Proc. Int. Cong. of Math. – 2018
Rio de Janeiro, Vol. 4 (4075–4092)

IN SEARCH OF THE SOURCES OF INCOMPLETENESS

Jan von Plato

Abstract

Kurt Gödel said of the discovery of his famous incompleteness theorem that he substituted “unprovable” for “false” in the paradoxical statement *This sentence is false*. Thereby he obtained something that states its own unprovability, so that if the statement is true, it should indeed be unprovable. The big methodical obstacle that Gödel solved so brilliantly was to code such a self-referential statement in terms of arithmetic. The shorthand notes on incompleteness that Gödel had meticulously kept are examined for the first time, with a picture of the emergence of incompleteness different from the one the received story of its discovery suggests.

1. Prelude to incompleteness. Kurt Gödel’s paper of 1931 about the incompleteness of mathematics, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, belongs to the most iconic works of the first half of 20th century science, comparable to the ones of Einstein on relativity (1905), Heisenberg on quantum mechanics (1925), Kolmogorov on continuous-time random processes (1931), and Turing on computability (1936). Gödel showed that the quest of representing the whole of mathematics as a closed, complete, and formal system is unachievable.

It has been said repeatedly that Gödel’s discovery has had close to no practical effect on mathematics: the impossibility to decide some questions inside a formal system has not surfaced except in a few cases, such as the question of the convergence of what are known as Goodstein sequences. On the other hand, the effect of methods Gödel developed to prove his theorems has been immeasurable: He invented the idea of a *formal syntax* coded through primitive recursive functions, from which Turing’s idea of machine-executable code arose.

I thank Thierry Coquand, Martin Davis, John Dawson, Warren Goldfarb, and Bill Howard for their suggestions and generous encouragement, Maria Hämeen-Anttila for moral support in moments of doubt during the study of Gödel’s manuscripts, and the splendid class of participants in my “Gödel detective” lecture course at the University of Helsinki in 2017, during which the results reported here were achieved.

MSC2010: primary 03F40; secondary 01A60, 03-03, 03A05.

Gödel was enormously lucky, or sagacious, to find in 1928 a very precise problem in logic to work with. It occurs in the textbook *Grundzüge der theoretischen Logik* (Basic traits of theoretical logic) by David Hilbert and Wilhelm Ackermann, actually written by Hilbert's assistant Paul Bernays (for which see Hilbert [2013, p. 49]). It gave, for the first time, a complete system of axioms and rules for the logic of the connectives and quantifiers and was the first important step in Hilbert's program that had set as its aim the formalization of mathematical reasoning within a logical language, with proofs of the consistency and completeness of the formalization. Predicate logic is such a language in which proofs in elementary arithmetic can be expressed as *derivations* in a formal system. Derivations, alongside *expressions* of a formal language, form an inductively defined class of objects. Hilbert's inspiration for the formalization of mathematical proofs within a logical language came from the three-volume *Principia Mathematica* 1927 that became known and read in Hilbert's Göttingen from 1917 on.

In [Hilbert and Ackermann 1928], the completeness of predicate logic is given as an important open problem (p. 68):

> Whether the axiom system is complete in the sense that all logical formulas correct in each domain of individuals really are derivable in it, is a question still unresolved. One can say only purely empirically that the axiom system has been always sufficient in all applications. The independence of the single axioms has not been studied yet.

Gödel set out to solve the completeness problem, a work that led to his doctoral thesis of 1929, with even proofs of the independence of the axioms and rules included.

I recently found in Gödel's Nachlass an 84 page notebook entitled *Übungsheft Logik* (Logic exercise notebook) based on his reading of Hilbert-Ackermann. Besides predicate logic in which one quantifies only over individuals, the book also presents higher-order logic, with short explanations of how the formalism could be applied in arithmetic and set theory. By the comprehension principle, higher-order predicates correspond to sets, or functions as well over which one can apply the quantifiers, with a very powerful formalism in which to express the principles and postulates used in mathematical theories as a result. One thing that transpires from the *Übungsheft* is that Gödel's main initial objective was to use higher-order logic for a formalization of proofs in arithmetic and set theory. The formalizations are in the "logicist" tradition of Gottlob Frege and Russell (see von Plato [2018] for a detailed account).

