Sources and Studies in the History of Mathematics
and Physical Sciences

Jan von Plato

Can Mathematics Be Proved Consistent?

Gödel's Shorthand Notes & Lectures on Incompleteness

Springer

Sources and Studies in the History of Mathematics and Physical Sciences

Sources and Studies in the History of Mathematics and Physical Sciences was inaugurated as two series in 1975 with the publication in Studies of Otto Neugebauer's seminal three-volume History of Ancient Mathematical Astronomy, which remains the central history of the subject. This publication was followed the next year in Sources by Gerald Toomer's transcription, translation (from the Arabic), and commentary of Diocles on Burning Mirrors. The two series were eventually amalgamated under a single editorial board led originally by Martin Klein (d. 2009) and Gerald Toomer, respectively two of the foremost historians of modern and ancient physical science. The goal of the joint series, as of its two predecessors, is to publish probing histories and thorough editions of technical developments in mathematics and physics, broadly construed. Its scope covers all relevant work from pre-classical antiquity through the last century, ranging from Babylonian mathematics to the scientific correspondence of H. A. Lorentz. Books in this series will interest scholars in the history of mathematics and physics, mathematicians, physicists, engineers, and anyone who seeks to understand the historical underpinnings of the modern physical sciences.

More information about this series at http://www.springer.com/series/4142

Jan von Plato

Can Mathematics Be Proved Consistent?

Gödel's Shorthand Notes & Lectures on Incompleteness

Springer

Jan von Plato
Department of Philosophy
University of Helsinki
Helsinki, Finland

ISSN 2196-8810 ISSN 2196-8829 (electronic)
Sources and Studies in the History of Mathematics and Physical Sciences
ISBN 978-3-030-50875-3 ISBN 978-3-030-50876-0 (eBook)
https://doi.org/10.1007/978-3-030-50876-0

Mathematics Subject Classification (2010): 03F40, 01Axx

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book contains all that is found in Gödel's preserved shorthand notebooks on his research that led to the famous incompleteness theorems of formal systems. The notes are followed by the original version of his article, before a dramatic change just a few days after it was handed in for publication, and six lectures and seminars in consequence of his celebrated result published in 1931. The notebooks and one of the lectures were written in German Gabelsberger shorthand that I have translated into English, usually from an intermediate transcription into German, but at places directly. I thank Tim Lethen for his help in the reading of many difficult shorthand passages, and Maria Hämeen-Anttila for her support, especially at the troublesome moment when I discovered Gödel's tricky change of his manuscript after it had been submitted for publication. Marcia Tucker of the Institute for Advanced Study was very helpful during my visit to the Firestone Library of Princeton University where the originals of Gödel's manuscripts are kept. Finally, I recollect with affection my mother's decision to challenge her little boy by enrolling him in the German elementary school of Helsinki, a choice without which I would not have learned to read Gödel's manuscripts.

Jan von Plato

Acknowledgment

The Kurt Gödel Papers on incompleteness that this book explores are kept at the Firestone Library of Princeton University. A finding aid with details about their contents is found at the end of the fifth volume of Gödel's *Collected Works.* The papers were divided by their cataloguer John Dawson into archival boxes and within boxes into folders. Folders can have a third division into documents, with a running document numbering system. The papers have been mainly accessed through a microfilm that is publicly available, but also directly in Princeton. References to specific pages of notebooks usually require the use of the reel and frame numbers of the microfilm and that is how the sources are mostly identified in this book. Passages from Gödel in the introductory Part I are identified *in loco*. The shorthand manuscript sources on which Gödel's 1931 article is based are described in detail in Part II, Section 2 of this book. The typewritten sources for his 1930 summary and the 1931 article are described in the following Section 3. The sources of the six lectures and seminars on incompleteness are described in the last Section 4 of Part II. These descriptions together with the frame and page numberings in Parts III–V allow the interested reader to identify the source texts with the precision of a notebook page.

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

Contents

Part I

Gödel's steps toward incompleteness

J. von Plato, *Can Mathematics Be Proved Consistent?*, Sources and Studies in the History of Mathematics and Physical Sciences, https://doi.org/10.1007/978-3-030-50876-0_1

1. The completeness problem

David Hilbert's list of 23 mathematical problems from the Paris international congress of mathematicians of 1900 had as the second problem the question of the consistency of analysis: to show that no contradiction follows from the axioms for real numbers. A slip of paper with an additional problem, to be placed last as a 24th one, was found some hundred years later, one that asked for criteria for the simplicity of proofs and in general, "to develop a theory of proof methods in mathematics." The development of a theory to this effect, what Hilbert called *proof theory*, started in Göttingen in 1917–18, when the First World War was coming to its end. Its main aim was to provide answers to such questions as consistency.

