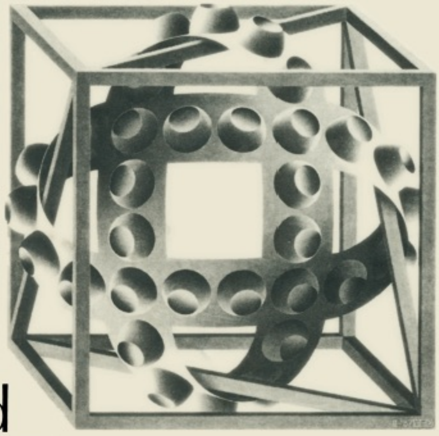


Kurt Gödel

# Unpublished Philosophical Essays

Edited by
**Francisco A. Rodríguez-Consuegra**

With a historico-philosophical
Introduction by the Editor

**Springer Basel AG**

By the same author also with Birkhäuser Publishing

**The Mathematical Philosophy of Bertrand Russell: Origins and Development (1991)**

# Kurt Gödel

# Unpublished Philosophical Essays

Edited by

**Francisco A. Rodríguez-Consuegra**

With a historico-philosophical Introduction
by the Editor

1995 Springer Basel AG

Author's Address: Universidad de Valencia
Departamento de Lógica y Filosofía de la Ciencia
Apartado 22.109
E-46071 Valencia/Spain

e-mail: FRANCISCO.RODRIGUEZ@UV.ES



Originally published in 1994 as "Kurt Gödel – Ensayos inéditos"
by Mondadori, Barcelona


A CIP catalogue record for this book is available from the Library
of Congress, Washington D.C., USA.

Deutsche Bibliothek Cataloging-in-Publication Data

**Gödel, Kurt:**
Kurt Gödel : unpublished philosophical essays ; with a historico-philosophical introduction / ed. by Francisco A. Rodríguez-Consuegra. – Basel ; Boston ; Berlin : Birkhäuser, 1995
[Sammlung]

ISBN 978-3-0348-9961-1 ISBN 978-3-0348-9248-3 (eBook)
DOI 10.1007/978-3-0348-9248-3
NE: Rodríguez-Consuegra, Francisco A. [Hrsg.]


Ursprünglich erschienen bei Birkhäuser Verlag 1995
Softcover reprint of the hardcover 1st edition 1995

Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Layout and design: Miriam Bussmann, Basel

9 8 7 6 5 4 3 2 1

# Contents

# Foreword

Gödel's famous incompleteness theorem shows that no formal proof procedure can reach every truth of mathematics, not even of the elementary theory of positive integers. His proof of this theorem, strictly mathematical of itself, wrought an abrupt turn in the philosophy of mathematics. We had supposed that truth, in mathematics, *consisted* in provability.

Gödel is celebrated in the philosophy of mathematics for this great discovery and also three other philosophically significant results, all strictly mathematical. Beyond these he published sundry brief notes in technical logic, many reviews, and some speculations on his friend Einstein's relativity physics, but of outright philosophy he published virtually nothing.

It was new for philosophers, therefore, when Gödel's *Nachlass* was found to contain manuscripts on the philosophy of mathematics, including even a substantial one on the philosophical bearing of his theorems. This is one of the two that Dr. Rodríguez-Consuegra has meticulously edited and presented in the present volume. At last we can glimpse Gödel's own philosophical adjustment to his bewildering discovery.

His philosophy of mathematics is at odds with the attitudes of most latter-day philosophers who deal with mathematics and logic. For him the abstract objects of mathematics are as real as sticks and stones, and their laws are objective matters of discovery on a par with those of physics. The other paper, written and repeatedly revised for a volume of commentaries on Carnap but never submitted, brings out these divergences very directly, for Carnap was a leading representative of the more dominant view.

Both manuscripts were tangles of revisions within revisions, labyrinthine transpositions, cryptic abbreviations, smudged erasures. These reflect Gödel's continuing sense of not having got the philosophy quite to his satisfaction.

Clearly the deciphering and linearizing of the manuscripts and the annotating of successive layers cost Dr. Rodríguez-Consuegra a lot of drudgery and demanded much scholarly ingenuity and a deep understanding of the subject. Rodríguez-Consuegra is a bright new light in the study of mathematical logic, set theory, and the philosophy of mathematics as these developed over the past twelve decades. His painstaking analysis *The Mathematical Philosophy of Bertrand Russell: Origins and Development* has already appeared as a book in English (Birkhäuser, 1991). With the present book he establishes yet another milestone in Spain's impressive latter-day progress in scientific philosophy.

W.V. Quine

# Acknowledgements

In the Spanish edition of these manuscripts (Barcelona: Mondadori, 1994) I already expressed my gratefulness to the following people, all of whom helped or encouraged me in connection with the project: Francine Abeles, José Alfredo Amor, Irving Anellis, Ken Blackwell, Irving Copi, John Corcoran, Mark Darby, John Dawson, Solomon Feferman, Eugenio Gallego, Alejandro Garciadiego, Ivor Grattan-Guinness, Nick Griffin, Jesús Hernández, Ramon Jansana, Faridah Kassim, Eckehart Köhler, Albert Lewis, Carlos Martín Vide, Greg Moore, Javier Moreno, Roman Murawski, Javier Ordóñez, Charles Parsons, Klaus Puhl, Elliot Shore, Friedrich Stadler, Carlos Solís, Roberto Torretti, Luís Vega, and Zbigniew W. Wolkowski.

Also, special thanks were due to Ignacio Jané, for reading and criticizing my introductory essay; José Manuel Sánchez Ron, who first gave me the opportunity to write a book like this; Hao Wang, for having read and approved my reconstruction of the Gibbs lecture, and for generously having sent me a copy of his *Conversations with Gödel*, still unpublished; and Willard V. Quine, for his illuminating remarks in conversation and correspondence, as well as for having written the Foreword. My final thanks were due to Kurt Gödel himself, for having written such fascinating essays.

Regarding the present edition I am grateful to Juan Climent and Enrique Trillas, who pointed out to me some errata in the previous edition and suggested improvements. Also, to George Boolos, Mic Detlefsen, and Ivor Grattan-Guinness, all of whom read parts – or the whole – of the introductory essay and sent me remarks useful for improving it. Also, I would like to express my thanks to Ms. Baer, Birkhäuser's language editor, for improving my English, and very especially to Doris Wörner, who approved the project of publishing this book and was very patient with my delay in the delivery of the manuscript. Finally, I am grateful to the help provided by the Spanish grant DGICYT PS93-0220.

*Departamento de Lógica y Filosofía de la Ciencia*
*Universidad de Valencia*

# Introduction

Kurt Gödel, together with Bertrand Russell, is the most important name in logic, foundations and philosophy of mathematics of this century. However, although Russell devoted many pages to articulating his ideas in these and many other fields, Gödel published very little apart from his well-known writings in logic, metamathematics and set theory. His introverted personality inclined him away from philosophical controversy, while his continuous search for definitively conclusive arguments in philosophy made it difficult for him to publish his less conclusive writings. Thus, the sum of his published philosophical remarks runs to what is found in the papers on Russell and Cantor and a few scattered remarks elsewhere.

Fortunately, Gödel the philosopher, who devoted many more years of his life to philosophy than to technical investigation, wrote hundreds of pages on philosophy of mathematics, as well as on other fields of philosophy. However, it seems that not even his closest colleagues and friends were allowed to read those manuscripts, let alone to discuss them. Only the opening of his literary estate in Princeton for scholarly research, following the catalogization made from 1982 to 1984, have made it possible that all these materials may someday appear in print.

The fundamental goal of this book is to make available to the scholarly public solid reconstructions and editions of three of the most important essays which Gödel wrote on the philosophy of mathematics. Since in chapter 1/part II give details about the character and origin of the manuscripts appearing here, as well as the particulars about their present edition, I shall devote the rest of this introduction to a summary of the book as a whole.

I thought that many – perhaps most – of the readers would be grateful if the Gödel essays were accompanied by an introductory apparatus, devoted not only to the manuscripts themselves but also to the philosophical context in which they were written. This led me to divide the book into two parts, the first providing the reader with that context, the second offering certain pertinent information regarding the background of the particular manuscripts which appear here, as well as the present edition.

