})].N[\mathfrak{b},\underline{\mathfrak{g}},\overline{\mathfrak{q}}'_{\mathfrak{b}}(\underline{\mathfrak{f}},\underline{\mathfrak{g}},\underline{\mathfrak{z}})]\}\supset K[\mathfrak{c},\overline{\mathfrak{p}}_{\mathfrak{c}}(\underline{\mathfrak{f}},\underline{\mathfrak{g}}),\underline{\mathfrak{z}}]$

$$B\rightarrow C = N[\mathfrak{b},\underline{\mathfrak{g}},\overline{\mathfrak{q}}''(\underline{\mathfrak{g}},\underline{\mathfrak{z}})]\supset K[\mathfrak{c},\overline{\mathfrak{p}}'(\underline{\mathfrak{g}}),\underline{\mathfrak{z}}]$$

(II) $A\rightarrow(B\rightarrow C) =$

$$M[\mathfrak{a},\underline{\mathfrak{f}},\overline{\mathfrak{q}}'''_{\mathfrak{a}}(\underline{\mathfrak{f}},\underline{\mathfrak{g}},\underline{\mathfrak{z}})]\supset\{N[\mathfrak{b},\underline{\mathfrak{g}},\overline{\mathfrak{p}}''_{\mathfrak{b}}(\underline{\mathfrak{f}})(\underline{\mathfrak{g}},\underline{\mathfrak{z}})]\supset K[\mathfrak{c},\overline{\mathfrak{p}}'''_{\mathfrak{c}}(\underline{\mathfrak{f}})(\underline{\mathfrak{g}}),\mathfrak{z}]\}$$

Now define

$$\overline{\mathfrak{p}}''_{\mathfrak{b}}(\underline{\mathfrak{f}})(\underline{\mathfrak{g}},\underline{\mathfrak{z}}) =_{Df} \overline{\mathfrak{q}}'_{\mathfrak{b}}(\underline{\mathfrak{f}},\underline{\mathfrak{g}},\underline{\mathfrak{z}});$$
$$\overline{\mathfrak{p}}'''_{\mathfrak{c}}(\underline{\mathfrak{f}})(\mathfrak{g}) =_{Df} \overline{\mathfrak{p}}_{\mathfrak{c}}((\underline{\mathfrak{f}},\underline{\mathfrak{g}}))$$

VII. Rule of universal quantifier

$$\frac{A\rightarrow B}{A\rightarrow(\Pi x)B}$$

I am reminding you of the definition of $(\Pi x)A$. Let $\overline{\mathfrak{f}}$ be the series of all existential variables of A in the order of their occurrence. Associate with each another existential variable of appropriate type. Let their series be $\overline{\mathfrak{p}}$. Then

$$(\Pi x)A = A\left(\genfrac{}{}{0pt}{}{x}{\underline{x}}\ \genfrac{}{}{0pt}{}{\overline{\mathfrak{f}}}{\overline{\mathfrak{p}}(\underline{x})}\right).$$

[54] Die $\mathfrak{a},\ \mathfrak{b},\ \mathfrak{c}$ sollten eigentlich *nicht* an $\mathfrak{q},\ \overline{\mathfrak{q}}' \dots$ etc. angehängt werden, sondern erst die $\mathfrak{q},\ \overline{\mathfrak{q}}'$ etc. durch *Konstanten* ersetzt werden.

99

$$A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$

We can assume B contains x as a free variable, hence $B = N[x, \mathfrak{b}, \overline{\mathfrak{g}}, \underline{\mathfrak{y}}]$, where $x \sim\varepsilon\, (\mathfrak{a}; \mathfrak{b})$. Then

(1) $\quad A \rightarrow B = M[\mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{y}})] \supset N[x, \mathfrak{b}, \overline{\mathfrak{p}}(\underline{\mathfrak{f}}), \underline{\mathfrak{y}}]$

$$(\Pi x)B = N[\underline{x}, \mathfrak{b}, \overline{\mathfrak{h}}(\underline{x}), \underline{\mathfrak{y}}]$$

(2) $\quad A \rightarrow (\Pi x)B = M[\mathfrak{a}, \bar{\mathfrak{f}}, \overline{\mathfrak{q}}'(\underline{\mathfrak{f}}, \underline{\mathfrak{y}}, x)] \supset N[\underline{x}, \mathfrak{b}, \overline{\mathfrak{p}}'(\underline{\mathfrak{f}})(\underline{x}), \underline{\mathfrak{y}}]$

Hence the corresponding formulas in Σ which are equipollent are

$$M[\mathfrak{a}, \mathfrak{f}, \varrho_{\mathfrak{a};\mathfrak{b};x}(\mathfrak{f}, \mathfrak{y})] \supset N[x, \mathfrak{b}, \sigma_{\mathfrak{a};\mathfrak{b};x}(\mathfrak{f}), \mathfrak{y}]$$

$$M[\mathfrak{a}, \mathfrak{f}, \varrho'_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}, \mathfrak{y}, x)] \supset N[x, \mathfrak{b}, \sigma'_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f})(x), \mathfrak{y}]$$

But the two expressions are the same except that the argument of ϱ and σ appear in another order in

100

the second than in the first and therefore it is easily seen that if the first can be proved in Σ for some constants ϱ, σ, the second can also be proved for some other constants ϱ', σ' defined in terms of these.

VIII. $(x)A \rightarrow A\binom{x}{t}$ for any term t ⟦addition: can be proved more easily from the simple substitution rule and invertibility of VII⟧.