After the war, German mathematicians were excluded from the international congress of mathematicians, held every four years. The reappearance of Germans on the international scene took place in the international congress of Bologna in 1928, with Hilbert lecturing on "Problems in the foundations of mathematics." In his lecture, Hilbert surveyed the development of mathematics in the past few decades, then listed four main problems in its foundation. There was behind the list the most remarkable period of research into logic and foundations of mathematics seen so far. Hilbert had realised that Bertrand Russell's *Principia Mathematica* of 1910–13 offered the means for formalizing, not just mathematical axioms as in geometry, but even the logical steps in mathematical proofs: "One could see in the completion of this grandiose Russellian enterprise of *axiomatization of logic* the crowning of the task of axiomatization as a whole." (Hilbert 1918, p. 153). Ten years later, the logic of the connectives and quantifiers had been brought to perfection, presented in the book *Grundzüge der theoretischen Logik* (Hilbert and Ackermann 1928). The formalization of arithmetic had also been accomplished, with recursive definitions of the basic arithmetic operations and an axiom system for proofs in arithmetic. Hilbert believed at this time Wilhelm Ackermann and Johann von Neumann to have solved the problem of consistency for a strong system of arithmetic, but there remained some doubts about it.

The first and second problems in Hilbert's Bologna list of 1928 are about the extension of Ackermann's proof to higher areas of mathematics. The list has as the third problem, from Gödel's reading notes on Hilbert's article:[1]

[1] Part of document 050135, reel 36, frames 377 to 385 in the Gödel microfilms.

III Completeness of the axiom system of number theory

i.e., to be proved:

1. $\mathfrak{S}$ and $\overline{\mathfrak{S}}$ to be shown as not both free from contradiction.
2. When $\mathfrak{S}$ is free from contradiction, it is also provable.

After a digression on the role of the principle of induction, there follows:

IV Completeness of logic (would follow from the completeness of number theory) "Are all generally valid formulas provable?"

So far proved only for propositional logic and the logic of classes.

Gödel had studied the Hilbert-Ackermann book in late 1928, then began to formalize proofs in higher-order arithmetic, for which purpose he invented an impeccable system of linear natural deduction. A whole long notebook, the "Übungsheft Logik" (exercise notebook logic) is devoted to this purpose, with formal derivations of unprecedented complexity, more than eighty steps and up to four nested temporary hypotheses made (cf. von Plato 2018b).

Gödel's interest shifted soon from the actual construction of formal derivations to the completeness of a system of proof. The completeness of quantificational logic is clearly formulated, independently of arithmetic, in Hilbert-Ackermann. There is a shorthand notebook with the title "Diss. unrein" (Dissertation draft), fifty pages long, with an outline of the first ten chapter headings on page 16, slightly abbreviated here (document 040001):

1. Introduction
2. Notation and terminology
3. Basic theorems about the axiom system
4. Reduction to denumerable domains [Denkbereiche]

   ...

7. Independence of propositions and rules
8. Extension for the case in which = is incorporated

9. Extension for axiom systems with finitely and infinitely many propositions

10. Systems with a finite basis and monomorphic systems

These items from 1 to 10 are detailed in the rest of the notebook.

Gödel's proof of completeness for the "narrower functional calculus," i.e., first-order classical predicate logic, has disjunction, negation, and universal quantification as the basic notions. The simplest case of quantification is the formula $\forall x F(x)$ with $F(x)$ a propositional formula. Gödel states in a shorthand passage that if such a formula is "correct," i.e., becomes true under any choice of domain of individuals and relations for the relation symbols of the formula, then the instance *with a free variable x* must be a "tautology" of propositional logic. In the usual "Tarski semantics" that is—unfortunately—included in almost every first course in logic, the truth of universals is explained instead by the condition that *every instance* be true, an explanation that with an infinite domain of objects becomes infinitely long.

In Gödel, by contrast, with the free-variable formula $F(x)$ a tautology, it must be provable in propositional logic by the completeness of the latter, a result from Paul Bernays' *Habilitationsschrift* of 1918 and known to Gödel from Hilbert-Ackermann. That book is also the place in which the rules of inference for the quantifiers appear for the first time in an impeccable form (p. 54, with the acknowledgment that the axiom system for the quantifiers "was given by P. Bernays"). With the free-variable formula $F(x)$ provable in propositional logic, the rule of universal generalization gives at once that even $\forall x F(x)$ is derivable. The step is rather well hidden in Gödel's completeness proof in the thesis that proceeds in terms of satisfiability. At one point, he moves to provability of a free-variable formula, then universally quantified "by 3," the number given for the rule of generalization.