Another, astonishing feature of the *Übungsheft* is the way Gödel overcomes the main difficulty of the axiomatic logic of Frege and Russell which is that formal derivations are, as a practical matter, impossible to construct in axiomatic logic. With no explanation or hesitation whatsoever, he starts to use a *system of natural deduction* in place of the hopelessly clumsy axiomatic calculus for his formal proofs, in a format in which the formulas follow in a linear vertically arranged succession. Gerhard Gentzen is generally considered to be the inventor of natural deduction, a logical calculus in which derivations are presented in a tree-form that allows for a deep analysis of their structure. His way to natural deduction is detailed out on the basis of his manuscripts in the book *Saved from the Cellar* [von Plato 2017a].

Gödel changed his objective from the actual formalization of mathematics as in the *Übungsheft* into a study of the properties of any such formalization: On 6 July 1929, Gödel handed in a brilliant doctoral thesis with the title *Über die Vollständigkeit des Logikkalküls* (On the completeness of the calculus of logic). A shorter version got published in 1930.

Gödel used the term *completeness* (Vollständigkeit) for the case of predicate logic. The contrary was not Unvollständigkeit, but *undecidability* (Unentscheidbarkeit). In the 1931 article , he uses "formal undecidability," even in the title and writes then specifically of a sentence A "axiomatically undecidable," i.e., undecidable or unsolvable in an axiomatic system in the sense of unprovability of both A and its negation $\neg A$ within the system. There is some danger of confusing this term with another notion of undecidability, the one of Hilbert's *Entscheidungsproblem*, as found in Turing's 1936 title. The two senses of undecidability have a somewhat tricky relation with all four combinations possible: Classical propositional logic is decidable and complete, but if you leave out one axiom you get an incomplete though still decidable axiom system. Similarly, classical predicate logic is complete but undecidable, and Peano arithmetic incomplete and undecidable.

2. The sources of incompleteness. Gödel was a maniacal keeper of notebooks in which he recorded his thoughts, from his earliest school years on, and there is an enormous amount of material left behind and kept in Princeton. These notebooks are written in the Gabelsberger shorthand that was regularly taught at schools at the time, but rendered obsolete since its substitution by the "unified shorthand" in 1925. Work on these archival sources was done in connection with the publication of the third of Gödel's five-volume *Collected Works*, which led to the incorporation of two very important manuscripts in the volume: a lecture on the consistency of the continuum hypothesis of December 1939, and another on Gödel's well-known *functional interpretation* of arithmetic in April 1941. Since those days, with publication in 1995, Gödel's notebooks have lain dormant except for his collection of fifteen philosophical notebooks, the *MaxPhil* series (Philosophical Maxims).

I started to work with the Gödel notes on 20 March, 2017, after a volume of shorthand notes by Gödel's younger contemporary Gentzen, the mentioned *Saved from the Cellar*. Here I report on my findings that concern the sources of Gödel's work on incompleteness, with a typewritten manuscript of *Über formal unentscheidbare Sätze der Principia*

Mathematica und verwandter Systeme submitted on 17 November and published in March 1931.

The Gödel archives contain three suites of notes in preparation of the incompleteness paper. The third version comes close to the published paper and bears the cover title *Unentsch unrein*. This text comes first in the microfilm and is in 39 pages. There follow about the same number of additional pages, some of which with remarkable connections to the main text. A second notebook follows that has the same title, this time written inside, and it is even a bit longer, with a break and what seems a new start in the middle. Finally, there is a third notebook with no title and with the first preserved notes on incompleteness, some 45 pages followed by a dozen pages that are very similar to the introduction of the printed paper.

3. The crucial point of the completeness proof. 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 propositional. 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.[1] In the usual "Tarski semantics" that is–unfortunately–included in almost every first course in logic, the truth of universals is explained by the condition that *every instance* be true, an explanation that with an infinite domain of objects leads to circles.

In Gödel, in 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 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 generalisation gives at once that even $\forall x F(x)$ is derivable. The step is rather well hidden in Gödel's 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 generalisation.

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 comparison: Rudolf Carnap, whose course he 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

[1] My notes are incomplete at this point and I have so far not found again this passage in Gödel's manuscripts.

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 1934. We shall see that Gödel was way ahead of him in understanding these matters by the summer of 1930.

