The Writing of *Introduction to Metamathematics*

Stephen C. Kleene

It was suggested that I talk here on the writing of my book *Introduction to Metamathematics* (IM).

In the summer of 1936 I drew up an outline for a graduate course on the foundations of mathematics, the first to be given at the University of Wisconsin, Madison. It ran in the fall semesters of 1936–37, 1938–39, and 1940–41, as well as more recently.

In the fall of 1939, Saunders MacLane suggested to Rosser and me that we jointly write a Carus monograph on foundations. But instead of doing that (the Carus monographs are rather compact books, tending to be of less than 200 pages), each of us separately wrote a book of over 500 pages. My *Introduction to Metamathematics*, was published in 1952. The ninth reprint by the Dutch publishers was in 1988.[1] According to incomplete records, about 17,500 copies were sold through 1986, not counting sales of a reprint in Taiwan and of two in Japan, of two printings of the Russian translation (the first consisting of 8000 copies), of one of the Spanish translation, and of one of the Chinese translation (in two volumes).[2] I must leave to others the assessment of the role the book may have played in the teaching of mathematical logic and foundations over the years.

I calculated that all of my spare time for $7\frac{1}{2}$ years went into the composition of the book. But the earlier preparation of my logic course was determining for the first ten chapters, which essentially followed that course. Subsequent chapters largely contain material used in seminars given in the spring semesters following the course beginning with 1938–39, except for the last two chapters (XIV and XV).

So what went into the course, the seminars, and the book? Mostly topics that stood out in the landscape of mathematical logic for me as an observer in the 1930s.

In 1983, I wrote, "Gödel's paper *1931* was undoubtedly the most exciting and the most cited article in mathematical logic and foundations to appear in the first 30 years of the century."[3] I learned of that paper, shortly after its publication, while I was a graduate student at Princeton. In subsequent research, I used Gödel's method (which I learned from it) of numbering the formal objects, and the Herbrand–Gödel notion of "general recursive functions" (from Gödel *1934*), in giving another version of Gödel's first *1931* incompleteness theorem in *1936*, and more simply in *1943* (abstract published in 1940) in connection with my arithmetical hierarchy (given independently by Mostowski in *1947*). I had originally intended to publish the details on my arithmetical hierarchy (§57 in the book) only through the book, hoping thereby to promote its sales. But I decided at the beginning of 1942 to publish them separately, when I realized that the publication of the book was far off. The writing of it had to be set aside during my military service in 1942–1945, while the uncompleted manuscript reposed in Saunders MacLane's office in Widener Library at Harvard.

What came to my mind in 1936 for my new course on the foundations mathematics? Gödel's *1931* results applied to Hilbert's uncompleted program for vindicating classical mathematics (which had been shaken by the paradoxes of set theory) by embodying a suitable portion of it in a formal system and proving that system consistent by safe ("finitary") methods in a new mathematical discipline to be called "proof theory" or "metamathematics." It struck me that the most exciting developments in foundations then centered around this program, and results such as Gödel's which came out of contemplating it. So I drew up a syllabus for the course which would give the student the whole broad picture of Hilbert's program (with the context in which it arose and with comparisons with the other outstanding foundational approaches), and which would allow one to present Gödel's theorems conveying a full understanding of them.

Thus I had to present the idea of a formal system, and train the student intensively in working with one. As I remarked in the preface of the book, the simplest formal system for the purpose of exhibiting such results as Gödel's is one for elementary number theory (based on first-order logic).

I put into the course (and book) three introductory chapters to set the enterprise in its broad historical context, and to present necessary fundamental concepts.

Then there followed (in Chapters IV–VIII) my specimen of a formal system, studied in stages. I think students of logic, confronted with the axioms and rules of inference of a formal system, had to rack their brains rather hard to construct proofs of formal theorems in it. The deduction theorem of the first-order predicate calculus (first proved by Herbrand in *1930*) provides great assistance in proving implications. It was emphasized in another context in Church's logic course in the fall of 1931–32. I used a similar treatment of disjunctions (proof by cases) in my *1934* paper. In brief, it is an easy step from the deduction theorem to a similar treatment of proofs of formulas involving each of the logical constants. Thus (originally in my 1936–37 course) I came to my treatment of logic in IM, §23 via a set of derived rules for the introduction and elimination of logical symbols. (One also finds versions of it in Jaśkowski *1934*, Gentzen *1934–35*, and Bernays, *1936*.) I followed von Neumann *1927* when I was treating logic as a subsystem of formal number theory, in using axiom schemata instead of propositional and predicate variables with formal substitution rules for them.