If A doesn't contain x, trivial. So let be:

$$A = M[x, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$

$t = T[\mathfrak{b}]$ (t contains no other free variables)

Then

$$(x)A = M[\underline{x}\ \mathfrak{a}, \overline{\mathfrak{G}}(\underline{x}), \underline{\mathfrak{x}}]$$

$$A\tbinom{x}{t} = M[T[\mathfrak{b}], \mathfrak{a}, \bar{\mathfrak{f}}, \mathfrak{x}]$$

$$(x)A \rightarrow A\tbinom{x}{t} =$$

$$M[\underset{\varrho_{\mathfrak{a};\mathfrak{b}}}{\overline{\mathfrak{Q}}(\underline{\mathfrak{G}}, \underline{\mathfrak{x}})},\ \mathfrak{a},\ \underset{\varrho_{\mathfrak{a};\mathfrak{b}}}{\underline{\mathfrak{G}}(\overline{\mathfrak{Q}}(\underline{\mathfrak{G}}, \underline{\mathfrak{x}}))},\ \underset{\varrho'_{\mathfrak{a};\mathfrak{b}}}{\overline{\mathfrak{Q}}'(\underline{\mathfrak{G}}, \underline{\mathfrak{x}})}] \supset M[T[\mathfrak{b}], \mathfrak{a}, \underset{\sigma_{\mathfrak{a};\mathfrak{b}}}{\overline{\mathfrak{P}}(\underline{\mathfrak{G}})}, \underline{\mathfrak{x}}]$$

101

$\varrho_{\mathfrak{a};\mathfrak{b}}(\mathfrak{G},\mathfrak{x}) \doteq T[\mathfrak{b}]$
$\sigma_{\mathfrak{a};\mathfrak{b}}(\mathfrak{G}) \doteq \mathfrak{G}(T[\mathfrak{b}])$
$\varrho'_{\mathfrak{a};\mathfrak{b}}(\mathfrak{G},\mathfrak{x}) \doteq \mathfrak{x}$

IX. Finally: Rule of Induction

$$\frac{\begin{array}{l} A\binom{x}{0} \\ A \rightarrow A\binom{x}{\nu(x)} \end{array}}{A}$$

⟦where⟧ $A = M[x, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$ $\qquad x \sim\varepsilon\, \mathfrak{a}$

$$\frac{A\binom{x}{0} = M[0, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]}{A\binom{x}{\nu(x)} = M[\nu(x), \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]}$$

⟦is⟧ assertable in Σ by assumption;

$$A \rightarrow A\binom{x}{\nu(x)} = M[x, \mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{Q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{x}})] \supset M[\nu(x), \mathfrak{a}, \overline{\mathfrak{P}}(\underline{\mathfrak{f}}), \underline{\mathfrak{x}}]$$

⟦is⟧ assertable in Σ by assumption, i.e. for certain constants α, ϱ, σ

102

(1) $M[0, \mathfrak{a}, \alpha(\mathfrak{a}), \mathfrak{x}]$ ⟦is assertable⟧ in Σ

(2) $M[x, \mathfrak{a}, \mathfrak{f}, \varrho_{x,\mathfrak{a}}(\mathfrak{f}, \mathfrak{x})] \supset M[\nu(x), \mathfrak{a}, \sigma_{x,\mathfrak{a}}(\mathfrak{f}), \mathfrak{x}]$ ⟦is assertable in Σ⟧

One can find a series of constants γ for which

$$\begin{cases} \gamma(0)(\mathfrak{a}) \doteq \alpha(\mathfrak{a}) \\ \gamma(\nu(x))(\mathfrak{a}) \doteq \sigma_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a})) \end{cases}$$

can be proved in Σ.

By induction ⟦the following are⟧ demonstrable:

in (1): $\binom{\alpha(\mathfrak{a})}{\gamma(0)(\mathfrak{a})}$ $\qquad M[0, \mathfrak{a}, \gamma(0)(\mathfrak{a}), \mathfrak{x}]$

in (2): $\binom{\mathfrak{f}}{\gamma(x)(\mathfrak{a})}$ $\qquad M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \varrho_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a}), \mathfrak{x})] \supset M[\nu(x), \mathfrak{a}, \gamma(\nu(x))(\mathfrak{a}), \mathfrak{x}]$

in (2): $\binom{\sigma_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a}))}{\gamma(\nu(x))(\mathfrak{a})}$

But the second member of the implication is obtained from $M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \mathfrak{x}]$ exactly by replacing x by $\nu(x)$ and the first by replacing some of the other variables, namely the term $\mathfrak{x}$, by some terms (it makes no difference which ones).

102′[55]

⟦Assume that⟧ it has been proven[56] that in Σ: $M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \mathfrak{x}]$. For an expression A in the variables $\mathfrak{a}, x, \mathfrak{x}$, $A(x, \mathfrak{a}, \mathfrak{x})$ is given, namely: $M(x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \mathfrak{x})$. For a proof we have:

1. $\underline{A(0, \mathfrak{a}, \mathfrak{x})}$
2. For certain functions $F_{x,\mathfrak{a}}(\mathfrak{x}) = \varrho_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a}), \mathfrak{x})$ it is proven ⟦that⟧

   $$M(x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), F_{x,\mathfrak{a}}(\mathfrak{x})) \supset M(\nu(x), \mathfrak{a}, \gamma(\nu(x))(\mathfrak{a}), \mathfrak{x}),$$

   that is,

   $$A(x, \mathfrak{a}, F_{x;\mathfrak{a}}(\mathfrak{x})) \supset A(\nu(x), \mathfrak{a}, \mathfrak{x})$$

---

[55] This page has been written in shorthand German.

[56] In Σ: $M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \mathfrak{x}]$ bewiesen. Dann:

Denn es ist ein Ausdruck A in den Variablen $\mathfrak{a}, x, \mathfrak{x}$ gegeben: $A(x, \mathfrak{a}, \mathfrak{x})$, nämlich $M[x, \mathfrak{a}, \gamma(x)\mathfrak{a}, \mathfrak{x}]$.