The first part is composed of three chapters, of which the first tries to achieve two goals. The first of these is to describe briefly the intuitive kernel of and the philosophical motivations behind Gödel's celebrated metamathematical results. Gödel took these as supports for a realistic (Platonic) philosophy of mathematics – the belief in the objectivity of mathematical entities – and we are therefore concerned to determine the quality of that support. To this end, I have successively examined the theses according to which Gödel's realism should be regarded as (i) part of the philosophical implications of those results; (ii) a heuristic principle leading to them; (iii) a philosophical "hypothesis" which can be "verified" by means of them. The second goal has been to provide the reader with a sort of overall view of the content of the manuscripts appearing here, as well as with a general context of ideas and authors. Here, it seemed to me that Gödel's last version of his essay on Carnap is very useful as a sort of summary of the basic ideas of the manuscripts.

The second chapter is devoted to one of Gödel's main theses in the philosophy of logic and mathematics, namely that these sciences are of an analytic – although not tautological – character. Thus, I begin by presenting the essential arguments given by other philosophers who supported this thesis before Gödel; authors such as Frege, Russell, Wittgenstein, Carnap and Quine. Their ideas provide the context needed to explain Gödel's arguments, published and unpublished, and to evaluate some of the advantages and disadvantages of his overall view.

The third chapter provides the reader with the context and basic arguments for another main philosophical thesis, namely that mathematics is essentially similar to physics, both in its objectivity and method. Here, again, similar ideas by several authors well known to Gödel are examined. The list includes such thinkers as Russell, Hilbert, Carnap, Tarski and Quine. Curiously enough, the philosophical ideas of the last three authors can be correctly understood only when seen in light of the impact of Gödel's results. These results, together with Gödel's particular arguments (both published and unpublished), are also briefly touched upon in the chapter.

As regards mathematical intuition, since most arguments used to defend its existence are of the same sort as those already used by Gödel in relation to the former theses, no separate chapter has been devoted to it. Concerning analyticity, mathematical intuition is, for Gödel, the faculty providing us with the guarantee of the truth of mathematical axioms, its fallibility notwithstanding. Regarding the analogy between mathematics and physics, Gödel viewed mathematical intuition as a faculty parallel to sense perception. Indeed, there seems to be no other way to develop Gödel's conception of mathematical intuition.

These three chapters can be seen as an outline of those aspects which I consider to be the most interesting ones of Gödel's philosophical writings (published and unpublished) in the field of the philosophy of logic and mathematics. Still, they do not address all of his philosophical ideas – for example, those in his Gibbs lecture concerning mechanistic theories of mind and those in some of his writings concerning the nature of time – for reasons of space.

The bibliography appears at the close of the third chapter as it seemed a fitting accompaniment to the introductory essay in the first part of the book. However, it contains items pertaining to the fourth chapter as well and also some non-standard references to such things as a symphonic composition and a videotape both devoted to Gödel.

The second part of the book, comprising chapters 1–4, is devoted entirely to Gödel's unpublished essays. Its first chapter, the fourth of the book, should adequately equip the reader to fit the essays into the larger framework of Gödel's unpublished work. It includes basic information concerning both Gödel's *Nachlass* as well as the origins of the essays published here. I also briefly describe the conditions in which I found the pieces published, and state the criteria which guided me in my choice of what to include here, as well as in my treatment of it.

The second chapter contains the text of the Gibbs lecture, which Gödel, invited by the American Mathematical Society, read in Providence, Rhode Island, in 1951. Officially entitled "Some basic theorems on the foundations of mathematics and their philosophical implications", it was devoted to non-technical description and philosophical discussion of Gödel's famous incompleteness results. Among the arguments presented, the most important are those asserting the impossibility of developing a reductionist program in the philosophy of mathematics and the necessary failure of any mechanist conception of human reasoning. Here I should perhaps call the reader's attention to the fact that the long and interesting footnotes included in the manuscript of this lecture are given in their entirety, but only following the main text, so as to make the reading easier and more fluid.

Between 1953 and 1959 Gödel wrote up to six versions of an essay on Rudolf Carnap's syntactical conception of mathematics, originally intended to appear in the Carnap volume of the Schilpp series (*The Library of Living Philosophers*) under the title "Is mathematics syntax of language?", although none of them were actually submitted for publication. The third chapter (part II) contains the text of the second version of that essay.

Gödel attacked the thesis that mathematics can be reduced to formal syntax of language, thereby being shown to be tautological and thus void of content. It is Gödel's only openly polemical essay. However, the positive part of the essay is also remarkable, with implications for a variety of related issues in and connected to the foundations of mathematics. The essay contains many references to the relevant literature and a number of footnotes. These latter are printed at the end of the main text.

The fourth and final chapter (part II) contains version VI of the same essay. It is a short text and certainly not very convincing. Still, it seemed worth printing because, although version V is clearer and more philosophical, version VI represents the culmination of a long evolution of thought. Moreover, nothing is essentially lost in so doing, for a comparison between both versions is offered in a systematic series of editorial footnotes.

Finally, there are the reproductions of various pages of the original manuscripts printed here. They are included to give the reader an idea of Gödel's actual style of working. It is a style borne of insecurity and perfectionism – a fact that makes the task of reconstructing and editing his manuscripts a difficult one. I apologize for the poor quality of the reproductions. They had to be made from photocopies since the original documents cannot be taken outside the Princeton Firestone Library.

# Part I

# Kurt Gödel and the philosophy of mathematics

Quand je vous aimerai?
Ma foi, je ne sais pas.
Peut-être jamais, peut-être demain!
Carmen

# 1

# Realism, metamathematics, and the unpublished essays

This initial chapter is divided into two sections. The first is devoted to a brief exposition of the intuitive essence and the philosophical motivation of Gödel's main metamathematical results, namely his completeness theorem for elementary logic (1930) and his incompleteness theorems for arithmetic (1931). Thereafter some discussion of the different ways to confront the relationship between those results and Gödel's philosophical realism in logic and mathematics is offered. Thus, mathematical realism will be successively regarded as (i) a philosophical consequence of those results; (ii) a heuristic principle which leads to them; (iii) a philosophical hypothesis which is "verified" by them. In the second section Gödel's philosophy of mathematics, such as it can be derived from his published writings, is briefly expounded upon. Then the final version of his essay on Carnap is summed up, in order to see how his unpublished philosophical ideas might throw some light on Gödel's published doctrines. Finally, other relevant ideas and authors are briefly surveyed.

## Gödel's results

Gödel's celebrated results referred to here have to do with basic properties of certain logical and mathematical formal systems. When those systems are studied in themselves, their properties are called metalinguistic or metatheoretical, namely metalogical or metamathematical. By "formal system" we mean a set of symbols and the corresponding rules of formation of acceptable sequences of symbols, or formulas. These formulas can be divided into two classes: axioms, which are taken as the starting point, and theorems, which are derivable (demonstrable, provable) from the axioms, that is to say, which can be obtained from the axioms by means of the application of certain rules of inference explicitly formulated. Thus, provability will always be understood as referring to a precise set of axioms and rules of inference. The main interest of formal systems lies in their capacity to be used to formalize certain languages. In this way the structure of these languages turns out to be perfectly specified, so it is easier to avoid certain problems that, like the one concerning the celebrated paradoxes, led to several crises in the foundations of mathematics at the turn of this century. In particular, the several attempts to formalize logical and mathematical theories, such as elementary logic and number theory, were the field in which Gödel's main results first appeared.

Among the metatheoretical properties of formal systems are consistency, completeness and decidability. A system is *consistent* when it is not contradictory, i.e., when it is not the case that both a formula and its negation are provable in the system. Thus, if we obtain a consistency proof, we will have the guarantee that we will never come to the point where we derive contradictory theorems from the axioms. A system is *complete* when every one of its formulas, or its negation, is provable. Therefore, to prove the incompleteness of the system it suffices to exhibit – or prove the existence of – a well-formed formula of the system, such that neither it nor its negation is derivable in it. Finally, a system is *decidable* when there is an algorithmic (mechanical) procedure through which we can determine, in a finite number of steps, whether or not each well-formed formula is provable in it, i.e., whether or not each formula is one of its theorems.

Understanding the relationships between such properties helps to clarify their importance. First of all, it has to be noted that completeness and decidability are by no means equivalent properties: a decidable system is not necessarily complete, and a system can be both complete and undecidable. At the same time, it is worth investigating the completeness and decidability of a system only under the hypothesis of its consistency, for a contradictory system allows us to prove any theorem, so although it is always complete and decidable, it is so in a trivial way. In principle, the ideal situation for every formal system would be to satisfy all these properties. Hence Hilbert's metamathematical program consisted in reaching a formal system for classical mathematics (intuitionist mathematics is more difficult, for it rejects the application of concepts which cannot be determined in a finite way) which was at once consistent, complete and decidable. As we shall see, Gödel's results showed the impossibility of such an ideal.