Gödel's profound understanding of predicate logic, especially the need for rules of inference for the quantifiers without which no proof of completeness is possible, is evident through a comparison: Rudolf Carnap, whose course he had followed in Vienna in 1928, published in 1929 a short presentation of Russell's *Principia*, the *Abriss der Logistik*, but one searches in vain for the quantifier rules in this booklet. Other contemporaries who failed in this respect include Ludwig Wittgenstein and Alfred Tarski. The former was a dilettante in logic who thought that truth-tables would do even for predicate logic. With the latter, no trace of the idea of the provability of universals through an arbitrary instance is found in his famous tract on the concept of truth of 1935.

Gödel's actual aim in his doctoral thesis was a proof of completeness of arithmetic, as is witnessed by the last third of the planned contents of his dissertation. It should be noted that Hilbert's Bologna address that listed the problem got published after Gödel had finished the thesis:

> 11. Application to geometry and arithmetic – connection between the two – inclusion of functions over objects
>
> a.) for the case of completeness b.) for the case that no finite basis is at hand
>
> 12. General construction of resultants and solution of the problem whether real roots are at hand
>
> holds for which number systems?, decision procedure
>
> 13. Resolution of the Archimedean axiom, proof of the completeness of the arithmetic axiom system
>
> 14. There is no finite basis for arithmetic propositions
>
> 15. Independence of the concepts ?

There seem to be no traces of how Gödel in 1929 thought he would prove the completeness of arithmetic, though I have not studied the long notebook in every detail yet—perhaps the above already indicates some doubt? There is instead his announcement of the failure of any such proof the next year, found at the end of the lecture he gave at the famous Königsberg conference on the foundations of mathematics on 5–7 September 1930. The conference is remembered for its presentation of the main approaches to the foundations of mathematics, logicism, formalism, intuitionism, in three widely read lectures by Rudolf Carnap, Johann von Neumann, and Arend Heyting, respectively.

Gödel's short and readable lecture about the completeness of predicate logic—just twenty minutes were allotted for it— is preserved in shorthand and very slightly changed in a typewritten form that got first published in the third volume of Gödel's *Collected Works*. Close to the end of that lecture, we find the following passage (p. 28):

> If one could prove the completeness theorem even for the higher parts of logic (the extended functional calculus), it could be shown quite generally that from categoricity, definiteness with respect to decision follows.[2] One knows for example that Peano's axiom system is categorical, so that the solvability of each problem in arithmetic and analysis expressible in the *Principia Mathematica* would follow. Such an extension of the completeness theorem as I have recently proved is, instead, impossible, i.e., there are mathematical problems that can be expressed in the *Principia Mathematica* but which cannot be solved by the logical means of the *Principia Mathematica*.

It is clear from these remarks that Gödel's first thought was to extend the completeness result to higher-order logic, a point emphasised in Goldfarb (2005). The above is an indication of his way to the first incompleteness theorem from the time when the actual work was done, namely through a failed attempt that led to the insight about undecidability.

The shorthand version of the Königsberg talk ends with (reel 24, frame 311):

> I have succeeded, instead [of extending the completeness theorem to higher-order logic], in showing that such a proof of completeness for the extended functional calculus is impossible or in other words, that there are arithmetic problems that cannot be solved by the logical means of the PM even if they can be expressed in this system. These things are, though, still too little worked through to go into more closely here.[3]

In the typewritten version, we read somewhat differently about his proof of the failure of completeness (document 040009, page 10):

> In this [proof], the reducibility axiom, infinity axiom (in the formulation: there are exactly denumerable individuals), and even the axiom of choice are allowed as axioms. One can express the matter also as: The axiom system of Peano with the logic of the

[2] Literal translation of the German "Entscheidungsdefinitheit."

[3] The last sentence reads in German: "Doch sind diese Dinge noch zu wenig durchgearbeitet, um hier näher darauf einzugehen."

> PM as a superstructure is not definite with respect to decision. I cannot, though, go into these things here more closely.[4]

Then this last sentence is cancelled and the following written: "It would, though, take us too far to go more closely into these things."[5] It would seem that matters concerning the incompleteness proof had cleared in Gödel's mind between the writing of the shorthand text for the lecture and the typewritten version.

The shorthand text for the Königsberg lecture is found fairly early in Gödel's two notebooks about incompleteness. There is, about sixty pages later, a shorthand draft for his two-page note on the two incompleteness results that he had prepared just before departing for Königsberg, with publication in October 1930. Whatever he had done about incompleteness by that point must have been before early September 1930, and some of it clearly earlier: Just a few pages before the Königsberg lecture text, Gödel writes that the formally undecidable sentences have "the character of Goldbach or Fermat," i.e., of universal propositions such that each of their instances is decidable. These examples suggest that a formally undecidable proposition $\forall x F(x)$ can have each of its numerical instances $F(n)$ provable, but still, addition of the negation $\neg\forall x F(x)$ does not lead to an inconsistency. Were the free-variable instance $F(x)$ provable, universal generalization would at once give a contradiction.