Für den Beweis ist

1. $\underline{A(0, \mathfrak{a}, \mathfrak{x})}$
2. Für gewisse Funktionen $F_{x,\mathfrak{a}}(\mathfrak{x}) = \varrho_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a}), \mathfrak{x})$ ist bewiesen $M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), F_{x,\mathfrak{a}}(\mathfrak{x})] \supset M[\nu(x), \mathfrak{a}, \gamma(\nu(x))(\mathfrak{a}), \mathfrak{x})]$ i.e. $A(x, \mathfrak{a}, F_{x;\mathfrak{a}}(\mathfrak{x}) \supset A(\nu(x), \mathfrak{a}, \mathfrak{x})$.

Also folgt $A(x, \mathfrak{a}, \mathfrak{x})$ d.h. $\underline{M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \mathfrak{x}]}$. <u>q.e.d.</u>

<u>Bemerkung:</u> Statt $\gamma(x)(\mathfrak{a})$ sollte überall stehen $\gamma(x, \mathfrak{a})$. Dann ist das Rekursionsschema

$$\underline{\gamma(0, \mathfrak{a}) = \alpha(\mathfrak{a})}$$
$$\underline{\gamma(x+1, \mathfrak{a}) = \beta[\gamma(x, \mathfrak{a}), x, \mathfrak{a}]}$$

genau erfüllt für $\beta(\mathfrak{g}, x, \mathfrak{a}) = \sigma_{x,\mathfrak{a}}(\mathfrak{g})$. Aber wenn das Rekursionsschema ohne Parameter $\mathfrak{a}$ verwendet werden soll, dann müsste man definieren

$$\underline{\gamma(0) = \alpha}$$
$$\underline{H(x, \mathfrak{g})(\mathfrak{a}) = \sigma_{x,\mathfrak{a}}(\mathfrak{g}(\mathfrak{a}))}$$
$$\underline{\gamma(x+1) = H(x, \gamma(x))}$$

denn daraus folgt

$$\gamma(x+1)(\mathfrak{a}) = \sigma_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a}))$$

Thus it follows that $A(x, \mathfrak{a}, \mathfrak{x})$ i.e. $\underline{M[x, \mathfrak{a}, \gamma(x)(\mathfrak{a}), \mathfrak{x}]}$. *q.e.d.*

Note: Instead of $\gamma(x)(\mathfrak{a})$ we should have everywhere $\gamma(x, \mathfrak{a})$. Then the recursion scheme

$$\underline{\gamma(0, \mathfrak{a}) = \alpha(\mathfrak{a})}$$
$$\underline{\gamma(x + 1, \mathfrak{a}) = \beta[\gamma(x, \mathfrak{a}), x, \mathfrak{a}]}$$

is precisely satisfied for $\beta(\mathfrak{g}, x, \mathfrak{a}) = \sigma_{x,\mathfrak{a}}(\mathfrak{g})$. However, if one applies the recursion scheme without the parameter $\mathfrak{a}$, then one must define

$$\underline{\gamma(0) = \alpha}$$
$$\underline{H(x, \mathfrak{g})(\mathfrak{a}) = \sigma_{x,\mathfrak{a}}(\mathfrak{g}(\mathfrak{a}))}$$
$$\underline{\gamma(x + 1) = H(x, \gamma(x))}$$

As it follows from there that:

$$\gamma(x + 1)(\mathfrak{a}) = \sigma_{x,\mathfrak{a}}(\gamma(x)(\mathfrak{a}))$$

103

Axioms

1. $A \rightarrow A \vee B$ (103)

3. $A\binom{x}{t} \rightarrow (\exists x)A$ (105) becomes superfluous by the inverse of 4

$(A \rightarrow B \vee A)$

Rules of inference

2.[57]

$$\frac{A \rightarrow C \qquad B \rightarrow C}{A \vee B \rightarrow C} \quad 104$$

4.

$$\frac{A \rightarrow B}{(\exists x)A \rightarrow B} \quad 106$$

$$A = M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$
$$B = N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}]$$
$$C = K[\mathfrak{c}, \bar{\mathfrak{h}}, \underline{\mathfrak{z}}]$$

1.

$$M[\mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \rightarrow N[\mathfrak{b}, \bar{\mathfrak{g}}, \underline{\mathfrak{y}}] \cdot 0 = \overline{u} \vee M[\mathfrak{a}, \bar{\mathfrak{h}}, \underline{\mathfrak{z}}] \cdot 1 = \overline{u}$$

---

[57] ⟦Added:⟧ Instead of this also: axiom $(A \vee A) \supset A$ (does *not* hold for the simple v!!) and axiom $(A \vee B) \supset (B \vee A)$. Rule: $A \supset B$ ⟦apparently meant to signify Modus Ponens⟧.

$$M[\mathfrak{a},\underline{\mathfrak{f}},\overline{\mathfrak{q}}(\underline{\mathfrak{f}},\underline{\mathfrak{y}},\underline{\mathfrak{z}})] \supset N[\mathfrak{b},\overline{\mathfrak{p}_1}(\underline{\mathfrak{f}}),\underline{\mathfrak{y}}]\cdot 0=\overline{\mathfrak{p}}(\underline{\mathfrak{f}}) \ \vee \ M[\mathfrak{a},\overline{\mathfrak{p}_2}(\underline{\mathfrak{f}}),\underline{\mathfrak{z}}]\cdot 1=\overline{\mathfrak{p}}(\underline{\mathfrak{f}})$$

$$M[\mathfrak{a},\mathfrak{f},\varrho_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f},\mathfrak{y},\mathfrak{z})] \supset N[\mathfrak{b},\sigma^{(1)}_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}),\mathfrak{y}]\cdot 0=\sigma(\mathfrak{f})\vee M[\mathfrak{a},\sigma^{(2)}_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}),\mathfrak{z}]\cdot 1=\sigma(\mathfrak{f})$$

$$\begin{cases}\sigma(\mathfrak{f})=1 \qquad\qquad \sigma^{(1)}_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}) \text{ arbitrary}\\ \sigma^{(2)}(\mathfrak{a};\mathfrak{b})(\mathfrak{f})\doteq\mathfrak{f}\\ \varrho(\mathfrak{a};\mathfrak{b})(\mathfrak{f},\mathfrak{y},\mathfrak{z})\doteq\mathfrak{z}\end{cases}$$