Gödel's thesis, but not its short published version of 1930, contains a deep remark by which the proof of completeness cannot be finitary because such a proof would give a decision method for predicate logic.

4. Encounter in Königsberg. There exists a very short and readable lecture about completeness in Gödel's hand, namely the one he gave in a conference in Königsberg in early September 1930. Close to the end of that lecture, we find the following passage [Gödel 1930c, 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 [Entscheidungsdefinitheit] follows. 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, not later reconstruction.

The second version of the incompleteness paper has, after some fifteen pages, the title "Meine Damen und Herren!" Then comes the text of the Königsberg lecture on completeness in shorthand. The ending is:

> 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.

The last sentence reads in German: “Doch sind diese Dinge noch zu wenig durchgearbeitet um hier näher auf einzugehen.” In the typewritten version, we read somewhat differently about his proof of the failure of completeness:

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

The German is: “Auf diese Dinge kann ich aber hier nicht näher eingehen.” Then this last sentence is cancelled and the following written: “Doch würde es zu weit führen, auf diese Dinge näher einzugehen” (It would, though, take us too far to go more closely into these things). 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. This must have been in the summer of 1930.

Just a few pages before the Königsberg outbreak, Gödel writes that the formally undecidable sentences have “the character of Goldbach or Fermat,” i.e., of universal propositions that can be refuted by a numerical counterexample. 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 generalisation would at once give a contradiction.

Among Gödel’s audience in Königsberg sat Johann von Neumann, who reacted at once and wanted more explanations. Gödel gave such in a discussion among the two and most likely during his stay in Berlin immediately after. The most detailed account of these events is [Wang 1996], section “Some facts about Gödel in his own words” [ibid., p. 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 formalised to show the existence of undecidable propositions without giving any individual instances.

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.”

Von Neumann lectured from late October 1930 on in Berlin on “Hilbert’s proof theory” of which Carl Hempel, later a very famous philosopher, has 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; for example, it is known that von Neumann got the proofs of Gödel's paper around the tenth of January 1931. The lectures of late 1930 were based on other sources to be presented below.

Jacques Herbrand 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 from late spring 1931 to July in Hamburg and Göttingen. 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.[2]

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 tell about 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.

Gödel's account, as reported by Wang, suggests that he had found this result right after the Königsberg meeting; it is further clear that he must have explained it to von Neumann during his visit to Berlin right after.

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 $\mathfrak{W}$.

[2] [von Plato 2017b], section 8.3 on two "Berliners" contains a detailed account of Herbrand's stay in Germany and his relation to von Neumann.

> …
>
> I show now: $\mathfrak{W}$ is always unprovable in systems free of contradiction, i.e., a possible effective proof of $\mathfrak{W}$ could certainly be transformed into a contradiction.

Gödel must have explained to von Neumann the essential point, not just a blunt statement of incompleteness, namely that provability of a formula in a system can be expressed as a formula of that system, here provability of $1 = 2$.

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 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."

5. Gödel's lost reply to von Neumann's letter.

Gödel's final shorthand version for his incompleteness paper occupies the first 39 pages of a notebook. It begins very closely the way the typewritten version does. Even footnotes are numbered consecutively until number 29 on page 24 of the manuscript. 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. Every 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 the following letter-sketch:[3]

[3] A word about the nature of shorthand sources is in place here: The transcription of shorthand is by the very nature of the script, with missing endings of words and abrupt shortenings–a single letter can stand for different words that have to be figured out from the context–also error-bound interpretation and guesswork. There are in addition uncertainties for reasons such as faded sources, badly written or heavily cancelled passages, etc. I have no pretense to a grammarian's exact reading, word for word, but offer my English translations as accounts of what Gödel wrote down about 87 years ago, in the hope that they appear consonant with Gödel's thought, with the suggestion to anyone who should like to quote them to give their own interpretation of the text. At places, I