After these preliminary notions, we can already say that two of Gödel's results are relevant here; first, the demonstration that the formal system for elementary logic (the logic in which our quantifiers range just over individuals, not over sets of these or their properties) is complete (although now in the sense that every formula universally valid is provable in it). This is what it is known as "Gödel's completeness theorem", which improved on former ones by Post and Bernays. According to it sentential logic (the logic with no quantifiers) is complete. This theorem was formulated in 1930; shortly thereafter, however, Church proved that elementary logic is not decidable (1936).

Second, Gödel proved, in 1931, that every formal system – like Russell's or Hilbert's – strong enough to try to embrace – to prove from its axioms and rules – every arithmetical truth is necessarily incomplete (see below for a more accurate presentation). It will always be possible to construct a sentence which, although true, is undecidable in the system, namely neither it nor its negation is a theorem. This is known as "Gödel's incompleteness theorem", or simply as "Gödel's theorem". At the same time, Gödel proved that it is impossible to find a proof for the consistency of such systems, because a sentence asserting that consistency would be one of those undecidable sentences. The final result was that the search for an "ideal" (see above) formalism was forever abandoned. Thus, formal systems of this type *cannot* be neither complete nor decidable, and, although they can be consistent, their consistency *cannot* be proved.

In the rest of this section I will make more detailed comments about both results, but mainly with the aim of emphasizing their philosophical interest. For more detailed expositions, the reader can turn to Hao Wang 1987 (chapter 10), Nagel/Newman 1958 or Hofstadter 1979. Also, a short, clear and thorough exposition of the incompleteness results appears in Detlefsen 1987 (see especially pp. 94-97). As for the specialist, I hope that she may be patient with the informal way in which I am going to point out some of the philosophical import of Gödel's technical work.

The completeness theorem for elementary logic, which was already established by Gödel in his doctoral dissertation (1929), asserts that every valid formula of that logic is a theorem, i.e., it is provable from the axioms and rules of the system. Gödel based that result on two techniques. The first came from a former result by Skolem according to which every formula of elementary logic is equivalent to another exhibiting a "normal form", which is characterized by the place of the quantifiers, located at the beginning. With that, the task is limited to proving that every formula logically valid of this class is demonstrable. The second shows that all these formulas can be put into correspondence with valid formulas of the sentential logic, from which they can be derived.

The basic idea of the proof is that for every formula X of elementary logic it is possible to construct a formula Y of sentential logic, such that (i) if X is a formal truth, then so is Y, and (ii) X is provable from Y in the system. Now, if Y is a formal truth, then it is provable – it is a theorem. Thus, X is a theorem of the system, for X is provable from Y in it. (I follow Sacristán 1964, pp. 183-4.) The conclusion is that the formal system of elementary logic is complete. (This does not mean that it is also decidable: Church proved in 1936 that there is no decision procedure for all its formulas.) Hence a formula of elementary logic is logically true if and only if it is logically demonstrable. In other words that means that in languages of this kind logical truth is equivalent to provability.

Several things have to be immediately noted. For one thing, it is not the case that in every formal system every true formula is provable, as Gödel himself proved with his incompleteness theorem: there are always undecidable formulas in sufficiently strong systems. On the other hand, it is not the case either that every provable sentence in any system is true. It is obvious that it would depend on the truth of the sentences which are taken as axioms. Therefore, a general equivalence between truth and provability cannot be sustained. To explore those limitations, together with their philosophical implications in connection with the concept of consistency, I will refer now to the context of Gödel's philosophical ideas of that time.

Gödel's introduction to his 1930 doctoral dissertation is very interesting, but was generally known only after its recent publication in his *Collected Works*, vol. I (henceforth referred to as CWI). Its philosophical character is likely to have been the reason why he did not publish it as it stood. Assuming this to be true, we would thus have the first sign of his extreme reluctance to make his philosophical realism public. A reason for this reluctance could be that his form of realism was interpreted as being openly opposed to Hilbert's formalism, according to which the proof of the consistency of a certain system would be sufficient to guarantee the existence of the corresponding mathematical concept, as well as the truth of its axioms. (See, however, the section on Hilbert in chapter 3 of this part.)

Gödel's main argument in that introduction is precisely directed against the belief that the consistency of a system of axioms implies the existence of the corresponding mathematical concept. That belief, says Gödel, presupposes that every mathematical problem is solvable, namely, that there cannot be undecidable formulas. However, Gödel suggests the possibility of proving the existence of unsolvable mathematical problems, so he seems to be foreshadowing the incompleteness result of the following year, according to which every formal system with certain properties necessarily contains undecidable formulas.

This argument can be used to reveal the extent to which we can defend the equivalence between truth and provability, or even better, between the consistency of a system and the existence of the corresponding concept. To do that, Gödel applied the completeness theorem to axiomatics at the end of his doctoral dissertation by extending his semantical methods (which are based on the construction of models satisfying – making true – certain formulas) to axiom systems. The result reached there is that every first-order axiom system has a model or is inconsistent, i.e., every consistent set of axioms has a model (CWI, p. 101). Thus, the consistency of the system implies the existence of a model. Hence the completeness theorem showed, too, that consistency is somehow equivalent to existence. Therefore, Gödel's reluctance to admit the general equivalence between consistency and existence could perhaps be interpreted as simply meaning that we cannot assume the equivalence before obtaining that theorem, and then only in the precise terms in which it appears as a consequence of the theorem itself (see Feferman 1984, pp. 551–2).

The following step towards achieving the incompleteness theorem can therefore be seen as pointing out the limit of the existential import of consistency, a step which took place in the Königsberg symposium of 1930. As Gödel said there, that consistency is not enough can be seen simply by considering a true but undemonstrable sentence in classical mathematics; if we add the negation of such a sentence to the axioms of classical mathematics we obtain a consistent system in which false sentences can be proved (CWI, p. 203). In this way the first incompleteness theorem comes to light; some months later Gödel stated the existence of true but undecidable sentences in every sufficiently strong formal system.

Thus, we can already make the conjecture that probably a great part of Gödel's early interest in marking off the bounds of completeness (which embraces elementary logic, but not arithmetic) lay in an attempt to provide his mathematical realism with a strong support. With such a support, the opposition to Hilbert's formalism – and, later, to Carnap's syntactical view of mathematics – could be better articulated. For those conceptions of mathematics it was not necessary to accept the objective reality of mathematical concepts. Thus, that sort of reality was for them somehow reducible to the recourses available in the mere formal system; hence there was no need to postulate the existence of a mathematical intuition in order to accede to those objects, and a strong link between consistency and existence was essential, at least for Hilbert.

That is why Gödel insisted so explicitly on the limitations of that link: if we can prove that in a consistent mathematical formal system it is possible to derive a false theorem, then the consistency proof is not enough to guarantee the existence of the corresponding concept, which cannot be "contradictory". Now, if consistency is no guarantee of truth, let alone of existence, then we cannot create "objective" concepts simply by means of the construction of formal systems; on the contrary, those systems should represent concepts existing by themselves. For similar reasons, it can be conjectured that even by that time Gödel believed that every attempt to interpret mathematics in a formalist way is a failure. Such an attempt forgets the unavoidable role of mathematical intuition, which leads us in the construction of formal systems, in order to give an account of those objective concepts. Thus, for Gödel mathematical intuition cannot be replaced by any proof of consistency, or by any other "reductive" recourse: there is an ultimate content of mathematics which cannot be eliminated by means of any formal system. (The essays appearing here are especially useful to see what could be for Gödel that ultimate content.)

We thus come to the incompleteness theorem, about which we can be briefer, once we have established its philosophical interest for Gödel. Its kernel is that every formal system strong enough to try to formalize classical mathematics necessarily contains true but undecidable sentences. (As a matter of fact the theorem holds for weaker systems, for all it needs to apply is that the recursively enumerable set of natural numbers be representable in the system in a certain way, but its great impact took place on the stronger systems which were being used to try to completely formalize mathematics.) Such systems are necessarily incomplete, for if they are consistent they cannot contain a proof of all sentences of their language that are true. Also, they are incompletable, for no number of sentences which we may add to the system could make it complete: the resulting system would contain, once again, true but undecidable sentences. Moreover, Gödel proved that a sentence expressing the consistency of one of such system is not provable in the system (if the system is consistent); that sentence will be precisely an undecidable sentence.