104

2.

$$M[\mathfrak{a},\underline{\mathfrak{f}},\overline{\mathfrak{q}}^{(1)}(\underline{\mathfrak{f}},\underline{\mathfrak{z}})] \supset K[\mathfrak{c},\overline{\mathfrak{p}}^{(1)}(\underline{\mathfrak{f}}),\underline{\mathfrak{z}}]$$
$$N[\mathfrak{b},\underline{\mathfrak{g}},\overline{\mathfrak{q}}^{(2)}(\underline{\mathfrak{g}},\underline{\mathfrak{z}})] \supset K[\mathfrak{c},\overline{\mathfrak{p}}^{(2)}(\underline{\mathfrak{g}}),\underline{\mathfrak{z}}]$$

i.e.,

$$M[\mathfrak{a},\mathfrak{f},\varrho^{(1)}_{\mathfrak{a};\mathfrak{c}}(\mathfrak{f},\mathfrak{z})] \supset K[\mathfrak{c},\sigma^{(1)}_{\mathfrak{a};\mathfrak{c}}(\mathfrak{f}),\mathfrak{z}]$$
$$N[\mathfrak{b},\mathfrak{g},\varrho^{(2)}_{\mathfrak{b};\mathfrak{c}}(\mathfrak{g},\mathfrak{z})] \supset K[\mathfrak{c},\sigma^{(2)}_{\mathfrak{a};\mathfrak{c}}(\mathfrak{g}),\mathfrak{z}]$$

$$M[\mathfrak{a},\bar{\mathfrak{f}},\underline{\mathfrak{x}}]\cdot 0=\overline{u}\vee N[\mathfrak{b},\overline{\mathfrak{g}},\underline{\mathfrak{y}}]\cdot 1=\overline{u}\rightarrow K[\mathfrak{c},\overline{\mathfrak{h}},\underline{\mathfrak{z}}]$$

$$M[\mathfrak{a},\underline{\mathfrak{f}},\overline{\mathfrak{q}}^{(1)}(\underline{\mathfrak{f}},\underline{u},\underline{\mathfrak{g}},\underline{\mathfrak{z}})]\cdot 0=\underline{u}\vee N[\mathfrak{b},\underline{\mathfrak{g}},\overline{\mathfrak{q}}^{(2)}(\underline{\mathfrak{f}},\underline{u},\underline{\mathfrak{g}},\underline{\mathfrak{z}})]\cdot 1=\underline{u}\rightarrow K[\mathfrak{c},\overline{\mathfrak{p}}(\underline{\mathfrak{f}},\underline{u},\underline{\mathfrak{g}}),\underline{\mathfrak{z}}]$$

$$M[\mathfrak{a},\mathfrak{f},\alpha^{(1)}_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g},\mathfrak{z})]\cdot 0=u\vee N[\mathfrak{b},\mathfrak{g},\alpha^{(2)}_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g},\mathfrak{z})]\cdot 1=u\supset K[\mathfrak{c},\beta_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g}),\mathfrak{z}]$$

$$\beta_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g})\doteq\sigma^{(1)}_{u;\mathfrak{c}}(\mathfrak{f}) \text{ if } u=0$$
$$\beta_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g})\doteq\sigma^{(2)}_{u;\mathfrak{c}}(\mathfrak{f}) \text{ if } u\neq 0$$

$\beta=$ the second because $u\neq^{0}_{1}$ i.e. $u=1\supset\ldots$

$$\alpha^{(1)}_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g},\mathfrak{z})\doteq\varrho^{(1)}_{u;\mathfrak{c}}(\mathfrak{f},\mathfrak{z})$$
$$\alpha^{(2)}_{\mathfrak{a};\mathfrak{b};\mathfrak{c}}(\mathfrak{f},u,\mathfrak{g},\mathfrak{z})\doteq\varrho^{(2)}_{u;\mathfrak{c}}(\mathfrak{g},\mathfrak{z})$$

$u=0\rightarrow(\Phi\equiv$ ⟦line ends here⟧

105

3.

$$A\binom{x}{t} \to (\Sigma x)A = \Phi \qquad \begin{aligned} A &= M[x, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \\ t &= T[\mathfrak{b}] \end{aligned}$$

$$A\binom{x}{t} = M[T[\mathfrak{b}], \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$
$$(\Sigma x)A = M[\overline{x}, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}]$$

$$\Phi = M[T[\mathfrak{b}], \mathfrak{a}, \underline{\mathfrak{f}}, \overline{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{x}})] \supset M[\overline{\mathfrak{h}}(\underline{\mathfrak{f}}), \mathfrak{a}, \overline{\mathfrak{p}}(\underline{\mathfrak{f}}), \underline{\mathfrak{x}}]$$

$$M[T[\mathfrak{b}], \mathfrak{a}, \mathfrak{f}, \sigma(\mathfrak{a}; \mathfrak{b})(\mathfrak{f})(\mathfrak{x})] \supset M[\pi_{\mathfrak{a};\mathfrak{b}}(\mathfrak{f}), \mathfrak{a}, \varrho(\mathfrak{a}; \mathfrak{b})(\mathfrak{f}), \mathfrak{x}]$$

$$\pi(\mathfrak{a}; \mathfrak{b})(\mathfrak{f}) \doteq T[\mathfrak{b}]$$
$$\varrho(\mathfrak{a}; \mathfrak{b})(\mathfrak{f}) \doteq \mathfrak{f}$$
$$\sigma(\mathfrak{a}; \mathfrak{b})(\mathfrak{f}, \mathfrak{x}) \doteq \mathfrak{x}$$

106

4.