2. From Skolem's paradox to the Königsberg conference

Later in his life, Gödel gave various explanations of how he found the incompleteness results. He often repeated that he was thinking of self-referential statements, as in the liar paradox: *This sentence is false.* Replacing unprovable for false, one gets a statement that expresses its own unprovability. The explanation is good as far as it goes, and indeed given as a heuristic argument in Gödel's 1931 paper, but it gives little clue as to how one would start thinking along such lines in the first place. Gödel's meticulously kept notes and other material point at interesting circumstances that concern his discovery of the undecidable sentences.

As a first source from the time Gödel had begun work on incompleteness in the early summer of 1930 (by Wang 1996, p. 82; I would say perhaps

[4] The last sentence is: "Auf diese Dinge kann ich aber hier nicht näher eingehen."

[5] "Doch würde es zu weit führen, auf diese Dinge näher einzugehen."

May) there is Fraenkel's *Einleitung in die Mengenlehre* that, as is seen from Gödel's preserved library request cards, he had taken out in early April. Fraenkel discusses the question of decidability in principle of any mathematical problem, remarking that not a long time ago, every mathematician would believe in such solvability (p. 234):

> It is a fact that until today, no mathematical problem has been proved to be "unsolvable." The discovery of such a problem would without doubt present an enormous *novum* for mathematics, and not only for it.

Fraenkel is very clear about *Skolem's paradox:* The propositions of a truly formal system form a denumerably infinite class, and therefore in particular the provable propositions, i.e., the theorems. The seemingly paradoxical consequence is that formal (first-order) theories of real numbers and of set theory admit of interpretations in which the domain is only denumerably infinite. In particular, it can be taken to be the domain of natural numbers.

Further down, Fraenkel notes that "there should be nothing absurd in imagining that the unsolvability of a problem could even be *proved*" (p. 235).

A second early source bears the date 13 May when Gödel borrowed Skolem's "Über einige Grundlagenfragen der Mathematik." This 49-page article was published as a separate issue of an obscure Norwegian journal. There Skolem gives a striking version of his paradox: The denumerable infinity of propositions of a truly formal system can be *ordered lexicographically*. "Propositions about natural numbers," in particular, can be likewise thus ordered, but by contrast the properties of natural numbers cannot be so ordered, by which (p. 269):

> It would be an interesting task to show that every collection of propositions about the natural numbers, formulated in first-order logic, continues to hold when one makes certain changes in the meaning of "numbers."

Among the wealth of ideas in Skolem's paper, there is an outline for a proof that the consistency of classical arithmetic reduces to that of intuitionistic arithmetic (p. 260), a result Gödel proved in 1932 through his well-known double negation translation.

Next to the library slips, two early notebooks give indications of Gödel's reading through his summaries of papers by others. It has turned out recently that these were begun around August 1931, when Gödel accepted the task of writing together with Heyting a short book on "Mathematical Foundational Research." The first notebook has two articles listed on each page, on top and half-way down, at times with notes, at times not, altogether over a hundred articles that relate to the topics Gödel was supposed to write as his part of the book project. Then there is the earliest preserved and clearly written-out notebook with the text *Altes Excerptenheft I* (1931—...) on the cover and a continuous page numbering (document 030079). This *Heft* gathers together some of his most important sources at a time when there were no photocopiers.

In his three-page summary of Skolem's long article (*Excerptenheft*, pp. 25–27), Gödel begins with Skolem's § 2, "proof of set-theoretical relativism" in Gödel's words, and then comes § 1, "enumeration of possible properties (therefore also sets) in Fraenkel's as well as Skolem's separation axiom." The last item in Gödel is for Skolem's §7, with the condition $ah - bk = 1$ pointing at the unique decomposition into prime elements in principal ideal domains:

> §7 Example of a domain that is not isomorphic with the number sequence even if it is an integral domain and even if for every two relatively prime $h, k,\ ah - bk = 1$.
>
> Conjecture that the number sequence is not at all characterisable by propositions of first-order logic.

At the end of this section, Gödel paraphrases Skolem's conclusion: "There is no possibility to introduce things nondenumerable as anything else but a pure dogma."

Gödel's summary was written down after his work on incompleteness had been finished and published. Still, Skolem's paper contains important ideas he had seen before that work. The way from these ideas to a first intimation of incompleteness is not long. One would likely think along the following lines:

Properties of natural numbers can be given as arithmetic propositions $F(x)$ with one free variable, and they can be listed in a lexicographical order, $F_1(x), F_2(x), \dots F_n(x) \dots$. Each of these properties $F_i(x)$ corresponds to a set of natural numbers, those for which the property holds and usually written as $M_i = \{x \, \varepsilon \, N | F_i(x)\}$. These sets form a denumerable sequence, but the sets of natural numbers as a whole form a continuum; each of them corresponds to a real number. The M_i give just a denumerable sequence of real numbers that one can diagonalise by the familiar argument of Cantor. Then we have a set D of natural numbers that is different from all of the M_i. Could we describe the diagonalization procedure within arithmetic itself, to form an expression in the language of arithmetic that corresponds to the diagonal set D, i.e., some free-variable formula $G(x)$ such that $D = \{x \, \varepsilon \, N | G(x)\}$?