So fascinating a theorem was reached by means of a novel method of proof which constitutes one of Gödel's greatest ideas, and is now called "arithmetization of syntax". Broadly speaking, the method consists in the correlation of syntactical objects with numbers, and properties of syntactical objects with properties of numbers, within a given formal system. This can be further explained this way: (i) the assignment of natural numbers to every expression of the language used (that is, to every symbol, to every finite string of symbols, and so on); (ii) the selection of number-theoretic properties (e.g., to be a prime number) to represent metamathematical properties; and (iii) a choice of formal expressions (expressions of the formal system used) for these numbers and properties of numbers. In this way the formulas are represented by numbers and the properties of formulas and string of formulas are represented by properties of numbers, so those numbers can be used to make assertions about the formulas. By using this device, an expression in a formal system for classical mathematics can play, at the same time, the role of an arithmetic formula and that of a mathematical sentence attributing certain metamathematical property to that formula. Yet this can be done without the need for any language of a higher level – a metalanguage, for both are formulated in the same language.

Gödel systematically and accurately applied such an idea, in conjunction with others which cannot be explained here, to construct a special formula. This formula, which we can call G, is both an arithmetic formula and an expression of the metamathematical proposition asserting that G is not provable; hence G represents precisely this mathematical proposition. However, the task was done without succumbing to anything similar to the paradox of the Liar, who said the truth when lying and lied when saying the truth. In our case, although G is true and non-provable, it does not say of itself that it is true, but only that it is not provable. (If G would say of itself that it is true, it would be in need of a higher language in order to avoid the paradox.) Thus, although G refers to itself, this is done through a perfectly acceptable type of self-reference. The following informal reasoning tries to give at least an idea of the way things happen in Gödel's first incompleteness theorem. It shows that G is true, while neither G nor its negation are provable in the system, so G is undecidable.

Let us assume a system which is consistent and such that every sentence provable in it is true. Now, if we suppose that G is provable, then G is true, so that what G asserts is true, and what it asserts is precisely that G is not provable. On the other hand, if G is provable, then G is false, for G asserts that "G is not provable". We can therefore assert at the same time that if G is provable then G is true, and that if G is provable, then G is false. But this is a contradiction, so G is not provable; hence G is true, for what G asserts is precisely that G is not provable. The conclusion is that if the corresponding formal system is consistent (if it is not, everything is provable in it), then neither G nor its negation is provable in it, so there is at least an undecidable formula in the system, and thus it is incomplete (Gödel's first theorem). Also, it can be proved that a sentence expressing the consistency of the system is undecidable as well (Gödel's second theorem).

Of course, it may always be objected that G is provable in another formal system, but this amounts to very little, for that formal system would necessarily contain undecidable formulas as well. For similar reasons, although it can be said that the consistency of our original formal system can be proved in another, stronger system, the problem of the consistency of the new system would immediately arise. Thus Gödel's results have to be interpreted as essential – and therefore impossible to overcome – limitations of certain formal systems. But these systems represent the ones which were believed to be able to give an account of classical mathematics, so the hope in that kind of formalization had to be definitively abandoned. (However, see Detlefsen 1987 for an interesting challenge to the use of Gödel's theorems to overthrow Hilbert's program. Chapters 4 and 5 of that book give reasons for thinking that Gödel's second theorem does not refute Hilbert's program, and the appendix to the book criticizes an argument claiming that Gödel's first theorem refutes Hilbert's program. Detlefsen 1990 extends the argument of the appendix.)

## The "philosophical implications"

The list of philosophical implications of the incompleteness results which have been pointed out from different viewpoints is impressive: objective mathematical truth is opposed to mere provability (as opposed to formalism and logical positivism); it is impossible to build up a unique formal system to give an account of mathematics (as opposed to logicism and formalism); if it is maintained that mathematics is analytic, then the sentence G should be synthetic, for although we know that it is true, it is not provable in the system (as opposed to Gödel's view of mathematics as being analytic); it is not possible to reduce mathematics to any algorithm, then the human mind should somehow surpass any formalism (as opposed to certain forms of materialism); etc.

However, Gödel, both in his published and unpublished writings, insisted mainly on the realistic implications, according to which the objective character of mathematical concepts (number, set, etc.) goes beyond any attempt of complete formalization. Nevertheless, if we join Gödel's incompleteness results together with one of the outcomes of the Löwenheim-Skolem theorem (according to which every characterization of natural numbers has non-standard models), then we can also say that some definitely non-realistic implications may be drawn from Gödel's discoveries.

As Dummett has pointed out (1963, 1978b), we cannot say that Gödel's theorem proved that, although we possess the concept of natural number, no finite description of this concept can give a complete account of its structure. For Dummett this would be to apply the notion of model in an incomprehensible way, for we can be given a model only by means of a description of it; hence "if we cannot be given a complete characterization of a model for number theory, then there is not any other way in which, [...], we could nevertheless somehow gain a complete conception of its structure" (Dummett 1978b, p. 191). Thus, the usual conclusion that there is a standard model for natural numbers in our mind, in spite of our inability to completely characterize it formally, is simply erroneous. To say that we cannot communicate unequivocally our intuition of natural numbers through a formal system could be acceptable only if we had available another way to communicate it; but this is simply not the case, so the Platonic notion of model seems to be in considerably difficulty.

We tend, says Dummett, to interpret the incompleteness of a formal system for arithmetic in terms of non-standard models. This simply means, for the Platonist, that the formal system fails to capture our intuition of the structure of the concept of number, for this structure makes other non-standard interpretations possible. Yet every formal system makes those interpretations possible, so we can never be sure that what someone refers to as the standard model is in fact isomorphic with the standard model we have in mind. Thus, the supposed objective content of mathematical intuition does not seem sure to the grasp.

Gödel never published anything on non-standard models in arithmetic, but in the record of his conversations with Hao Wang in the 1970s some relevant fragments can be found. These are given below (I retain Wang's original notations to locate the fragments, as well as his remarks, in square brackets):

> G says that there should be a more fruitful manner of developing non-standard analysis. According to him, it is wrong interpretation of Skolem's theorem to say that it makes the characterization of integers by logic impossible, because one can use the theory of concept. [I believe Skolem's theorem here refers to his construction of nonstandard models of arithmetic] [112.5]; The axioms correspond to the concepts and the models which satisfy them correspond to the objects; the representations give the relation between concepts and objects [21.4]; Numbers appear less concrete than sets: they have different representations [by sets] and are what is common to all representations [21.5]; I say in the draft that for set theory 'we do not even have an equally fixed intended model as with natural numbers.' G says that in intention the model is equally fixed, only we do not perceive it as clearly [21.28]. 'The iterative concept is admittedly vague and undetermined.' G objects to the formulation, saying that we just don't see it with precision [25.3]; Intuition is not a proof; it is the opposite of proof. We do not analyze intuition to see a proof but by intuition we see something without proof. We only describe in what we see those components which cannot be analyzed any further. We do not distinguish between intuition *de re* and *de dicto*, the one is contained in the other. To understand a proposition we must have an intuition of the objects referred to. If we leave out the formulation in words, something general comes in anyhow. We can't separate them completely. [511.6]

It suffices to say here that such fragments appear as Gödel's defence against certain philosophical implications of his celebrated results, or even a kind of reversion to former well-known views (such as

Peano's, for whom numbers were merely that which all possible interpretations of formal arithmetic have in common). At any rate the supposed existence of an intuition which allows us the direct access to mathematical objects seems to be the main basis for Gödel's philosophical conception at this point. As we shall see in the following sections, the same basis comes to light in the rest of the apparently conflicting views we are going to examine.

## The heuristic principle

In letters to Hao Wang (1974), Gödel provided us with a second philosophical interpretation of the relationship between his metamathematical results and his realist conception of mathematics. According to this interpretation, that philosophical conception, used as a sort of "heuristic principle", made it possible for him to obtain those results. In particular, Gödel tries to show that there is a link between transfinite concepts – such as the concept of objective mathematical truth – and his results, and his summed up that link by speaking of a particular epistemological attitude to metamathematics and the sort of reasoning having to do with the infinite (non-finitary reasoning), which, for Hilbert, was reducible to the finite.

In respect to the completeness result Gödel wrote (Hao Wang 1974):

> ... the completeness theorem, mathematically, is indeed an almost trivial consequence of Skolem 1922. However, the fact is that, at that time, nobody (including Skolem himself) drew this conclusion (neither from Skolem 1922 nor, as I did, from similar considerations of his own).... The blindness (or prejudice, or whatever you may call it) of logicians is indeed surprising. But I think the explanation is not hard to find. It lies in a widespread lack, at that time, of the required epistemological attitude toward metamathematics and toward non-finitary reasoning. Non-finitary reasoning in mathematics was widely considered to be meaningful only to the extent to which it can be 'interpreted' or 'justified' in terms of a finitary metamathematics. (Note that this, for the most part, has turned out to be impossible in consequence of my results and subsequent work.)