$$\frac{A \to B}{(\Sigma x)A \to B} \qquad \begin{aligned} (\Sigma x)A &= M[\overline{x}, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \\ A &= M[x, \mathfrak{a}, \bar{\mathfrak{f}}, \underline{\mathfrak{x}}] \\ B &= N[\mathfrak{b}, \overline{\mathfrak{g}}, \underline{\mathfrak{y}}] \end{aligned}$$

$$A \to B = M[x, \mathfrak{a}, \underline{\mathfrak{f}}, \underbrace{\overline{\mathfrak{q}}(\underline{\mathfrak{f}}, \underline{\mathfrak{y}})}_{\varrho(\mathfrak{a};\mathfrak{b},x)}] \supset N[\mathfrak{b}, \underbrace{\overline{\mathfrak{p}}(\underline{\mathfrak{f}})}_{\sigma(\mathfrak{a};\mathfrak{b},x)}, \underline{\mathfrak{y}}]$$

$$(\Sigma x)A \to B = M[\underline{x}, \mathfrak{a}, \underline{\mathfrak{f}}, \underbrace{\overline{\mathfrak{q}}'(\underline{\mathfrak{f}}_1, \underline{x}, \underline{\mathfrak{f}}_2, \underline{\mathfrak{y}})}_{\varrho'(a;b)}] \supset N[\mathfrak{b}, \underbrace{\overline{\mathfrak{p}}'(\underline{\mathfrak{f}}_1, \underline{x}, \underline{\mathfrak{f}_2})}_{\sigma'(\mathfrak{a};\mathfrak{b})}, \underline{\mathfrak{y}}]$$

Notes on recursive functions[58]

107′[59]

A shorter outline of the following:

1. General recursive functions (not everywhere defined).
2. Numbers associated with a general recursive function (not unique) but
3. With each type t associated a class of integers $\varphi(t)$.

Definition by recursion.

1. $\varphi(I) =$ all integers.
2. Denote by $\tau(t_1, \ldots, t_k)$ the type whose complete argument series is of type $t_1, \ldots, t_k$.

Then $\tau(t_1, \ldots, t_k)$ is the class of numbers of recursive functions f with k arguments $f(x_1, \ldots, x_k)$ such that f ⟦is⟧ defined whenever $x_i \,\varepsilon\, \varphi(t_i)$ and f ⟦is⟧ extensional i.e. ⟦if⟧ $x_i \sim x_i'$, then $f(\ldots x_i \ldots) = f(\ldots x_i' \ldots)$ where $\sim$ is also defined by induction on the levels, i.e.,

$$x_i \underset{\tau(t_1 \ldots t_k)}{\sim} x_i' \equiv (y_1, \ldots, y_k)(f_{x_i}(y_1, \ldots, y_k) = f_{x_i'}(y_1, \ldots, y_k))$$

I call $x \,\varepsilon\, \varphi(t)$ a number of type t.

4. Interpretation of $\overline{\Sigma}$:

1. Function of type $t =$ number $\varepsilon\, \varphi(t)$.
2. Application $n(m)$ is a certain recursive function whose definition[60]
3. Constants (Greek letters) = certain numbers satisfying the definitional postulates (whose existence demonstrable).

---

[58] The third part of the Princeton lectures consists of 13 loose pages filed with the two first notebooks.

[59] This page follows the page 107 in the notes. However, it seems to be a summary of the introduction to recursive functions that Gödel refers to as what was done “last time” on p. 107.

[60] Text ends here; Gödel returns to the definition of application on p. 110 of the lecture notes.

5. Then demonstrable:

   1. If A is an axiom of intuitionistic number theory then A' true for this model M.

107

1. **Last time** I mentioned the notion of a partially recursive function of integers. This is a function $f(x)$ with one variable definable by a system of equalities which allow to calculate $f(x)$ for the argument den[[oted by]] x which belong to a certain class K called the domain of definition of f.

2. All partially recursive functions can be numbered by integers. Call the nth function $\underline{f_n}$ and call n the associated number of f. To each integer [[corresponds]] exactly one function, but many numbers to one integer.

3. [[Added: *Kleene*]] Perhaps good if I make a list of the facts about recursive functions which I need.

   0. Introduce once for all an operation of pairing $\langle uv \rangle$.

   Primitive recursion of two integers enumerating all pairs and triples $\langle x, y, z \rangle = \langle x, \langle y, z \rangle \rangle$. n-tuples, partially recursive functions of n variables [[are]] defined in this manner. Then

   1. There is a partially recursive function W such that $W\langle n, x \rangle = f_n(x)$.

   2. [[If]] f, g [[are]] partially recursive then $f(g(x))$ [[is]] partially recursive [[where the]] domain of definition [[is appropriate]].

   3. Term composed of partially recursive functions of special integers and of variables defines a partially recursive function.

108

   4. In particular $f\langle n, x \rangle$ is partially recursive if f is and you can find a primitive recursive function $p(k, n)$ such that if k is an associated number of f then $p(k, n)$ is an associated number of $f\langle n, \hat{x} \rangle$.

   5. If f is a partially recursive function such that for each $x \,\varepsilon\, K$ there exists a y [[such that]] $f\langle y, x \rangle = 0$, then there exists a partially recursive function $g(x)$ such that $f\langle g(x), x \rangle = 0$ for $x \,\varepsilon\, K$.

If for each $x \varepsilon K$ there ⟦exists⟧ $\langle y, x\rangle \varepsilon \mathfrak{D}(f)$ for every y, then there exists even a partially recursive function $g(x)$ which picks out for every $x \varepsilon K$ the smallest y for which this is true.

109

4. And now by means of these partially recursive functions I defined last time a function $\varphi(t)$ which associates with each type a class of integers called integers of this type and at the same time a relation or rather a set of relations (one for each type t) $\underset{t}{\sim}$ ⟦such that⟧ $x \underset{t}{\sim} y$ ⟦is⟧ defined for $x, y \varepsilon \varphi(t)$.