I arrived at the formulation and basic discussion of Gödel's two famous *1931* incompleteness theorems at the end of Chapter VIII. In the next two chapters (IX and X) I gave a treatment of primitive recursive functions (what Gödel in *1931* called simply "recursive functions") and completed the proof of his first incompleteness theorem.

I continued in seminars, and thence in the book, with chapters on general recursive functions (Gödel *1934*, adapting a suggestion of Herbrand, and Kleene *1936*), partial recursive functions (Kleene *1938*, discovered late in 1936), and computable functions (Turing *1936–37* and Post *1936*).

My treatment of computable functions, considerably reworked from Turing's account and in some respects closer to Post *1936*, was first introduced in my seminar in the spring of 1941. (That seminar was honored by the presence of Richard Brauer.) Turing applied his machines primarily to compute the dual expansions of real numbers x ($0 \leq x \leq 1$), the successive digits being printed *ad infinitum* on alternate squares of a one-way infinite tape, while the intervening squares were reserved for temporary notes used as scratch work in the continuing computation. (Many of Turing's technical details were incorrect as given, and a person who wishes to follow his treatment in detail will profit from the critique in the appendix to Post *1947*.) I respected and took over Turing's brilliant conception of the kinds of operations a human, or physical computer can perform, and thus the basic mode of operation of his machines, But, continuing from my Chapters IX, XI, and XII, I wanted rather to consider the computability of number-theoretic functions. I chose to use a machine of his sort to compute a given such function by supplying it with a tape on which is printed an argument (or n-tuple of arguments) of the function, and asking that the machine eventually come to a stop with the corresponding function value printed on the tape following the argument(s). A natural number n, as argument or value, I represented by $n + 1$ tallies printed on consecutive squares, with blank squares separating the representations of two numbers.

In his §10, Turing suggests that possibly the simplest way to use a machine of his to compute, e.g., a function $\phi(n)$ of "an integral variable," n would be to let it compute the sequence of 0s and 1s with the number of 1s between the nth and $(n + 1)$th 0 being the value of $\phi(n)$.[4] That, in my opinion, would certainly be more cumbersome to work with than my method. Moreover, that method breaks down when one seeks to apply it to the partial recursive functions which I introduced in *1938* (and §63). Wang (*1974*, p. 84) says, "Gödel points out that the precise notion of mechanical procedures is brought out clearly by Turing machines producing partial rather than general recursive functions. In other words, the intuitive notion does not require that a mechanical procedure should always terminate or succeed. A sometimes unsuccessful procedure, if sharply defined, still is a procedure, i.e., a well determined manner of proceeding. Hence we have an excellent example here of a concept which did not appear sharp to us but has become so as a result of a careful reflection. The resulting definition of the concept of mechanical by the sharp concept of 'performable by a Turing machine' is both correct and unique. Unlike the more complex concept of always-terminating mechanical procedures, the unqualified concept, seen clearly now, has the same meaning for the intuitionists as for the classicists. Moreover, it is absolutely impossible that anybody who understands the question and knows Turing's definition should decide for a different concept."

I remember that, when in the spring of 1940 I used the words "partial recursive function" in a conversation with Gödel, he immediately asked, "What is a partial recursive function?". Church in *1936* had avoided the concept of "partial recursive function" by using "potentially recursive function."

Having started out to build the course (and the first ten chapters of the book) toward getting Gödel's incompleteness theorems, it was essentially irresistible to proceed to these next three chapters (XI, XII, and XIII). There had been such exciting developments in the foundational research at Princeton (involving Church, Gödel, Rosser, and myself) just before my coming to Wisconsin in the fall of 1935, culminating in Church's thesis (proposed in 1934 and 1935 and published in 1936) and its applications, and in Turing's *1936–7* work, and Post's *1936*, of each of which I learned a bit later! For the history of these developments, see Kleene (*1981*) and Davis (*1982*).