As for Skolem's results and its relationship to Gödel's completeness theorem, this is a very complex technical and historical problem which we cannot deal with here (see the introduction by van Heijenoort and Dreben in CWI, pp. 51 ff.). However, I think Gödel's reference in this passage, concerning non-finitary reasoning in connection to the completeness results, is not difficult to clarify. Hao Wang wrote that this remark applies not only to non-finitary reasoning, but also to the formation of certain concepts, "and the very concept of completeness involves the nonfinitary component of being true in 'arbitrary domains'" (Wang 1987, p. 269). The fact that Gödel believed that his theorem supposed "a theoretical completion of the usual method of proving consistency" (CW1, p. 61), is derived from a well-known application of the theorem. According to this application, every consistent first order axiom system has a model. As we have pointed out above, for Gödel we can only accept the equivalence between consistency and existence if we presuppose that every mathematical problem is solvable. That is to say, the completeness theorem proves only the existence of a model, but does not provide us with any means to discover that model. On the contrary, if every problem is solvable, then the particular problem of specifying every model in an effective way would also be solvable (Wang 1987, p. 272).

Yet I think it is possible to defend another interpretation of the same remark. According to this interpretation, Gödel's remark would actually refer to Hilbert's presupposition of the solvability of every problem, but in a different sense. By following the line of thought I have already indicated in the first section of this chapter, what Hilbert was really presupposing is that consistency implies truth. As we have seen, Gödel pointed out in the Königsberg symposium that once he indicated the existence of true but non- provable sentences, if one added an undecidable false sentence to a consistent axioms system, then one could obtain from it false theorems (CWI, p. 203). Thus, consistency would not be enough as a guarantee for truth, unless we presupposed that every problem is solvable, namely that there is no undecidable sentence. With that we could think that Gödel's "heuristic principle" might really have depended on a former result, which, although not yet formally reached, may well have been seriously conjectured: the incompleteness of arithmetic, which was later proved through the existence of true but undecidable sentences.

As for Gödel's incompleteness theorem, Gödel offered, in the same letters to Hao Wang, a similar explanation in terms of the correct epistemological attitude to mathematics: "How indeed could one think of *expressing* metamathematics *in* the mathematical systems themselves, if the latter are considered to consist of meaningless symbols which acquire some substitute of meaning only *through* metamathematics?" With that he seems to be referring to the arithmetization of syntax (see above), which provided Gödel with the possibility to express metamathematical properties through arithmetic sentences. Also, he could be referring to the supposed guarantee that a proof of consistency would bring to Hilbert's "ideal" propositions (see the section on Hilbert in chapter 3 of this part). However, to construct undecidable sentences Gödel preferred to resort to what he calls objective mathematical truth, regarded as a heuristic principle. According to him, objective mathematical truth is a highly transfinite concept which used to be confused with provability: "Again the use of this transfinite concept eventually leads to finitarily provable results, e.g., the general theorems about the existence of undecidable propositions in consistent formal systems".

Now, let us see how this is compatible with the "heuristic principle" interpretation. First of all, the existence of an objective transfinite truth, as something different from provability, is a consequence of the incompleteness theorem. Yet Gödel is now presenting objective truth as the idea which *led* him to such famous a result. On the other hand, we now tend to say that the ultimate difference between truth and provability is a part of the kernel of what we now know as "Tarski's theorem" (e.g., the indefinability of arithmetic truth), and the essentials of this theorem were well known to Gödel at least as early as 1930 (see next section in this chapter). Thus, the interpretation in terms of something so vague as heuristic principles seems to be in trouble. In fact, it seems that Gödel's progress depended upon precise results, rather than upon ideas (or "attitudes") which, in an informal and almost psychological way, merely suggested to him other ideas.

We said before that Gödel's rejection to any equivalence between consistency and existence seemed to depend rather on a former, although still not fully reached, result (incompleteness) than on a mere heuristic principle. We can add now, although still provisionally, that the very incompleteness result might have depended more on another former result (indefinability of arithmetic truth, e.g., Tarski's theorem) than on a new heuristic principle. In the next section we shall face this problem directly, while we explore the third and final philosophical interpretation under consideration: realism as an hypothesis to be verified by means of its consequences.

**The philosophical hypothesis**

The Gödel thesis in the letters to Hao Wang was that the formalists were incapable of distinguishing between mathematical truth and provability, for they analyzed the first in terms of the latter. Once the incompleteness result was available, it could be shown that the analysis was faulty. This particular fact exemplifies of a more general pattern: when it can be proved mathematically that a philosophical thesis is in error, then an alternative thesis must be true. This new interpretation of the relationships between Gödel's philosophical viewpoint (mathematical realism) and his metamathematical results opens the way to our third approach: realism, regarded as a philosophical hypothesis, can be mathematically verified by means of its consequences.

It seems that Gödel, in discovering on his own Tarski's theorem (the indefinability of arithmetic truth, through the difference between truth and provability), was immediately aware of the existence of undecidable sentences in certain strong formal systems. Then he began the search for a proof of this fact which was acceptable for the intuitionists, who required "effective", finitary methods. Thus, such a proof should provide us with some effective instance of an undecidable sentence, while it should not require the concept of "every true formula", which is clearly infinitary. There is a line of thought in section 7 of Gödel's Princeton lectures of 1934 (first published in 1965, now in CWI) which is relevant for that proof. There we can find

Gödel's first reference in print to Tarski's theorem in connection with the necessary existence of undecidable sentences (CWI, p. 363):

> So we see that the class $\alpha$ of numbers of true formulas cannot be expressed by a propositional function of our system, whereas the class $\beta$ of provable formulas can. Hence $\alpha \neq \beta$ and if we assume $\beta \subseteq \alpha$ (i.e., every provable formula is true) we have $\beta \subset \alpha$, i.e., there is a proposition A which is true but not provable. $\sim A$ then is not true and therefore not provable either, i.e., $A$ is undecidable.

This reasoning had already been expressed in a the letter to Zermelo of 1931 about his incompleteness theorem (see Grattan-Guinness 1979). It is not clear, however, that it can be seen as an example of the application of a genuine "heuristic principle". Rather, it seems to me simply a rigourous argument which *proves* the necessary existence of undecidable sentences. In fact, Tarski's theorem and Gödel's incompleteness theorems are completely rigourous mathematical results, and it seems obvious that Gödel arrived at the latter from the former. Therefore, when he wrote that the Princeton argument of 1934 constitutes a "heuristic argument" for the existence of undecidable sentences (CWI, p. 363) it seems we should interpret him as saying that (i) earlier he was philosophically convinced that objective mathematical truth does not coincide with provability; (ii) the difference between mathematical truth and provability resulted confirmed by his original discovery of Tarski's theorem, which led him immediately to state the distinction itself in a rigourous, mathematical way; (iii) all this led him to the incompleteness result.

At any rate Gödel did not develop the philosophical hypothesis view in his publications, so it is not possible to resort to them at this point. Fortunately, in his later conversations with Hao Wang of the 1970s certain fragments appear where it seems that Gödel interpreted that view by speaking of the fruitfulness of verifiable consequences of his realist viewpoint (again, I retain Wang's original notations to locate the fragments, as well as his remarks, in square brackets):

> ... the Platonic view helps in understanding things: this fact illustrates the possibilities of verifying philosophical theory [110.2]; The hypothesis stating that Platonic ideas give shape to the universe is the most natural and, philosophically, the most economical [110.28]; It is an assumption even made by the positivists that if a hypothesis [objectivism in this case] leads to verifiable consequences (which could be reached in another way) or to theorems provable without this hypothesis, such a state of affairs makes the truth of the hypothesis likely. However, mathematicians [probably represented for G by Abraham Robinson] like to take the opposite position: It is correct to take objectivism (or Platonism) to be fruitful but it need not be true. This position is opposite to the nature of truth or even science and positivists [61.26]; People might choose not to adopt the objectivistic position but merely do their work 'as if' the position were true, if they are able to produce such an attitude. But then they only take this 'as if' point of view toward this position after it has been shown to be fruitful [21.8].