> Dear Mr von Neumann
>
> Many thanks for your letter of [20 November]. Unfortunately I have to inform you that I am already since about three months in possession of the result you communicated. It is also found in the attached offprint of a communication to the Academy of Sciences. I had finished the manuscript for this communication already before my departure for Königsberg and had presented it to Carnap. I gave it over for publication in the *Anzeiger* of the Academy on 17 September. [Cancelled: The reason why I didn't make any presentation [written heavily over: didn't tell anything] of the above result is that the precise proof is not suited to oral communications and an approximate indication could easily arouse doubts about the correctness…that would not convince]. As concerns the publication of this matter, there will be given only a shorter sketch of the proof of impossibility of freedom from contradiction in the *Monatshefte* that will appear in the beginning of 1931 (the main part of this treatise will be filled with the proof of existence of undecidable sentences). The detailed carrying through of the proof appears in a *Monatsheft* only in July or August. I can send you proofs in a few weeks.
>
> I shall include a part of my work that concerns the proof of freedom from contradiction so that you can state to what extent your proof matches mine.
>
> The carrying out of the proof appears together with my proof of undecidability in the next volume of the *Monatshefte*. I didn't want to talk about it further provisionally because this thing (even more than the proof of undecidability) must arouse doubt about its executability before it is laid out in a concrete way.

There are eight pages between the first and second versions of the letter, filled with Gödel's attempts at formulating the second incompleteness theorem in various ways and how it should be proved, until a second letter sketch:

> Dear Mr von Neumann!
>
> Hearty thanks for your letter of 20·/XI. The result of which you write to me is known to me since already about three months, but I didn't want to talk anything about it before I had done it in a print-ready form. I send you enclosed an offprint in which the proved theorem gets expressed. The manuscript of

have left a question mark in place of a word or two that I failed to read. A sentence can be informative even if one doesn't see what precise verb is used in it, say. There are very many cancelled passages in the letter sketches; I have included just some of these details, with the aim of a readable result in mind, as close to something Gödel may have intended to send to von Neumann.

this communication to the Academy was finished already before my departure to Königsberg and presented it to Carnap. I gave it over for publication in September. The carrying through of the proof will appear together with the proof of undecidability in a near *Monatsheft* (beginning of 1931). I shall have proofs of this work in a few weeks and will then send them to you immediately.

Now to the matter itself. A basic idea of my proof can be described (quite roughly) like this. The sentence A that I have put up and that is undecidable in the formal system S asserts its own unprovability and is therefore correct. If one analyses precisely how this undecidable sentence A could still be metamathematically decided, it appears that this is possible only under the condition of the freedom of contradiction of S. That is, it is strictly taken not A but $W \to A$ that is proved (W means the proposition: S is free from contradiction). The proof of $W \to A$ lets itself be carried through, though, within the system S, so that if even W is provable in S, then also A which contradicts the undecidability of A.

As concerns the meaning of this result, my opinion is that *only* the impossibility of a proof of freedom from contradiction for a system *within this system* is thereby proved. For the rest, I am fully convinced that there is [cancelled: a finite] an intuitionistically unobjectionable proof of freedom of contradiction for classical mathematics [added above: and set theory], and that therefore the Hilbertian point of view has in no way been refuted. Only one thing is clear, namely that this proof of freedom from contradiction is in any case far more (?) complicated than had been assumed so far.

As concerns the question that remains, my opinion is that exists no formal system in which all [cancelled: intuitionistically unobjectionable constructive] finite proofs would be expressible.*[4] Still, I would like very much to hear about your contrary argument concerning the matter. I would be further interested whether your proof is built on the same thought as mine, something I hope all the same from what you intend in relation to publication, namely that you relate your work to mine.

Unfortunately, nothing seems to come of my travel to Berlin this year.

In the hope of a swift reply, I remain with

[4] [Ed: The asterisk directs to an addition at the end of the letter sketch:]

* From the treatise of P. Bernays on "Philosophie der Mathematik und die hilbertsche Beweistheorie" in the *Blätter für Deutsche Philosophie*, volume 4, issue 3/4, 1930, I gather that this is also the view of Hilbert and Bernays (cf. what is said on page 366).

best wishes, yours sincerely

6. HERBRAND'S TESTIMONY. 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 argument for the second 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 $\underline{\underline{\text{in T}}}$

Next in Herbrand's letter comes von Neumann's striking addition to Gödel's first theorem: with $D(x, A)$ standing as above for: proof number x demonstrates proposition A, Herbrand writes in the letter the magic 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 $\sim A$, i.e., $\sim\exists x\exists z(D(x,A)$ et $D(z,\sim A))$, or in a free-variable formulation, for each x and z, $\sim(D(x,A)$ et $D(z,\sim A))$.