To realise a possibility is one thing. To express provability in a formal system inside the system itself and to construct a proposition that expresses its own unprovability is, then, the real discovery. The Gödel notes show stages of the development of his ideas. The clearest turning point is one connected to the Königsberg conference. Before that, Gödel's argument was to give a truth definition for propositions of *Principia Mathematica*, then to prove that all theorems are true. If the proposition that expresses its own unprovability were provable, it would be true, hence unprovable, so it cannot be a theorem.

Gödel saw very clearly that the truth definition is the element in his proof that cannot be expressed within the formal system. He asked what it was that made his proof of undecidability possible. It was the said metatheorem about the truth of all the theorems, by which it could be decided that the constructed self-referential proposition is not simply false. If that decision could be made within the system, the unprovable proposition would follow. Therefore the truth of theorems is unprovable in the system.

The above argument is, in brief, a proof that the consistency of the system of *Principia Mathematica* cannot be proved within the system, or Gödel's original second incompleteness theorem. His later recollections dated its discovery to the times of the Königsberg conference. At that time, he prepared the mentioned short note of his results that appeared in October 1930, the

> *Some metamathematical results on definiteness with respect to decision and on freedom from contradiction*

This note was published in the *Anzeiger der Akademie der Wissenschaften zu Wien*, communicated by "corresponding member H. Hahn," Gödel's professor.

No trace of Gödel's original proof of the incompleteness theorems that uses a truth definition is left in his published article, but the idea surfaced from other quarters. Andrzej Mostowski knew Gödel from the late 1930s, from his stay in Vienna as recorded in Gödel's shorthand notes on the discussions they had. After the war, Mostowski became the author of the first book on Gödel's incompleteness theorems, the *Sentences Undecidable in Formalized Arithmetic: An Exposition of the Theory of Kurt Gödel* of 1952. There he describes two main ways of proving incompleteness, the first called *syntactic* and followed in Gödel's paper, the second *semantic*. The latter gives (p. 10) "an exact definition of what may be called the class of true sentences," with Gödel's theorem following from three conditions: "Every theorem of (S) is true," secondly the condition that no negation of a theorem be true, and as third the condition by which the truth predicate is equivalent to a condition of unprovability. A footnote on the next page states that "the idea of the semantical proof of the incompleteness theorem is due to A. Tarski," the long work on the concept of truth in formalized languages of 1935.

The second series of Gödel's notes contains, about six pages before the Königsberg break, the following (page 300R below):

> We go now into the exact definition of a concept "true proposition." The idea of such a definition has been expressed [cancelled: simultaneously] independently of me by Mr A. Tarski from Warsaw.

On the next page, we read:

> Now one arrives also quite exactly at proving (through complete induction) that
>
> *Each provable proposition is true.*

Tarski had visited Vienna in February 1930 and gave some lectures there that Gödel followed. A hint on their discussions is given by a letter Gödel wrote to Bernays on 2 April 1931. One finds there a "class sign" $W(x)$ read as "x is a *true proposition*," with truth of negation, disjunction, and universal quantification defined in the standard way (*Collected Works IV*, p. 96):

> The idea to define the concept of a "true proposition" along this way has been, incidentally, developed simultaneously and independently of me by Mr A. Tarski (as I gather from an oral communication).

The characteristic of Gödel's pre-Königsberg proof of the incompleteness theorems was that he used whatever means of classical mathematics, analysis and set theory included, in metamathematics. After the Königsberg meeting, the concept of truth and even the intuitive notion of "correctness" disappeared absolutely from his notes on incompleteness: one simply doesn't even find these words anymore, but instead an emphasis on the constructiveness of his proofs achieved through an "arithmetization of metamathematics" by elementary means. A hint of his original proof method is contained in the lectures Gödel gave on incompleteness in Princeton in the spring of 1934. There is a brief heuristic discussion of an arithmetic predicate $T(z_n)$ that expresses the "truth of the formula with number n," similar to the truth predicate W of his earlier writings.