With that we can see that Gödel was influenced by a verificationist scheme, in the old-fashioned style, which can be regarded as predating even Popper's falsationism, perhaps in the way of certain rather popular positivistic presentations, in the sense at least of resorting to their language.

In view of this, the only distinction which occurs to me between the fruitfulness conception and the one based on the heuristic principle is that the first is more similar to the usual method of science (the classical verification of hypotheses), while the second involves the very psychology of the researcher, who is led by his personality, prejudices, etc., to certain ideas rather than to others. But then the big difficulty lies in the fact that Gödel does not even refer to the need for making actual predictions according to the hypothesis under discussion, which constitutes the standard hypothetico-deductive scheme to verify that hypothesis. In this way fruitfulness seems to consist simply in the capacity to suggest some, and not others paths to follow. If we add this difficulty to the problem of distinguishing between the heuristic principle and the explicit reference to a rigourous result previously obtained, it seems to me that neither of the three views throws much light on Gödel's actual philosophy of mathematics. As we shall see in the two next chapters, his defence of the analytic-synthetic distinction, as well as his mathematics-physics analogy, proceed in a similar way. But first let us have a look at the way in which Gödel's global conception appears in the unpublished manuscripts.

## Realism and the basic content of the unpublished essays

Here I offer merely the skeleton of Gödel's main arguments presented in the sixth and final version of "Is mathematics syntax of language?", which appears in full in chapter 4 of part II of this edition, without offering any critical treatment (see my 1991b). But first let us recall Gödel's main theses in philosophy of mathematics, according to the three main topics involved (I follow Feferman, CWI, pp. 30–31):

1. *Mathematical objects (ontology)*:
   - they have an independent existence and reality;
   - their reality is analogous to that of physical objects.

2. *Mathematical axioms (semantics)*:
   - they refer to the reality of mathematical objects;
   - their truth depends upon objective facts, not upon our constructions;
   - they are analytic, not tautological (i.e., true in virtue of the meanings of the concepts involved, not of their definitions).

3. *Mathematical knowledge (epistemology)*:
   - we have something "similar" to perception of mathematical objects;
   - even if there is no direct perception of them, their existence is necessary to deduce immediate sense perception;
   - we need to assume mathematical objects and axioms to obtain a satisfactory system of mathematics (as physical objects and laws are necessary to explain appearances);

- we can regard as mathematical "sense data" those propositions whose generalizations require "transcendent" assumptions;
- mathematical intuition can provide mathematical knowledge, although it can be cultivated through study;
- mathematical objects and axioms can be also justified through their fruitfulness, but that is less certain than intuition.

Every one of these philosophical theses is somehow referred to in version VI of the essay on Carnap. However, in order to follow Gödel's own order of exposition in that essay, I will use his threefold analysis of the conventionalist conception. According to this analysis, conventionalism is based upon the following theses, which are presented by Gödel as capable of refutation: (i) mathematical intuition can be replaced by conventions; (ii) there are no mathematical objects or facts; (iii) mathematical conventionalism is compatible with empiricism. Yet at the end of the section I shall add something relative to the way in which all this material can be related to the three philosophical fields which have been pointed out above (e.g., ontology, semantics and epistemology).

Gödel's disfavour with these three assertions is formed by his view that the "philosophical terms" occurring in them (e.g. "replacing", "fact") are not well defined, so it can be shown that their meanings are artificial.

The detailed arguments follow (my textual remarks always within brackets). As for (i), the thesis that mathematical intuition can be replaced by conventions, Gödel admits that the application of certain syntactical conventions concerning the symbols (and the propositions containing them) can lead us to the same results as the application of mathematical intuition; but he insists that we need to know that these syntactical rules are consistent, for from inconsistent rules we can derive any proposition. However, already for proving that consistency mathematical intuition is needed.

Concerning thesis (ii), the denial of mathematical objects and facts, Gödel resorts to five arguments.

(1) If we say that mathematics or logic imply nothing about experiences, the same is true of laws of nature, because to make actual empirical inferences from laws of nature we need mathematics or logic. This is so in the sense that mathematics adds something to the laws of nature which is not expressed in them, or in the sense that for predicting the result of an observation general laws about an infinity of physical elements may be required. Thus, although it is true that mathematical propositions do not express physical properties, but properties of concepts, these last properties "are something quite as objective and independent of our choice as physical properties of matter".

(2) As for the possibility of disproving mathematical axioms, Gödel continues to resort to the same line of comparison with empirical science. Thus, he says those axioms can be disproved by an inconsistency derived from them, so they must have some content, such as all propositions which may be wrong. And this leads us to attribute existence to the objects of a successful, consistent science of mathematics, in the same way that we attribute existence to the objects of a successful science of physics.

The two following arguments are very different, for they are related to the meaning of the symbols used in mathematics, in the sense that this meaning involves something else than arbitrary convention. The arguments are difficult but brief, so I quote them in their entirety.

(3) As for the possible voidness of conventions Gödel writes (always in the same manuscript of version VI, unless otherwise indicated):

> Even if it were admitted that mathematics can be based on conventions about the use of symbols, its voidness of content still would not follow. For symbolic conventions are void of content only in so far as they *add* nothing to the theory in which they are made, but they may very well imply propositions of this theory. If, e.g., on the ground of the empirically known associativity of some physical operation a convention about the dropping of brackets is introduced, then from this convention the associativity of the operation in question, i.e., an empirical proposition, follows. If a mathematical convention is introduced on the

basis of its consistency, the situation is quite similar. For this fact of the consistency of the convention, again, is expressible in the main system in which it is made and the convention implies, although not this consistency itself, still certain only slightly weaker propositions, i.e., substantially the same facts as those which justified its introduction.

(4) Concerning the possible arbitrariness of conventions, we read:

> Even if mathematics is built on rules of syntax, this makes it not a bit more conventional (in the sense of "arbitrary") than other sciences. For according to the positivistic point of view the rules for the *use of a symbol are the definition* of its meaning, so that different rules simply introduce different meanings, i. e. different concepts. But the choice of the concepts is free also in other sciences. Moreover syntactical rules, which introduce new symbols not as mere abbreviations for combinations of symbols present already, must be consistent [the transcription of this word is uncertain] and compatible with all empirical possibilities and, therefore, are *very* far from arbitrary.

We come now to the argument (5). This argument, the last one related to thesis (ii), is clearly epistemological, as it is based on the "perception" of the mathematical objects involved in mathematical intuition, although, as usual, it is developed by returning to the comparison with "empirical" perception. Gödel begins by advancing the main assertion: "There exist experiences, namely those of mathematical intuition, in which we perceive mathematical objects and facts just as immediately as physical objects, or perhaps more so". The argument is that the difference between an empirical datum (e.g., "this is red") and a logico-mathematical datum (e.g., *modus ponens* or complete induction) "consists solely in the fact that in the first case a relationship between a concept and a particular object is perceived, while in the second case it is a relationship between concepts". Thus, in the same way that the syntactical conception tries to reduce mathematical intuition, which is the "mathematical sense", to conventions about the use of symbols, the same could be done with some "physical sense", in order to avoid its corresponding objects or facts. The parallelism is even more complete, for in the same way that we would need more and more conventions to relate numerous independent sense-impressions, the same has to be done in mathematics, where more and more axioms are needed to solve its problems, and these axioms are justified only by intuition or experience.

As for the last thesis (iii), the supposed compatibility of conventionalism with empiricism, Gödel is extremely brief: "it suffices to say that, if consistency and compatibility with ['empiricism', ?] (which must be known in order to be able to introduce the mathematical axioms as 'conventions') is based on empirical induction mathematics is not a priori true; on the other hand to prove it by mathematical intuition is not compatible with empiricism".

Gödel finishes by pointing out that the plausibility of the positivist viewpoint about the voidness of content of mathematics depends upon two circumstances: (1) that the logical concepts seem to belong not to the subject matter of the empirical propositions, but to the means of expression; (2) that no possibility is excluded by logically true propositions [presumably because they, in being always true, are compatible with any facts whatsoever]. However, it can be countered that the first argument "does not exclude that the logical concepts may be made the subject matter of non-empirical propositions", while the second forgets "that there are different levels of possibility".

Now we can take, provisionally, the above outline of this particular essay as an illustration of the sort of arguments contained in the manuscripts appearing here. Then the question is: how can these manuscripts help us to improve our understanding of Gödel's philosophical viewpoint? I think that they throw light on some of Gödel's theses, at least in comparison with the well-known details in the publications, which are often obscure. Yet in so doing Gödel's general views continue to be difficult to grasp. Let us see the way in which this is so, by following, in so far as this is possible, the three general philosophical fields which were mentioned above.