At the time Herbrand wrote to Vessiot, 28 November, the "absolutely unexpected results" he alludes to are perhaps an indication of von Neumann's version of the second theorem. By 29 November, von Neumann has read Gödel's letter of reply and that shows in Herbrand's letter to Chevalley of 3 December. Gödel had explained to von Neumann that the second theorem is proved by first showing an implication *within* the formal system. The details are found in the interim pages between the two letter sketches–with even references to the incompleteness paper. Here $\mathfrak{K}$ is any "recursive consistent class" of formulas:

> Let us now turn back to the undecidable proposition $17Gen\,r$. The proposition that $\mathfrak{K}$ is free from contradiction will be denoted by $Wid(\mathfrak{K})$ for the proof that $17Gen\,r$ is unprovable, and only the freedom of contradiction of $\mathfrak{K}$ is used (cf. 1.) on page 30) so we have
>
> $$Wid(\mathfrak{K}) \rightarrow \overline{Bew_{\mathfrak{K}}}(17Gen\,r)$$

If now $Wid(\mathfrak{K})$ were provable within the system, also the unprovable sentence $\overline{Bew_{\mathfrak{K}}}(17Gen\,r)$ would, which is impossible.

In von Neumann's second letter to Gödel, of 29 November, he writes:

> I believe I can reproduce your sequence of thoughts on the basis of our communication and can therefore tell you that I used a somewhat different method. You prove $W \rightarrow A$, I show independently the unprovability of W, though with a different kind of inference that likewise copies the antinomies.

Von Neumann's proof idea brings to mind Gödel's early formulations of the unprovability of consistency. More cannot be said unless notes for the course are found somewhere. The lectures must have been widely attended, but I have been able to secure only Hempel, Herbrand, and B. H. Neumann, and very likely Gerhard Gentzen as participants.

There is a third letter of von Neumann's of 12 January 1931, after he had received the page proofs of Gödel' article, in which he sketches what he describes as a "somewhat shorter carrying out of the unprovability of freedom from contradiction."

7. Gödel in panic. The sequence of events in and around Gödel's two sketches of letters is psychologically interesting. He was of course worried about von Neumann's plans: First he wants to assure von Neumann that he had both results, even mentioning Carnap

as witness and quoting 17 September as the date he sent in the short note to press. He writes that he will copy a part of his manuscript for the incompleteness paper, about the second theorem, etc. Then come eight pages of attempts at a satisfactory formulation, and the second letter sketch in which just the proofs of the incompleteness article are promised once they arrive.

The pages between Gödel's two letter sketches to von Neumann are his notes for section 4 of his incompleteness paper. An inspection of his typewritten manuscript shows that the last three lines of page 41 have been cancelled. They contain the beginning of his closing paragraph as in the shorthand manuscript. Pages 42–44 contain the added section 4. The first proofs have a "I" added in the end of the title, a paragraph that explains the second theorem added at the end of the introduction, and a long footnote on the second theorem added in another place. The original proofs have no mention at all of the second theorem before section 4 that Gödel wrote, and must have taken directly to the printer's, some time after he had received von Neumann's letter.

A shadow is cast on Gödel's great achievement; there is no way of undoing the fact that Gödel played a well-planned trick to persuade von Neumann not to publish. In his letter of reply, he reproduced details from section 4, freshly written after von Neumann's letter, but he also included his short note of October 1930 that contains a statement of the second theorem. The latter would have been enough, but Gödel panicked at the prospect of von Neumann publishing his second theorem. The writing is quite nervous, with cancellations and additions all over. Moreover, the first proofs that reveal his trick must have caused him quite a stress; nothing he could send to von Neumann who would have wondered why the magnificent second incompleteness theorem is not even mentioned in the lengthy introduction. He got page proofs for the article only around the tenth of January.