5. The definition went by induction on the level and implies this, that for $x \varepsilon \varphi(\tau(t_1, \ldots, t_n))$ and $y_i \varepsilon \varphi(t_i)$,

$$\begin{aligned} &f_x(\langle y_1, \ldots, y_n\rangle) \text{ is defined and} \\ = \; &f_x(\langle y'_1, \ldots, y'_n\rangle) \text{ if } y_i \underset{t_i}{\sim} y'_i. \end{aligned}$$

Call a function extensional if it satisfies ⟦this condition⟧.

Instead of $f_x, \langle y_1, \ldots, y_n\rangle$, we can write $W\langle x, \langle y_1, \ldots, y_n\rangle\rangle$. I call an integer $\varepsilon \varphi(t)$ an integer of the type t.

6. And now I want to define a model for the system Σ (and also for $\overline{\Sigma}$) such that

1. The objects of type τ will be the integers $\varepsilon \varphi(t)$.

110

2. The operation of application $A_{s,t}(m, n)$ to be defined for $n \varepsilon \varphi(t)$, $m \varepsilon \varphi(s\tau t)$ must now be defined in such a manner that $A_{s,t}(m, n) \varepsilon \varphi(s)$ and that all axioms of Σ will be satisfied.

3. Now this operation will be defined as follows:

   If $s = I \quad A_{s,t}(m, n) = W\langle m, n\rangle$

   If $s \neq I \quad A_{s,t}(m, n) =$ one of the associated numbers of the following recursive function $W\langle m, \langle n, \hat{x}\rangle\rangle$.

Now owing to one of the lemmas I quoted this function is recursive and in addition its associated number depends primitive recursively on n, m, i.e., $A_{s,t}(m, n) = V\langle m, n\rangle$ for $s \neq I$.

This is the operation of application in the model. Let's call this model M. Hence $W\langle m, \langle n, x\rangle\rangle = W\langle V\langle m, n\rangle, x\rangle$ whenever the left side ⟦is⟧ meaningful ⟦with respect to the⟧ right side and vice versa.

This equality can be generalized like this:

111

$$W\langle m, \langle x_1, x_2, x_3\rangle\rangle = W\langle V\langle V\langle m, x_1\rangle, x_2\rangle, x_3\rangle$$

because ⟦$\langle x_1, x_2, x_3\rangle$⟧$= \langle x_1, \langle x_2, x_3\rangle\rangle$

$$W\langle V\langle m, x_1\rangle, \langle x_2, x_3\rangle\rangle = W\langle V$$ [61]

and for any number of variables but this allows to prove the axioms of explicit definition of Σ for this model in this sense: If we have a term T composed of these operations $A_{s,t}$, of variables $x_1, \ldots, x_n$ of any types and of integers $n_1, \ldots, n_k$, where in the arguments of these $A_{s,t}$ only integers, not variables, of an appropriate type must occur, then there exists a number of type $\tau(x_1, \ldots, x_n)$ such that $a(x_1)(x_2)\ldots(x_n) = T$ for any numbers $x_1, \ldots, x_n$ of appropriate type, where application $a(x_1)$ is to be taken in the sense of $A_{s,t}$. The proof is immediate. At first there exists a number a such that

112

1. $W\langle a\langle x_1, \ldots, x_n\rangle\rangle = T$ because T as a term composed of recursive functions defines a recursive function of n variables.

2. The left side can be replaced by this expression, but this is exactly what we want owing to this definition of the operation of application.

That also the axiom of recursive definition holds in the same sense can be proved by a similar device. So with each Greek letter we can associate a certain number such that all recursive axioms of Σ will be satisfied for these numbers,and the other axioms and rules of Σ are likewise easily proved for this model M. Of course, universal quantification, expressed in Σ by free variables, is to be interpreted in the model as meaning to be true for all numbers of the appropriate types. This convention about the meaning of universal quantification defines under what circumstances an expression of Σ holds in the model M and in this sense all all axiom and rules are satisfied.

[61] Incomplete formula that should probably read as above, $W\langle V\langle V\langle m, x_1\rangle, x_2\rangle, x_3\rangle$.

113

Now as to the expressions of $\overline{\Sigma}$, say $N[\mathfrak{a}, \overline{f}, \underline{\mathfrak{x}}]$: We shall say that this is true for the model M if there are numbers ϱ of appropriate types such that $N[\mathfrak{a}, \varrho(\mathfrak{a}), \mathfrak{x}]$ is true in M (where this is now an expression of Σ) for which truth in M has been defined already.

Then the following things are demonstrable:

1. If A is a formal axiom of intuitionistic number theory or of Σ_I, A' ⟦is⟧ true for M.

2. If C is the result of applying a rule of inference of Σ_I to A, B and A', B' ⟦are⟧ true in M, then C' ⟦is⟧ true in M.

These two theorems are exactly the same as those I stated last time, only "demonstrable in $\overline{\Sigma}$" is replaced by "true for M," and the proofs are literally the same because the axioms of Σ are true in this model.

But 3.) we have now the following theorem:

114

3. There is a number-theoretic propositional function $\varphi(x)$ such that

   $$(\sim \underbrace{(x)[\varphi(x) \vee \sim\varphi(x)]}_{U})'$$

   true in M, namely $\varphi(x) \equiv x$ is not the number of an everywhere defined recursive function, i.e.

   $$\varphi(x) \equiv (\exists u)(v)[\sim v\, B\, x, u].$$

   $v\, B\, x, u$ (primitive recursive) is by definition: v is the number of a proof from a system of equations number x for a proposition of the form $f(u) = k$ or shortly, is a number of a computation for $f(u)$ where

$$U' = (\exists\, e\, u\, f)(v\, z\, x)[e(x) = 0 \cdot \sim v\, B\, x, u(x) \vee \underbrace{e(x) = 1 \cdot f(x)(z)\, B\, x, z]}_{\Psi(euf, xvz)=0}$$

   e, u ⟦are⟧ of type $I\tau I$ ⟦and⟧ f of type $(I\tau I)\tau I$.