*Ontology*. Mathematical objects cannot be eliminated, for we always need primitive terms and axioms to express their properties. Mathematical axioms must have content, for they can be rejected in case an inconsistency is derived from them. Primitive concepts would then be the raw matter with which to build up the rest of the concepts. Summing up: the role these concepts play in the formalism would be identical with the one for physical objects in physics: the assumption of their existence would act as a fruitful hypothesis.

*Semantics.* Mathematical axioms are not void of content, for they have physical consequences only in conjunction with others, in the same way as do natural laws. Hence mathematics and empirical science are similar; in fact we can even interpret formal derivation as some sort of observation. Moreover, also as it happens in empirical science, the rules which state the use of the symbols determine, at the same time, the definition of their meanings. This can be used to reject arbitrariness, for once we accept definitions, what can be derived from them is purely objective.

(iii) *Epistemology.* Mathematical intuition cannot be replaced by a set of conventions, for these should be consistent, and this consistency can be proved only by means of mathematical intuition. On the other hand, the knowledge of the truth of mathematical axioms depends on mathematical intuition although they are introduced to relate independent problems, in the same way that physical objects are introduced to relate independent sense data.

Unfortunately, in version VI of the Carnap essay Gödel did not use many interesting arguments and considerations which appeared in former versions and the Gibbs lecture (we shall refer to some of them in the next chapters). Nevertheless, it has to be regretted the fact that Gödel never agreed to the publication of this brief, useful essay, especially in view of the fact that the volume for which it was destined has been widely read.

## Other relevant ideas and authors: a brief survey

To finish, I shall attempt to relate these arguments to a few main ideas which can be found in the writings of other authors, but only with the aim to place Gödel's philosophy of mathematics into a more general historical framework. However, here I merely will point out some links, without developing them (for more details see chapters 2 and 3 of this part).

First of all, we have in the literature at least two more or less clear occurrences of Gödel's "vicious circle" argument (this argument appeared in print for the first time in Rodríguez-Consuegra 1991b). According to this argument we can say that, in a certain sense, the conventionalist view, which intends to dispense with logic and mathematics, actually presupposes them. A general form of the argument (or at least of a parallel one) can be found in Quine's well-known attacks against certain forms of conventionalism (see Quine 1936 and 1963; especially 1936, p. 352): logic is presupposed in conventions, as we need logic to infer whatever we need from conventions. In a more particular form the argument can been literally found in Benacerraf's and Putnam's introduction to their celebrated anthology (1983, p. 23; already in the 1964 edition): admissible conventions have ultimately to be consistent, and this can be interpreted as mathematical fact; unless we admit that we could make contradictions true by convention as well. I do not know whether Gödel had heard about these – or similar – ideas, for he hardly gives us references to the "philosophical" literature, but in ay event this points out to an interesting link with Quine, his famous rejection of the analytic-synthetic distinction, and some of the philosophical consequences. Let us then have a look at this link.

Quine's rejection of the analytic-synthetic distinction was precisely a result of his systematic comparison of mathematics and logic with empirical science. He finally came to see both types of science as the two extremes, only gradually separated, of a continuum. The rejection itself took place in print in White 1950, that is one year before Quine's 1951 publication of the legendary "Two dogmas of empiricism" (although this was obviously due to Quine's massive influence). However, the kernel of the idea can be traced back to the celebrated discussions between Carnap, Quine and Tarski in the early 1940s (and even to the writings of J. S. Mill). As Carnap wrote (1963, pp. 64–65), he was then convinced that there is a sharp distinction between logical and factual truth, while Tarski and Quine thought that the distinction could be, at best, a matter of degree. The recent publication of a letter of Tarski to White (White 1987), written in 1944, has thrown some more light on the point. Tarski, rather surprisingly, wrote that not only logical and mathematical truths do not differ in origin from empirical truths, for "both are results of accumulated experience", but also that, as a consequence of this fact, empirical changes could provoke changes in the underlying logic (White 1987, p. 31).

It is also well-known that Gödel himself participated in some of the discussions which took place at that time, especially with Carnap and Tarski. As we have seen, Gödel always maintained an important difference between mathematical and factual truth, yet a question naturally arises: to what extent did Gödel himself contribute to originate the idea of the rejection of the distinction? And how is it possible that Gödel did not arrive at the rejection itself, especially when he always pointed out the deep similarity between mathematics and empirical science?

There are relevant elements about the relationship between Gödel and Carnap, which have been preserved in Carnap's records, as studied by Wang (1987, pp. 50 ff) and Coffa (1987, pp. 552 f). According to them, it seems that Gödel had criticized Carnap's conception of analyticity already in 1932, that is, precisely when Carnap was working on the first versions of *The Logical Syntax of Language*. And this criticisms were made precisely to persuade Carnap away from the mere syntactical viewpoint of language to a more semantical one. Through the same records (Wang 1987, p. 52), it can also be seen that in 1948 Gödel regarded as usual the comparison between theoretical physics and set theory, in the sense that while physics is confirmed by sense perceptions, set theory is confirmed by its consequences in arithmetic. And we have seen how, in his manuscript on Carnap, Gödel regarded arithmetical data as very similar to sense data. Thus, it may seem that there was only a further step to be taken until arriving at the full recognition that no sharp distinction can be drawn between mathematical propositions and physical laws.

Yet, as we have seen by quoting Tarski's letter of 1944, this line of thought leads directly to empiricism, and empiricism was for Gödel completely unacceptable as it is unable to give an account of the a priori character of mathematical intuition. Therefore, when Gödel insisted upon the analogy between mathematics and empirical science, he must have thought not only of the empirical character of mathematics, but of the intuitive character of empirical science. Gödel admitted the distinction between the analytic and the synthetic to avoid empiricism, but he also needed to avert the danger of presenting mathematics as something merely tautological. Thus, he was apparently forced to give an account of analyticity through a semantic (rather than syntactic) concept of "meaning". However, as Quine has taught us, it is very difficult to achieve a clear account of analyticity by means of meaning or similar concepts.

Curiously enough, the other great logician of this century, Bertrand Russell, arrived at a completely opposed position in the philosophy of mathematics, in spite of the fact that he began precisely with the most exuberant Platonism, through an almost complete surrendering to the linguistic account of mathematics. Thus, in a manuscript written also in the 1950s, but left unpublished for many years, we can read (Russell 1973, p. 306):

> Our conclusion is that the propositions of logic and mathematics are purely linguistic, and that they are concerned with syntax. When a proposition "p" seems to occur, what really occurs is "'p' is true". All applications of mathematics depend upon the principle: "'p' is true" implies "p". All the propositions of mathematics and logic are assertions as to the correct use of a certain small number of words. This conclusion, if valid, may be regarded as an epitaph on Pythagoras.

As we have seen, for Gödel, Pythagoras should live on forever.

# 2

# The analytic-synthetic distinction

This chapter tries to throw light on the first of Gödel's two main theses in the philosophy of mathematics, namely that mathematical propositions are analytic. To this end, an overview of similar conceptions is presented first in which the views by Frege, Russell, Wittgenstein, Carnap and Quine are expounded. Then Gödel's view is analyzed, both in his publications and in the manuscripts which appear in this edition. The presentation of Carnap's detailed attempt to define analyticity in his *The Logical Syntax of Language* (1934) may seem rather long in comparison with the ones devoted to the other authors, but it should be recalled that the Gödel manuscripts appearing here were a direct philosophical reaction to Carnap's viewpoint, and there is no other sufficiently detailed presentation of this viewpoint in the literature. I am convinced that Gödel's manuscripts cannot be properly understood without having at least a summary of Carnap's construction available.

## Frege

G. Frege's *Die Grundlagen der Arithmetik* (1884), was mainly an investigation in search of the ultimate foundations of arithmetic in order to show that the rigour of its proofs was based on something other than the mere safety against contradictions. With that aim Frege tried not only to state the doubtless truth of arithmetic statements but also to reach "insight into the dependence of truths upon one another", in a way according to which "the further we pursue these enquiries, the fewer become the primitive truths to which we reduce everything" (Frege 1884, §2). Thus, for Frege, the celebrated distinctions a priori/a posteriori and analytic/synthetic are not referred to the content of judgement, but to "the justification for making the judgement", that is to say, to "the ultimate ground upon which rests the justification for holding it to be true", in complete independence of the psychological circumstances (§ 3). Therefore, the main objective would be to find out the proof of every statement, in order to show it to depend on the primary truths.