There has been some lament about Gödel not acknowledging Tarski's approach to incompleteness. In the light of the above, the matter was *dejà vu* for Gödel, and not original to Tarski. From what has come out above, Gödel had begun work on incompleteness in May or June 1930. How does this fit together with Tarski's visit several months earlier, if the concept of "true proposition" was developed simultaneously? Gödel had arranged for an opportunity to discuss with Tarski and knew in that way about Tarski's ideas. In February 1930, he was in need of a truth definition for the aim that comes out so clearly from the Königsberg lecture, namely for the completeness of higher-order logic, the type theory of Russell's *Principia*, to be a well-posed problem. Such a concept would cover his system of proof in the 1928/29 *Übungsheft*, also to decide what axioms to accept in higher-order logic. The topic of a truth definition was of great systematic value for Gödel who mentions in his shorthand notes from the 1930s several times a folder named "The concept of truth" (Mappe "Wahrheitsbegriff").

3. From the Königsberg conference to von Neumann's letter

Among Gödel's audience in Königsberg sat Johann von Neumann who reacted at once and wanted more explanations. The two had discussions at the conference and in Berlin, where Gödel stayed for a few days immediately after the conference. The most detailed account of these events is Wang (1996), section "Some facts about Gödel in his own words," that describes the first approach to incompleteness as follows (pp. 82–84):

> I represented real numbers by predicates in number theory and

> found that I had to use the concept of truth to verify the axioms of analysis. By an enumeration of symbols, sentences, and proofs of the given system, I quickly discovered that the concept of arithmetic truth cannot be defined in arithmetic.
>
> . . .
>
> Note that this argument can be formalized to show the existence of undecidable propositions without giving any individual instances.

Gödel's words are different from those of his notebooks of 1930; The "verification of the axioms of analysis" means that a concept of truth is established by which the axioms turn out true and the rules of inference maintain truth. The formulation of 1930 was that each provable proposition of Russell's type theory is true.

Von Neumann suggested in the discussion to transform undecidability "into a proposition about integers." Gödel then found "the surprising result giving undecidable propositions about polynomials."

An edited account of the Königsberg discussion was published in the journal *Erkenntnis* (vol. 2, 1931, pp. 135–151). It contained also a brief summary of the incompleteness result by Gödel with the title "Nachtrag" (addendum, pp. 149–151), written some time in 1930/31. A typewritten version, not essentially different from the published one, is found in reel 24, frames 240–242.

Gödel's library loan cards show that he stayed in Berlin right after the Königsberg meeting and that he requested again Skolem's long paper of 1929, on 12 September from a library in Berlin. We are at the most crucial turning point in Gödel's work on incompleteness, the abandonment of the proof idea by which all theorems of the *Principia* are true, proved by methods of set theory. The first sign of this change is a set of 13 shorthand pages, 360L to 366L, in particular page 364L in which it is stated that the concept of "contentful correctness" can be restricted to instances of recursive predicates. These pages begin in exactly the same way as the final shorthand version and come close to the formulations in the introductory parts of the published article: "The development of mathematics in the direction of greater exactness has, as is well known, led to wide areas of it being formalized." Another sign of change from a set-theoretic approach that uses the concept of truth to one that uses primitive recursive arithmetic is that

Gödel writes ω-consistency instead of $\aleph_0$-consistency, the latter still found in the *Anzeiger* note handed by Gödel's account in on 17 September.

Von Neumann lectured from late October 1930 on in Berlin on "Hilbert's proof theory." Carl Hempel, later a very famous philosopher, recollected the excitement created, even evidenced by contemporary letters for which see Mancosu (1999). The account is (Hempel 2000, pp. 13–14):

> I took a course there with von Neumann which dealt with Hilbert's attempt to prove the consistency of classical mathematics by finitary means. I recall that in the middle of the course von Neumann came in one day and announced that he had just received a paper from... Kurt Gödel who showed that the objectives which Hilbert had in mind and on which I had heard Hilbert's course in Göttingen could not be achieved at all. Von Neumann, therefore, dropped the pursuit of this subject and devoted the rest of the course to the presentation of Gödel's results. The finding evoked an enormous excitement.

These are later recollections; It is known that von Neumann got the proofs of Gödel's paper around the tenth of January 1931. As we shall soon see, what von Neumann received during his lecture course are the October 1930 note with the first and second theorem stated, and the manuscript of section 4 of Gödel's paper.

One of the few known participants in von Neumann's lecture course was Jacques Herbrand. He was born in 1908 and received his education at the prestigious *Ecole normale superieure* of Paris. He finished his thesis *Recherches sur la théorie de la démonstration* at the precocious age of 21 in the spring of 1929. He went to stay for the academic year 1930–31 in Germany, first Berlin from October 1930 on, then Hamburg and Göttingen from late spring 1931 to July. These stays were in part prompted by his work on algebra, where Emil Artin in Hamburg and Emmy Noether in Göttingen were the leading figures. Herbrand's life ended in a mountaineering accident in July 1931.[6]

There is a letter of Herbrand's of 28 November 1930 to the director of the *Ecole normale* Ernest Vessiot in which he mentions von Neumann's "absolutely unexpected results," then writes that for now he will write about

[6] My *Formal Machinery Works*, section 8.3 on "two Berliners" contains a detailed account of Herbrand's stay in Germany and his relation to von Neumann.

the

> extremely curious results of a young Austrian mathematician who succeeded in constructing arithmetic functions Pn with the following properties: one calculates Pa for each number a and finds $Pa = 0$, but it is impossible to prove that Pn is always zero.