In brief, when I was invited in 1936 to teach a graduate course in the foundations of mathematics, and later (first in 1939) to extend it by a seminar, these were the topics which seemed to me the most exciting. Naturally, I was influenced by where my own research interests lay. There was no existing *connected* treatment of them in an English text book that I could just follow. (Quite a bit of the material was available in German in Fraenkel *1928*, Hilbert and Ackermann *1928*, Heyting *1934*, and Hilbert and Bernays *1934* and *1939*.) So I designed the course (giving my students dittoed notes), expanded it in seminars, and subsequently wrote it all into my own textbook.

Incidentally, the symmetric form of Gödel's theorem (§61), which I discovered in 1946–47, was in the manuscript of my book when I showed it to Mostowski on his visit to Madison in April 1949, on whose urging I published it separately in *1950*. Kreisel wrote me on 4 October 1983 that Gödel on several occasions remarked on this symmetric form as being a very significant improvement of his incompleteness results. Kreisel continued, "The reasons are so patently clear that I did not bother to ask him to elaborate." I did not ask Kreisel to elaborate. But I will attempt now to do so myself.

Fundamental requirements on a formal system are as follows:

(1) It should be effectively recognizable what linguistic objects (formulas) of the system express certain well-defined propositions of mathematics—those of a domain of intuitive mathematics that we are choosing the formal

system to include a formalization of. If this domain includes some elementary number theory, then by Church's *1936* thesis discussed in §62, to any well-defined one-place number-theoretic predicate (propositional function of one natural number variable x) of this domain, the Gödel number of the formula expressing the proposition taken as its value for a given natural number as argument (i.e., as value of x) will be a general recursive function of the argument x. We are assuming Gödel numbers assigned to symbols, strings of symbols, and strings of strings of symbols in the standard way.

(2) It should be possible to check effectively without fail whether a string of formulas we have before us constitutes a proof in the system. Then, by Church's thesis, to be the Gödel number of a proof is a general recursive predicate. There is a primitive recursive function which, applied to the Gödel number of a proof, gives the Gödel number of the formula proved by it (its last formula).

(3) And of course, for any propositions to which the qualities of being true or false apply in a clear way, we want a proof of the formula asserting one of those propositions to exist only when the proposition is true. For my symmetric form of Gödel's theorem, I use only as much of this as is assured by simple consistency (after Rosser *1936* and my *1943*, top p. 64).

In the *1931* version of Gödel's first incompleteness theorem, the formal systems to which it applied were "*Principia Mathematica* (Whitehead and Russell *1910–12*, second edition *1925*, *1927*) and related systems." It was far from clear then that the theorem would apply to all formal systems satisfying in reasonable measure the aforesaid requirements (inclusive of systems quite remote in their details from Gödel's examples), except ones so weak (unrobust) as to be of slight interest.

That the theorem applies to all conceivable simply consistent slightly robust systems is given by my symmetric form of Gödel's theorem. For the robustness, what I ask is that they include a formalization of the small piece of elementary number theory which I now describe.

I use two recursively enumerable sets $C_0 = \check{x}(Ey)W_0(x, y)$ and $C_1 = \check{x}(Ey)W_1(x, y)$ (the variables range over the natural numbers $0, 1, 2, \ldots$), where $W_0(x, y)$ and $W_1(x, y)$ are two particular primitive recursive predicates (defined on p. 308). For each of $i = 0$ or 1, and any given x such that $x \in C_i$ (i.e., such that $(Ey)W_i(x, y)$), a proof of that fact in intuitive elementary number theory exists, resting simply on our ability (because W_i is primitive recursive) to verifty for the given x and a suitable y that $W_i(x, y)$ is true. Furthermore, an easy argument in intuitive elementary number theory shows that my two sets C_0 and C_1 are disjoint.

For a formal system to come under my symmetric form of Gödel's theorem, it should formalize the foregoing, thus:

(a) For each of $i = 0$ and 1 and each x, it shall have a formula (which we can find effectively from the i and x) expressing the proposition $x \in C_i$ (i.e.,

$(Ey)W_i(x, y))$, and likewise one expressing $x \notin C_i$. I shall denote these formulas by "$\exists y W_i(\mathbf{x}, y)$" and "$\neg\exists y W_i(\mathbf{x}, y)$," respectively, avoiding placing any restriction on the actual symbolism of the system; and for simple consistency I merely understand that for no i and x are both of those formulas provable.