   $(\sim U)' = \neg(U') \equiv (\exists\, V\, Z\, X)$

associating v, z, x with every triple of functions e, u, f such that

$$\Psi(euf, xvz) \neq 0$$

114′

Let us denote $(x)[\varphi(x) \vee \sim\varphi(x)]$[62] by U and calculate U'. For this purpose, replace

$$\varphi'(x) = (\exists u)(v)[\sim v\, B\, x, u]$$
$$\neg\varphi'(x) = (\exists f)(z)[f(z)\, B\, x, z]$$
$$\varphi'(x) \text{ v } \neg\varphi'(x) = (\exists euf)(vz) \begin{cases} e = 0 \cdot \sim v\, B\, x, u \\ \vee\, e = 1 \cdot f(z)\, B\, x, z \end{cases}$$

$$\begin{aligned} U' &= (\Pi x)[\varphi'(x) \vee \neg\varphi'(x)] \\ &= (\exists\, e\, u\, f)(v\, z\, x) \underbrace{\begin{cases} e(x) = 0 \cdot \sim v\, B\, x, u(x) \\ \vee\, e(x) = 1 \cdot f(x)(z)\, B\, x, z \end{cases}}_{\Psi(euf, xvz) = 0} \end{aligned}$$

115

That this holds for M means: There are three partial recursive and extensional functions V, Z, X which, applied to any recursive e, u, f of these types, give numbers x, v, z for which this is true. We prove

1. For any e, u, f of these types, there exist ⟦such⟧ numbers x, v, z. ⟦The proof is⟧ indirect because if for some e, u, f such numbers did not exist, then this would be true for any numbers v, z, x and e would be a recursive procedure to decide whether x is a number of an everywhere defined recursive function or not (but that doesn't exist).

2. But now Ψ is itself a recursive relation and we had the theorem that if $(x)(\exists x)R(x, y), x\, \varepsilon\, M$, then there is a recursive function f ⟦with the domain⟧

[62] Top of page has an indication for an addition that is missing. The bottom of the page reads:

$$\neg A(\overline{\mathfrak{y}}, \underline{\mathfrak{x}}) = \sim A(\underline{\mathfrak{y}}, \overline{f}(\underline{\mathfrak{y}}))$$
$$\neg(\exists \underline{\mathfrak{x}})(\overline{\mathfrak{x}})A = (\exists \overline{f})(\mathfrak{y}) \sim A(\mathfrak{y}, f(\mathfrak{y}))$$

$D(f) \supset M$ [[such that]] $R(x, f(x))$. So there exist recursive functions X, V, Z of the variables e, u, f associating these x, v, z for each e, u, f. q.e.d.

And [[there are]] even such as associate the smallest numbers x, v, z ($\langle x, v, z\rangle$ is smallest). But these functions V, Z, X will be extensional.

116

I want to make an addition to what I said last time. We had: If A is demonstrable in Σ_I, then A' is demonstrable in $\overline{\Sigma}$. If in particular A is an existential proposition $(\exists x)B(x)$ then this will allow to construct an example (because the system $\overline{\Sigma}$ is trivially constructive). The example will have the form of a constant term α of Σ. But from this argument it does not follow yet that $B(\alpha)$ is demonstrable in Σ_I, only that $B'(\alpha)$ is demonstrable in $\overline{\Sigma}$. The transformation from a proof for $B'(\alpha)$ in $\overline{\Sigma}$ to a proof of $B(\alpha)$ in Σ_I is not quite so easy; only in one case, namely where $B(\alpha)$ contains no quantifiers because then $B'(\alpha) \equiv B(\alpha)$ and $\overline{\Sigma} \subseteq \Sigma_I$. But for unquantified expressions B even more is true.

117

Namely, even if $\sim(x)B(x)$ is demonstrable in Σ_I, then $\sim B(\alpha)$ is demonstrable in Σ_I for some constant because $(\sim(x)B(x))' = ((\exists x)\sim B(x))'$ as I remarked. But now this expression contains no $\vee, \exists$, therefore it is demonstrable in Σ_I if it is demonstrable in the corresponding classical system Σ_K. So you have [[the]] theorem.

References

Ackermann, W. (1940). Zur Widerspruchsfreiheit der Zahlentheorie. *Mathematische Annalen 117*, 162–194.

van Atten, M. (2015). Gödel and intuitionism. In *Essays on Gödel's Reception of Leibniz, Husserl, and Brouwer*, pp. 189–234. Dordrecht: Springer.

Avigad, J. and S. Feferman (1998). Gödel's functional interpretation. In S. R. Buss (Ed.), *Handbook of Proof Theory*. Amsterdam: Elsevier.

Bernays, P. (1935). Sur le platonisme dans les mathématiques. *L'Enseignement Mathematique 34*, 52–69.

Diller, J. and W. Nahm (1974). Eine Variante zur Dialectica-Interpretation der Heyting-Arithmetik endlicher Typen. *Archiv für mathematische Logik und Grundlagenforschung 16*, 49–66.

Feferman, S. (1998). Gödel's Dialectica interpretation and its two-way stretch. In *In the Light of Logic*, pp. 209–225. New York: Oxford University Press.

Ferreira, G. and P. Oliva (2010). On various negative translations. In *Proceedings Third International Workshop on Classical Logic and Computation, CL&C 2010, Brno, Czech Republic, 21-22 August 2010*, pp. 21–33.

Gentzen, G. (1934-35). Untersuchungen über das logische Schliessen I/II. *Mathematische Zeitschrift 39*, 176–210, 405–431.

Gentzen, G. (1935/1974). Der erste Widerspruchsfreiheitsbeweis für die klassische Zahlentheorie. *Archiv für mathematische Logik und Grundlagenforschung 16*, 97–118.

Gentzen, G. (1936). Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*.

Gödel, K. (1932). Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien 69*, 65–66.

Gödel, K. (1933a). The present situation in the foundations of mathematics. Printed in Gödel 1995, 45–53.

Gödel, K. (1933b). Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums 4*, 34–38. Page numbering refers to the English translation in Gödel 1986, 286–295.