As noted above, the pre-Königsberg part of Gödel's second notebook mentions that undecidable problems can have "the character of Goldbach or Fermat." There is a difference, though, for Goldbach's conjecture, if false, can be refuted by a counterexample. With Gödel's undecidable propositions, it happens that each instance $F(n)$ of a property of natural numbers is provable, by which there is no counterexample. Still, $\forall x F(x)$, classically equivalent to $\neg\exists x \neg F(x)$, or the impossibility of a counterexample, need not be provable within the system. Gödel hardly thought that Goldbach's conjecture would be a "Gödel sentence."

Gödel states that he found the arithmetical form of incompleteness right after the Königsberg meeting. Here are his own words about the change (from Wang 1996, pp. 83–84):

> To von Neumann's question whether the proposition could be expressed in number theory I replied: of course they can be mapped into the integers but there would be new relations. He believed that it could be transformed into a proposition about integers. This suggested a simplification, but he contributed nothing to the proof, because the idea that it can be transformed into integers is trivial. I should, however, have mentioned the suggestion; otherwise too much credit would have gone to it.[7] If today, I would have mentioned it. The result that the proposition can be transformed into one about polynomials was very unexpected and done entirely by myself.

Herbrand's letter shows that von Neumann knew about the polynomial formulation–the "arithmetic functions Pn" for which $Pa = 0$ is provable for each number a–therefore the matter must have surfaced during their discussions in Berlin.

[7] The wording of Wang's notes seems somewhat awkward here, as if Gödel needed to protect himself against a priority claim by von Neumann, deceased two decades earlier.

Looking at the notebooks, one realizes that Gödel's "arithmetization of metamathematics" was initially that natural numbers can be used as the basic symbols of a formal system and that formulas then correspond to series of numbers. This representation appears first on page 294R:

> We replace the basic signs of the *Principia* (variables of different types and logical constants) in a one-to-one way by natural numbers, and the formulas through finite sequences of natural numbers (functions over segments of the number sequence of natural numbers).[8]

The famous Gödel numbering through the uniqueness of prime decomposition is seen first on page 293R, but just in the margin. There is no explanation of these expressions, $2^x3^y5^z7^u11^v$, 2^u3^v, and p_n, the last the n-th prime, by which they must be later additions.

Incidentally, Gödel's page 299L gives a clue to the origin of the idea of coding formulas and proofs through the uniqueness of prime decomposition: Gödel had used the numbers 0–7 as arithmetic representations of his basic signs, then needed an unlimited supply of numbers to represent variables of all finite types. He took numbers greater than 7 divisible by exactly one prime as propositional variables, and those divisible by exactly $k+2$ primes as variables of type k.

The cancelled page 329L, written well before the Königsberg meeting, develops the idea of Gödel numbering, with the comment that by the mapping of series of numbers to numbers through a product of powers of primes, "metamathematical concepts earlier defined that concern the system S, go over into properties and relations between natural numbers." This mapping is put aside, however, and series of numbers continue to represent formulas and proofs until the final shorthand version that was written after the Königsberg meeting. There, on pages 254L-R, Gödel writes that by taking products of powers of primes, "a natural number is associated in a one-to-one way, not just to each basic sign but also to each finite series of basic signs" – an idea described as "trivial" in Gödel's recollections about his meeting with von Neumann.

---

[8] The German is: Belegungen von Abschnitten der natürlichen Zahlenreihe mit natürlichen Zahlen. The English wording is from the printed article in Van Heijenoort (1967), as approved by Gödel.

Eight days before Herbrand's letter, von Neumann had written to Gödel about his proof:

> It can be expressed in a formal system that contains arithmetic, on the basis of your considerations, that the formula $1 = 2$ cannot be the endformula in a proof that starts from the axioms of this system—and in this formulation in fact a formula of the formal system mentioned. Let it be called $\mathfrak{W}$.
>
> ...
>
> I show now: $\mathfrak{W}$ is always unprovable in systems free from contradiction, i.e., a possible effective proof of $\mathfrak{W}$ could certainly be transformed into a contradiction.

Gödel must have explained how undecidable propositions are constructed to von Neumann in Berlin, not just a blunt statement of incompleteness, namely the way in which the provability of a formula in a system can be expressed as a formula of that system, and likewise with unprovability. In particular, the unprovability of a contradiction, say $1 = 2$, becomes expressed through an arithmetic formula.