(b) Corresponding to the intuitive provability of $x \in C_i$ whenever true, in it for each of $i = 0$ and each x:

$$(*) \qquad \text{If } x \in C_i, \text{ then } \exists y W_i(\mathbf{x}, y) \text{ is provable.}$$

(c) Corresponding to the intuitive proof of the disjointness of C_0 and C_1, in it for each x:

$$(**) \qquad \begin{array}{l} \text{If } \exists y W_0(\mathbf{x}, y) \text{ is provable, then } \neg\exists y W_1(\mathbf{x}, y) \text{ is provable.} \\ \text{If } \exists y W_1(\mathbf{x}, y) \text{ is provable, then } \neg\exists y W_0(\mathbf{x}, y) \text{ is provable.} \end{array}$$

My symmetric form of Gödel's theorem shows that, for each formal system which encompasses this small piece of number theory and is simply consistent (no matter what novelties of symbolism, or of methods of proof, it may have), we can find a number f such that the formula $\neg\exists y W_0(\mathbf{f}, y)$ is true but unprovable while $\exists y W_0(\mathbf{f}, y)$ is also unprovable (and symmetrically, a number g such that $\neg\exists y W_1(\mathbf{g}, y)$ is true while it and $\exists y W_1(\mathbf{g}, y)$ are both unprovable).

Thus, for each such formal system, we have as a formally undecidable proposition $(Ey)W_0(f, y)$ a respective value of one preassigned predicate $(Ey)W_0(x, y)$, a result which Gödel did not have in *1931*. Just the theory of that particular predicate provides inexhaustible scope for mathematical ingenuity.[5]

The symmetric form of Gödel's theorem, as I stated it in abstract terms in *1950*, says simply that to any two disjoint recursively enumberable sets D_0 and D_1 with $D_0 \supset C_0$ and $D_1 \supset C_1$, a number f can be found such that $f \notin D_0 \cup D_1$. Applying this by taking D_0 to be the x's such that $\exists y W_0(\mathbf{x}, y)$ is provable (so by (*), $D_0 \supset C_0$) and D_1 to be the x's such that $\neg\exists y W_0(\mathbf{x}, y)$ is provable (so by the simple consistency, D_1 is disjoint from D_0; and by (*) and (**), $D_1 \supset C_1$), the f we get does the job.[6]

Chapter XIV picked up some standard topics of mathematical logic which one would recognize in the 1930s as important and which had not found a place in Chapters I–XIII. Those earlier chapters provided background material, using which these topics could be treated compactly. Among these topics were some outstanding model-theoretic results (Gödel's *1930* completeness theorem, including the theorem of Löwenheim *1915* and Skolem *1920* extended to include "compactness," in §72; and Skolem's *1922–23* paradox, and his *1933*, *1934* nonstandard arithmetics, in §75), as well as some more topics in proof theory (metamathematics). After all, I had to live up to the title of the book; and one of the jewels of proof theory was eliminability theory after Hilbert and Bernays *1934*; so I put in §74. Even Gödel's completeness theorem for the predicate calculus (§72) has a proof-theoretic version after Hilbert and Bernays *1939* (IM, Theorem 36) and a relation to my arithmetical hierarchy (IM, Theorems 35 and 38).

As to the final Chapter XV, I first became aware of Gentzen's elegant *1934–35* paper in 1947, and used it to expound some results on consistency proofs in §79. And I concluded with §§80–82 on intuitionistic systems, the last giving an exposition of research initiated by me in 1941, and published by me in *1945* and David Nelson (in his Ph.D. thesis written under my direction) in *1947*. The first pages of Kleene (*1973*) give the history. This research established a connection between intuitionism and the subjects of general and partial recursive functions and Church's thesis, which were in the main line of the book (Chapters XI and XII). I remember Tarski telling me that he found this an interesting connection, which he had not thought of searching for himself. I was not under the spell of Brouwer, except to the extent that, when something can be done intuitionistically (not just classically), I feel it has more concrete content. Indeed, §82 emphasizes this. I had many interesting conversations with Brouwer from 1948 on, and felt he was sympathetic to my work. Many logicians not identified as intuitionists have paid attention to intuitionism (e.g., Kolmogorov (*1932*); Gödel *1932*, *1932–33*; Gentzen *1934–35*; Jaśkowski *1936*; Gentzen *1936*, p. 532 and Bernays (cf. IM, p. 495)). I had from the beginning made a point of identifying (at very little cost of space) which results hold for the intuitionistic versions of the formal systems being considered.