Concerning the October 1930 one-page notice to the Vienna academy, the last page of the shorthand manuscript instructs to add to page 1 a reference to this note. There is in the title (!) of Gödel's article a footnote that points to it, without further explanations. The microfilms contain a typewritten copy with a stamp "Akademie der Wissenschaften in Wien, Zahl 721/1930 eingefangt: 21.X.1930." The wording of "Satz II" is well known:

> Even when one allows in metamathematics all the logical means of the *Principia Mathematica* (especially therefore the extended functional calculus with the axiom of reducibility or without ramified type theory and the axiom of choice), there is *no proof of freedom from contradiction* for the system S (and even less if one restricts the means of proof in some way). Therefore, a proof of freedom from contradiction of the system S can be carried through only by methods that lie *outside* the system S, and the case is analogous for other formal systems, say the Zermelo-Fraenkel axiom system for set theory.

Having sent the note to von Neumann, it is clear that the latter had no new result to publish, and there would have been no need for Gödel to change anything, at most mention the results in the short notice. The formulation also confirms what I said above, namely that Gödel's early metamathematics used strong methods. Moreover, the printed text mentions ω-consistency, but in the manuscript and in the notes before Königsberg, Gödel always wrote $\aleph_0$-consistency, the latter a distinctly set-theoretic notation.

The typewritten manuscript with the typesetters' leaden fingerprints on it contains three lines at the end of page 41, and the rest exists only in his shorthand:

> To finish, let us point at the following interesting circumstance that concerns the undecidable sentence A put up in the above. By a remark made right in the beginning [page 41 ends here, in the shorthand the letter S is used instead of A], S claims its own unprovability. Because S is undecidable, it is naturally also unprovable. Then, what S claims is correct. Therefore the sentence S that is undecidable in the system has been decided with the help of metamathematical considerations. An exact analysis of this state of affairs leads to interesting results that concern a proof of freedom from contradiction of the system P (and related systems) that will be treated in a forthcoming continuation of this work.

Gödel shows here a cautiousness the editor of his *Collected Works* Sol Feferman liked to emphasise about him, just "interesting results" about consistency. The thought of von Neumann publishing the second theorem must have haunted him and led to the hasty addition of a section on results so far "zu wenig durchgearbeitet" as he put his closing words in the Königsberg lecture. In fact, Gödel was unable to prove the second theorem to his satisfaction and no "Part II" of the incompleteness paper ever appeared, neither do the shorthand notes suggest any such work even in manuscript form.

8. INCOMPLETENESS BEFORE THE SECOND THEOREM? We have now looked ahead from the Königsberg meeting; let's look back also. Among Gödel's preserved notes for the incompleteness article, the last one is, as noted, very close to the printed paper. The first of these notebooks is a rather carefully composed set that seems to have been written for an article before Gödel had found the second theorem, so before August 1930 and before the part of the second notebook that was written down before the notes for the Königsberg lecture. This timing is in accordance with what Gödel wrote to von Neumann toward the end of November, namely that he had been in possession of the second theorem for some three months, and with what he told Hao Wang in 1976–77. Therein we find Gödel state that he discovered his second theorem "shortly after the Königsberg meeting." Therefore,

anything that precedes the Königsberg lecture notes in the second suite of notes for incompleteness, must be before the second theorem about the unprovability of consistency had surfaced.

The first version of the incompleteness article opens with the words:

> The question whether every mathematical problem is solvable, i.e. whether for every mathematical proposition A either A or non A is provable, was so far devoid of a concrete sense, because the words "mathematical proposition" and "mathematically provable" had not been made precise. The opinions of various mathematicians diverge strongly on this point, as is shown sufficiently by the discussions over the axiom of choice and the law of excluded middle. The way to make for precision that is at the basis of the investigation at hand is essentially the one given in the *Principia Mathematica.*

A detailed examination of this early incompleteness work has to await another occasion. Let us note two interesting remarks therein. Page 32 has:

> It is easy to convince oneself by complete induction about the correctness of the following theorem:
>
> *Every provable formula is true* because the axioms are obviously true and this property is not destroyed by the rules of inference. This result can be proved, though, only with the help of the axiom of choice.

The passage refers to a formal system that contains higher-order logic. On page 18, the nature of metamathematics is described:

> No limitations in the means of proof are required. One can use all the theorems and methods of analysis, set theory, etc in metamathematical proofs. A proof of a metamathematical theorem conducted in such a way is comparable to a proof in analytical number theory.