Gödel, K. (1938). Vortrag bei Zilsel. Printed in Gödel 1995, 86–113.

Gödel, K. (1941). In what sense is intuitionistic logic constructive? Printed in Gödel 1995, 189–200.

Gödel, K. (1958). Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica 12*, 280–287. Page numbering refers to the English translation in Gödel 1990, 240–252.

Gödel, K. (1986). *Collected Works. Vol. I. Publications 1929–1936*. Oxford: Oxford University Press.

Gödel, K. (1990). *Collected Works. Vol. II. Publications 1938–1974*. Oxford: Oxford University Press.

Gödel, K. (1995). *Collected Works. Vol. III. Unpublished Essays and Lectures*. Oxford: Oxford University Press.

Gödel, K. (2003a). *Collected Works. Vol. IV. Correspondence A–G*. Oxford: Oxford University Press.

Gödel, K. (2003b). *Collected Works. Vol. V. Correspondence H–Z*. Oxford: Oxford University Press.

Hämeen-Anttila, M. (2020). *Gödel on intuitionism and constructive foundations of mathematics*. Ph. D. thesis, University of Helsinki.

van Heijenoort, J. (1967). *From Frege to Gödel*. Cambridge, MA: Harvard University Press.

Heyting, A. (1930a). Die formalen Regeln der intuitionistischen Logik I/II. *Sitzungsberichte der Preussischen Akademie von Wissenschaften 16*, 42–56, 57–71.

Heyting, A. (1930b). Die formalen Regeln der intuitionistischen Mathematik. *Sitzungsberichte der Preussischen Akademie von Wissenschaften 16*, 158–169.

Heyting, A. (1931). Die intuitionistische Grundlegung der Mathematik. *Erkenntnis 2*, 106–115.

Heyting, A. (1934). *Mathematische Grundlagenforschung. Intuitionismus, Beweistheorie*. Berlin: Springer.

Heyting, A. (1978). History of the foundations of mathematics. *Nieuw Archief voor Wiskunde 26*, 1–21.

Hilbert, D. (1926). Über das Unendliche. *Mathematische Annalen 95*, 161–190.

Hilbert, D. and P. Bernays (1934). *Grundlagen der Mathematik I*. Die Grundlehren der mathematischen Wissenschaften. Berlin: Springer.

Howard, W. (1970). Assignment of ordinals to terms for primitive recursive functionals of finite type. In J. M. A. Kino and R. Vesley (Eds.), *Intuitionism and Proof Theory: Proceedings of the Summer Conference at Buffalo N.Y. 1968*, Volume 60 of *Studies in Logic and the Foundations of Mathematics*, pp. 443–458. Amsterdam: North-Holland.

Howard, W. A. (1968). Functional interpretation of bar induction by bar recursion. *Compositio Mathematica 20*, 107–124.

Kleene, S. C. (1945). On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic 10*, 109–124.

Kohlenbach, U. (2008). *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Berlin: Springer.

Kreisel, G. (1959). Interpretation of analysis by means of constructive functionals of finite types. In A. Heyting (Ed.), *Constructivity in Mathematics*, pp. 101–128. Amsterdam: North-Holland.

Kreisel, G. (1987). Gödel's excursions into intuitionistic logic. In P. Weingarter and L. Schmetterer (Eds.), *Gödel Remembered: Salzburg 10–12 July 1983*, pp. 65–179. Napoli: Bibliopolis.

Krivine, J.-L. (1990). Opérateurs de mise en mémoire et traduction de Gödel. *Archive for Mathematical Logic 30*, 241–267.

Kuroda, S. (1951). Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Mathematical Journal 2*, 35–47.

Luckhardt, H. (1973). *Extensional Gödel Functional Interpretation: A Consistency Proof of Classical Analysis*. Lecture Notes in Mathematics. Berlin: Springer.

Parsons, C. (1970). On a number theoretic choice schema and its relation to induction. In A. Kino, J. Myhill, and R. Vesley (Eds.), *Intuitionism and Proof Theory: Proceedings of the Summer Conference at Buffalo N.Y. 1968*, Volume 60 of *Studies in Logic and the Foundations of Mathematics*, pp. 459–473. Amsterdam: North-Holland.

Petersen, U. (2003). $L^i D^Z_\lambda$ as a basis for PRA. *Archive for Mathematical Logic 42*(7), 665–694.

von Plato, J. (2021). *Chapters from Gödel's Unfinished Book on Foundational Research in Mathematics*, Volume 8 of *Vienna Circle Institute Library*. Springer, forthcoming.

Shoenfield, J. R. (1967). *Mathematical Logic*. Reading, MA: Addison-Wesley.

Sieg, W. and C. Parsons (1995). Introductory note to *1938. In Gödel 1995, 62–85.

Spector, C. (1962). Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics. In J. Dekker (Ed.), *Recursive Function Theory, Symposia in Pure Mathematics*, Volume 5 of *Proceedings of Symposia in Pure Mathematics*, pp. 1–27. Providence: American Mathematical Society.

Tait, W. W. (1967). Intensional interpretations of functionals of finite type I. *Journal of Symbolic Logic 32*, 198–212.

Troelstra, A. S. (1973). *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Berlin: Springer.

Troelstra, A. S. (1990). Introductory note to 1958 and 1972. In Gödel 1999, 217–239.

Troelstra, A. S. (1995). Introductory note to *1941. In Gödel 1995, 186–189.

Urquhart, A. (2016). Russell and Gödel. *Bulletin of Symbolic Logic 22*, 504–520.

Wang, H. (1989). *Reflections on Kurt Gödel*. Cambridge, MA: MIT Press.

Weyl, H. (1921). Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift 10*, 39–79.

Weyl, H. (1925). Die heutige Erkenntnislage in der Mathematik. *Symposion 1*, 1–32.

Weyl, H. (1926). *Philosophie der Mathematik und Naturwissenschaft*. München: Oldenbourg.

Index of names in the Princeton Lectures