I have been asked why I did not include much model theory, or go further into set theory. Little of what is presented now in an introductory model theory course was known in 1952 when my book was published. As remarked in Chapter XIV, I did include some outstanding results. As to set theory, Gödel's work on the consistency of the continuum hypothesis first came out in *1938* (and Cohen's completion of an independence proof for it in *1963–64*), which was after the basic choices had been made as to the direction my course, seminars and book would take. Unlike the selections from model theory treated in Chapter XIV, this topic (involving details of the development of set theory formulated in logical symbolism, rather than just of logic and number theory) could not have been added without taking considerable space.

I intended to make the book essentially self contained, so that a reader with typical undergraduate-mathematics-major preparation could follow the story I was telling without having to supplement it by outside sources. At the same time I endeavored to be generous with references to the literature, both to give credit as due, and also (sometimes stating results) with the idea of encouraging the reader to enlarge his knowledge beyond what he could learn from the book.

I hoped all the topics I chose to work into the book would be of lasting significance. As I treated them, they made too bulky a manuscript to leave room for any other massive development.

Had I made the manuscript of the book any bigger, it might not have been published. *Proof*: The manuscript was accepted in principle on 4 April 1950 for publication by D. Van Nostrand Co. But on 31 August 1950, they wrote me backing out on the basis of the estimate by their printer that it would be

600 pages and would have to sell several thousand copies in the next few years to recoup the cost of printing. I replied that, by my calculation from a line count of my typescript, it would be 550 pages.[7] I then persuaded the North-Holland Publishing Company (with whom I had been in personal contact in Amsterdam in the spring of 1950 about printing the *Journal of Symbolic Logic*) to share with P. Noordhoff Ltd. the risk of printing my book, with Van Nostrand agreeing to bring out an American. edition from sheets printed in the Netherlands.

Notes

1. There has been no revision—only a few equal-space changes, two notes added in open spaces at ends of chapters, and the updating of eleven references which in 1952 were "to appear," as is indicated on p. vi of the sixth (1971) and later reprints.

2. On 31 December 1984 Moh ShawKwei, who translated IM into Chinese in 1965, wrote me "Now at last the Chinese translation of your work *Introduction to Metamathematics* comes off the press—only the first half of it yet. The other half, I hope, will be pressed in about a year."

3. Italicized dates, as here "*1931*," constitute references to the Bibliography of *Introduction to Metamathematics*, or (mainly for dates after 1952) to its Supplement at the end of the present article. IM was the first work, in any area of scholarship with which I am acquainted, to use dates (years) for the references.

4. This covers the case Turing's "integral variable" n ranges over the positive integers. If it ranges over the natural numbers, the number of 1's preceding the first 0 shall be the value of $\phi(0)$.

5. In *1943* and IM, §60, I used as the preassigned predicate $(y)\overline{T}_1(x, x, y)$ with the assumption of ω-consistency; and in *1936* $(y)(Ez)T_1(x, y, z)$ with a more complicated consistency assumption.

6. That D_0 and D_1 are recursively enumerable follows from the above fundamental requirements (1) and (2) using Church's thesis.

7. One may observe that, not counting the ten preliminary pages i–x with the preface and table of contents, it turned out to be exactly 550 pages, with the last page of the text, 516, filled to the last line.

Supplement to the Bibliography in IM

Cohen, P.J. (1963–4), The independence of the continuum hypothesis, and ibid., II. *Proc. Nat. Acad. Sci.* **50**, 1143–1148 and **51**, 105–110.

Davis, M. (1982), Why Gödel didn't have Church's thesis, *Inform. and Control*, **54**, 3–24.

Kleene, S.C. (1973), *Realizability*: A *retrospective survey*. Cambridge Summer School in Mathematical Logic, 1971, in A.R.D. Mathias and H. Rogers (eds.), Lecture Notes in Mathematics, No. 337, Springer-Verlag, Berlin, pp. 95–112.

Kleene, S.C. (1981), Origins of recursive function theory. *Ann. Hist. Comput.* **3**, 52–67. For six corrections, see Davis (1982), footnotes 10 and 12.

Kolmogorov, A.N. (1932), Zur Deutung der intuitionistischen Logik, *Math. Z.* **35**, 58–65.

Wang, H. (1974), *From Mathematics to Philosophy*, Routledge and Kegan Paul, London and Humanities Press, New York, xiv + 428 pp.