At the end of the more than forty pages of notes, there is a text for an introduction that begins with:

> In what follows, a proof is sketched in coarse outline by which Peano's axioms with the logic of the *Principia Mathematica* (natural numbers as individuals) don't form any system definite with respect to decidability, even allowing the axiom of choice. In other words, there are in the system unsolvable problems, even of a relatively simple structure.

The second set of Gödel's notes for incompleteness is, as its first page tells, "a provisional version." It gets a fairly good dating by the presence of the Königsberg lecture in it:

anything before the lecture text is before the Königsberg meeting. An "exact definition" of the notion of truth is given in the earlier part and the theorem stated that all provable formulas are true. A proposition is then constructed by arithmetic coding that states its own unprovability. If it is provable, it is true, so must in fact be unprovable.

The critical point in the truth definition is with the universal quantifier:

> $(x)F(x)$ shall be called true when and only when for every number n [of right type], $F(n)$ is true. This definition fails in that it presupposes that there are names for all classes and relations which certainly is not the case (there are especially only denumerably many names).
>
> …
>
> When one asks by what means not contained in the system S undecidability was concluded, the answer can be only: through the definition of truth that extends type theory into the transfinite.

A footnote tells: "The idea of such a definition has been expressed [cancelled: simultaneously] independently by Mr A. Tarski of Warsaw." Tarski had lectured in Vienna in February 1930 and a letter of Gödel's of 2 April 1931 to Bernays even recollects a discussion on the topic with Tarski. It seems clear that at that time, Gödel was trying to prove the completeness of higher-order logic and needed a truth definition for the soundness part. The other direction failed, though.

Gödel has seen clearly the critical point, namely that the syntactic condition of provability of $F(x)$ with a free variable suffices for $(x)F(x)$ in predicate logic, whereas universal quantification in higher-order logic becomes a transfinite notion.

Soon after the Königsberg lecture notes break in the shorthand, Gödel saw that one can restrict the methods used in metamathematics. This change was prompted by von Neumann's suggestion in Königsberg. Close to fifty years later, Gödel regretted not having mentioned the suggestion [Wang 1996, p. 84]. The proof of the first theorem in the final version is, as Gödel emphasised, carried out "constructively," and he planned undoubtedly to do the same with the second theorem. There would then be two versions each of the two incompleteness theorems.

9. Gödel's sources, cited and uncited. 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.* Substituting unprovable for false, one gets a statement that expresses its own unprovability. The explanation is good, 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. Another explanation was that he tried to prove the consistency of analysis relative to first-order

arithmetic. This explanation has an affinity with the early formulations of incompleteness.

The middle version, where the definition of truth is given, makes the following comment after the proposition that states its own unprovability is shown to be true and therefore unprovable:

> One recognises a close connection of this proof to the Richard antinomy and it can be expected that even other epistemic antinomies can be reorganized into analogous proofs, something that actually is the case.

Hilbert-Ackermann contains a lengthy discussion of such paradoxes and the analogy was therefore fresh in Gödel's mind. The effect of starting in the 1931 paper with the heuristic analogy gave the impression–whether planned or accidental–that the paradoxes were his way to the result, an impression that created an unprecedented aura of genius around the discovery and around him, shared by von Neumann and everyone else who read his finished paper.

Gödel's meticulously kept notes and other material point at interesting circumstances that concern his discovery of the undecidable sentences, to be treated in a forthcoming continuation of this work. Let me just refer to a couple of unmentioned sources: Gödel had begun work on incompleteness in the summer of 1930 by [ibid., p. 82]; I would say perhaps May). Gödel's library request cards show that he had taken out in April Fraenkel's *Einleitung in die Mengenlehre* in which the question of completeness of mathematical theories is discussed. The most poignant remark is that "there should be nothing absurd in imagining that the unsolvability of a problem could even be *proved*" (p. 235).

On 13 May and again right after Königsberg on 12 September while in Berlin, Gödel borrowed an obscure Norwegian journal issue, Skolem's separately published 49 page "Über einige Grundlagenfragen der Mathematik," of the previous year. There Skolem states a version of the "Skolem paradox," namely that the theorems of a truly formal system are denumerable, indeed they can be ordered lexicographically, but that the properties of natural numbers cannot be in that way ordered, by which (p. 269):

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

Gödel wrote down detailed summaries of the works he read. In his three page summary of Skolem's paper, we read 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.