Von Neumann writes next that if Gödel is interested, he would send the details once they are ready for print. He asks further when Gödel's treatise will appear and when he can have the proofs, with the wish to relate his work "in content and notation to yours, and even the wish for my part to publish sooner rather than later."

Herbrand had explained the post-Königsberg statement of incompleteness in terms of polynomials to Vessiot, and five days later he writes another letter, to his friend Claude Chevalley, in the worst handwriting imaginable, but full of sparkling ideas that seem to spring from nothing. In the letter, Herbrand explains von Neumann's presentation of the incompleteness theorem as follows:

> Let T be a theory that contains arithmetic. Let us enumerate all the demonstrations in T; let us enumerate all the propositions $Q\,x$; and let us construct a function $P\,x\,y\,z$ that is zero if and only if demonstration number x demonstrates $Q\,y$, Q being proposition number z.
>
> We find that $P\,x\,y\,z$ is an effective function that one can construct with arithmetic functions that are easily definable.

Let β be the number of the proposition $(x) \sim P\, x\, y\, y$ ($\sim$ means: not); let $A\, x$ be the proposition $\sim P\, x\, \beta\, \beta$

A the proposition $(x).A\, x$ ($A\, x$ is always true)

$A\, x$, equivalent to: demonstration x does not demonstrate the proposition β; so

$A\, x. \equiv .$ demonstration x does not demonstrate A

Let us enunciate:

$A\, x. \equiv . \sim D(x, A)$

1) $A\, x$ is true (for each cipher x); without it $D(x, A)$ would be true; therefore A; therefore $A\, x$; therefore $\sim D(x, A)$.

2) A cannot be demonstrated

for if one demonstrates A, $A\, x$ would be false; contradiction.

Therefore: $A\, 0, A\, 1, A\, 2 \ldots$ are true

$(x) A\, x$ cannot be demonstrated in T

Next in Herbrand's letter comes the striking second incompleteness theorem. With $D(x, A)$ standing as above for "proof number x demonstrates proposition A," Herbrand writes in the letter the key formulas:

3) $\sim A \rightarrow D(x, A)$ et $D(z, \sim A)$

therefore: $\sim (D(x, A)$ et $D(z, \sim A)) \rightarrow A$

The conclusion, for the unprovable proposition A, is that "if one proves consistency, one proves A": Consistency requires that for any proposition A, there do not exist proofs of A and $\sim A$. This inexistence can be expressed as the formula $\sim \exists x \exists z (D(x, A)$ et $D(z, \sim A))$, or in a free-variable formulation, as $\sim (D(x, A)$ et $D(z, \sim A))$ for each x and z.

The contrapositive of Herbrand's formula 3) states that consistency implies A, a formulation taken over from Gödel as we shall see.

Let us now turn to Gödel's final shorthand version of the incompleteness paper. It occupies the first 39 pages of a notebook (document 040014), with a beginning that is very similar to the typewritten version. The impressive list of 45 recursive relations in the published paper matches a similar list of 43 items, some ten pages, followed by the upshot of the laborious work in the form of a theorem:

VI. Each recursive relation is arithmetic.

After the text proper of the manuscript for the article ends, there are two attempts at a formulation of a title, like this:

> On the existence of undecidable mathematical propositions in the system of *Principia Mathematica*
>
> On unsolvable mathematical problems in the system of *Principia Mathematica*

There follow five pages with formulas, recursive definitions of functions, elementary computations, and a stylish layout for a lecture on the completeness of predicate logic given in Vienna on 28 November. Next the title "Lieber Herr von Neumann" hits the eye, with an unfinished letter draft that contains:

> Dear Mr von Neumann
>
> Many thanks for your letter of [20 November]. Unfortunately I have to inform you that I have been in possession of the result you communicated since about three months. It is also found in the attached offprint of a communication to the Academy of Sciences. I had already finished the manuscript for this communication before my departure for Königsberg and had presented it to Carnap. I gave it over to Hahn for publication in the *Anzeiger* of the Academy on 17 September. [Cancelled: The reason why I didn't inform you in any way [written heavily over: didn't tell anything] of my second result in Königsberg is that the precise proof is not suited to oral communications and that an approximate indication could easily arouse doubts about [heavily cancelled: correctness] executability (as with the first) that would not appear convincing.] Concerning the publication of this matter, there will be given only a shorter sketch of the proof of impossibility of a proof of freedom from contradiction in the *Monatsheft* that will appear in January[9] (the main part of this treatise will be filled with the proof of existence of undecidable sentences). The detailed carrying through of the proof

[9] [Despite its name, the *Monatshefte* (monthly notices) appeared in four yearly issues. January has been changed into "early 1931."]