References

P. Bernays (1918). "Beiträge zur axiomatischen Behandlung des Logik-Kalküls". Manuscript Hs. 973:193, Bernays collection, ETH-Zürich, printed in [Hilbert 2013].

R. Carnap (1929). *Abriss der Logistik*. Springer.

J. Dawson Jr. (1997). *Logical Dilemmas. The Life and Work of Kurt Gödel*. A K Peters, Ltd., Wellesley, MA, pp. xiv+361. MR: 1429389.

A. Fraenkel (1928). *Einleitung in die Mengenlehre*. 2nd ed. Springer.

K. Gödel (1930a). "Die Vollständigkeit der Axiome des logischen Funktionenkalküls". *Monatsh. Math. Phys.* 37.1, pp. 349–360. MR: 1549799.

– (1930b). "Einige metamathematische Resultate über Entscheidungsdefinitheit und Widerspruchsfreiheit". *Anzeiger Akad. Wiss. Wien* 67, pp. 214–215.

– (1930c). "Vorlesung über Vollständigkeit des Funktionenkalküls". In: vol. III. First printed in [Gödel 1986], pp. 16–29 (cit. on p. 4079).

– (1931). "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I". *Monatsh. Math. Phys.* 38.1, pp. 173–198. MR: 1549910.

– (1986). *Collected works. Vol. I*. Publications 1929–1936, Edited and with a preface by S. Feferman. Oxford University Press, New York, pp. xvi+474. MR: 831941 (cit. on p. 4092).

W. Goldfarb (2005). "On Gödel's way in: the influence of Rudolf Carnap". *Bull. Symbolic Logic* 11.2, pp. 185–193. MR: 2132114 (cit. on p. 4079).

C. Hempel (2000). "An intellectual autobiography". In: *Science, Explanation, and Rationality*. Springer, pp. 3–35 (cit. on p. 4081).

D. Hilbert (2013). *David Hilbert's Lectures on the Foundations of Arithmetic and Logic, 1917–1933*. Vol. 3. Edited by W. Ewald, W. Sieg and M. Hallett. Springer, pp. xxv+1062. MR: 3551948 (cit. on pp. 4076, 4092).

D. Hilbert and W. Ackermann (1928). *Grundzüge der theoretischen Logik*. Die Grundlehren der mathematischen Wissenschaften, Band 27. Springer, pp. viii+120. MR: 0351742 (cit. on p. 4076).

P. Mancosu (1999). "Between Vienna and Berlin: the immediate reception of Gödel's incompleteness theorems". *Hist. Philos. Logic* 20.1, pp. 33–45. MR: 1700712 (cit. on p. 4081).

J. von Plato (2017a). *Saved from the Cellar, Gerhard Gentzen's Shorthand Notes on Logic and Foundations of Mathematics*. Sources and Studies in the History of Mathematics and Physical Sciences. Springer, Cham, pp. x+315. MR: 3642843 (cit. on p. 4077).

– (2017b). *The Great Formal Machinery Works, Theories of Deduction and Computation at the Origins of the Digital Age*. Princeton University Press, Princeton, NJ, pp. viii+377. MR: 3676146 (cit. on p. 4081).

– (2018). "Kurt Gödel's first steps in logic: formal proofs in arithmetic and set theory through a system of natural deduction". *Bull. Symb. Logic* 24 (cit. on p. 4076).

T. Skolem (1929). "Über einige Grundlagenfragen der Mathematik". *Norsk videnskaps-akademi i Oslo. Skrifter I. Mat.-naturv. klasse* 4.

A. M. Turing (1936). "On computable numbers, with an application to the Entscheidungsproblem". *Proc. London Math. Soc. (2)* 42.3, pp. 230–265. MR: 1577030 (cit. on p. 4077).

H. Wang (1996). *A Logical Journey, From Gödel to Philosophy*. MIT Press, Cambridge, MA, pp. xiv+391. MR: 1433803 (cit. on pp. 4080, 4090, 4091).

A. Whitehead and B. Russell (1927). *Principia Mathematica*. 2nd ed. Vol. I–III. Cambridge (cit. on p. 4076).

Received 2017-12-07.

Jan von Plato
University of Helsinki
jan.vonplato@helsinki.fi