In applying that method to a particular statement two things can happen: (a) that we arrive at general logical truths and definitions, in which case we already have an analytic truth; (b) that we cannot build up the proof without the need for non-logical truths, in which case we would have a synthetic statement. As for the a posteriori truths, their proof would be valid by means of some recourse to facts, while the a priori ones should be proved from general laws, which should be non-provable in themselves (§ 3).

Then Frege faces the problem of the supposed vacuous character of the science of numbers, which was assumed for some philosophers to be reducible to "mere identities". First, he argues that the fact that we can make our calculations in a purely formal way does not imply that the symbols we use are devoid of meaning or content (§ 16). Second, he points out that "The truths of arithmetic would then be related to those of logic in much the same way as the theorems of geometry to the axioms. Each one would contain concentrated within it a whole series of deductions for future use..." (§17). With that Frege seems to point out to a difference between a logical a priori (i.e., a priori in itself), and an epistemological a priori (i.e., a priori to us).

Bertrand Russell's discovery of his celebrated paradox of set theory in 1901 led Frege to abandon the belief in logical objects and analyticity. After a few years of uncertainty, he seems to have searched for the genuine foundations of mathematics not in logic, but in geometry. Yet he saw this science as an intuitive and a priori one in the Kantian sense, although synthetic. Thus he moved the borderline between both terms of the distinction until all was made synthetic. As we shall see, Russell, too, was forced to move the borderline, although not so drastically.

## Russell

After having maintained in former works that the analytic-synthetic distinction is indefensible (following Bradley), and that the necessary does not coincide with the analytic (following Kant), B. Russell faced the enormous task of giving a logico-philosophical foundation to mathematics in his *Principles of mathematics* (1903). There he wrote: "...Kant never doubted for a moment that the propositions of logic are analytic, whereas he rightly perceived that those of mathematics are synthetic. It has since appeared that logic is just as synthetic as all other kinds of truth ..." (p. 457); and also: "All mathematics, we may say – and in proof of our assertion we have the actual development of the subject – is deducible from the primitive propositions of formal logic: these being admitted, no further assumptions are required" (p. 458).

The bases for these two fragments could be the following. In the first Russell might have been thinking of Moore's theory of proposition, according to which every truth is relational (then synthetic), in the sense that every proposition is a unique concept, which cannot be reduced to its components (this theory was partially inherited from Bradley; see Rodríguez-Consuegra 1990, 1991c). A possible, though unlikely alternative could be that Russell was already thinking of the problematic axioms which would later be needed to develop the logicist program (e.g., the axioms of infinity and reducibility). Those axioms would be rather synthetic than analytic, for they could not be proved, and only their consequences would make our acceptance of them advisable. But this alternative is practically excluded, for by that time Russell did not yet begin the detailed task of developing that program in the mathematical sense.

As for the second passage, it is not (contrary to Taylor 1981) that Russell recognized the analytic (the tautological), except in the purely logical sense that there is no need to resort to the Kantian intuition (the space-time intuition). Thus, it seems that for Russell there were two senses of the synthetic: the Kantian one, and a rather obscure one which would be relevant in the theory of proposition, and also for the pure access to certain axioms which would have a real content (they would not be mere tautologies). So the fact that in 1903 he dispensed with the Kantian intuition does not mean that the result is analytic, if by analytic we understand what is devoid of content.

In 1912 Russell wrote: "But Kant undoubtedly deserves credit for two things: first for having perceived that we have *a priori* knowledge which is not purely 'analytic', i.e. such that the opposite would be self-contradictory; and secondly, for having made evident the philosophical importance of the theory of knowledge" (1912, p. 82). Then he added that Kant also deserved credit for having perceived that "...all the propositions of arithmetic and geometry are 'synthetic', i.e., not analytic: in all these propositions, no analysis of the subject will reveal the predicate" (pp. 83–84). Here we have another interesting element.

On the one hand, Russell makes clear that for him "analytic" means basically "tautological", rather than "a priori", as is shown by his criticism of the subject-predicate general scheme for the analysis of propositions. This was really basic for him, for it would justify the fact that the process of deduction does provide us with new knowledge, and not mere tautologies. Thus, deduction is analytic in the logical sense, but synthetic in the sense that it leads us to something new which is not previously contained in the premises. On the other hand, the very mathematical statements are analytic, for they are provable by means of pure logic from the logical axioms, while they are synthetic in the sense that they are not tautological; namely, that they have a proper content which is not derivable from the subject-predicate scheme of analysis. There is then no contradiction; Russell had different reasons to call deduction, as well as the mathematical axioms themselves, both analytic and synthetic.

## Wittgenstein

According to L. Wittgenstein's *Tractatus logico-philosophicus* (1922), a tautology is one of the two extreme possible cases in which we can organize the truth possibilities of a non-elementary proposition. A proposition is tautological when it is true for any possible assignment of a truth value to the elementary propositions which are its components. Hence Wittgenstein writes that both tautologies and contradictions lack sense, although this does not mean that they are nonsensical. They belong to the symbolism itself (like the arithmetic zero), although they are not "pictures of reality" and do not represent any possible situation, as it happens with genuine propositions (*Tractatus,* 4.46...). From this conception Wittgenstein sketches a whole philosophy of logic (6.1...), according to which logical propositions are tautologies – they say nothing – and makes clear that he is referring to the analytic propositions in the traditional sense.

There is here a most interesting element, precisely the one which Carnap was to use to criticize Wittgenstein's apparent restriction to sentential logic (i.e., propositional logic, the logic with no quantifiers). It is the following: "It is possible – indeed possible even according to the old conception of logic – to give in advance a description of all 'true' logical propositions. Hence there can *never* be surprises in logic. One can calculate whether a proposition belongs to logic, by calculating the logical properties of the *symbol.*" (6.125–6.126). Now, Carnap interpreted (in *Logical Syntax,* see next section) this passage in the sense that Wittgenstein seemed to be thinking merely of sentential logic, namely the one which is decidable (see first section of chapter 1 of this part), i.e. the one in which there is an algorithmic method to decide whether any non-elementary proposition is tautological, contradictory or merely consistent: the truth tables (and similar effective recourses).

The interesting nuance which does not seem to have been seen by Carnap is that Wittgenstein is also referring to proofs, while Carnap seemed to criticize only the limitations of the truth table methods. Wittgenstein wrote: "And this is what we do when we 'prove' a logical proposition. For without bothering about sense or meaning, we construct the logical proposition out of others using only *rules that deal with signs.*" (6.126); and he finishes by saying that at any rate this is not essential in logic, for the starting point for proofs (the initial propositions) has to be accepted as being tautological without proof, and that proofs are only the mechanical means to obtain an easier recognition of tautologies.

This seems to me to be rather strange for several reasons. First, Wittgenstein seems to have extended what he said earlier on sentential logic to elementary logic (or first-order logic, the logic *with* quantifiers) through his recourse to proofs, although he does not make any explicit reference to quantifiers. Second, the reference to the initial propositions, which have to be accepted without proof, seems to coincide with the former point, for in first-order logic the method of the truth tables is very limited. Third, at any rate, it is clear to me that Wittgenstein seems to believe that elementary logic is decidable, for he insists that we can mechanically recognize any tautology (unless he is thinking, again, merely of sentential logic; but then it is unclear why he speaks about proof instead of merely about the usual semantic elementary methods – truth tables and the like). As is usual with Wittgenstein, it is quite difficult to throw light on what he wrote, but I think our analysis has been worth the bother for his ideas constitute a significant stage in our overview, as we shall see immediately in connection to Carnap.

### Carnap

Wittgenstein's stance, at any rate, made a new effort to define analyticity necessary, hence R. Carnap's task in *Logische Syntax der Sprache* (1934) to determine more rigourously and with more clarity the meaning of "analytic", and also the difference, if any, between "analytic" and "tautological". The problem was that, in the meantime, Gödel's results came to light (1930–31; see chapter 1 of this part). They completely changed the scenery and affected not only the current views on completeness and incompleteness, but also the underlying former presupposition that it is obvious that logical and mathematical propositions are analytic (after Frege, Wittgenstein and, partially, Russell). In particular, Carnap tried not only to "completely" define the notion of analyticity, in such a way that every mathematical proposition was shown to be either analytic or contradictory, but also to prove thereby that there cannot be "special" mathematical propositions which are not analytic. With that he seemed to have been thinking of Russell's *Principia* problematic axioms (infinity, reducibility, choice), or even struggling against the danger of interpreting Gödel's celebrated undecidable sentence exhibited in his proof of 1931 – which truly says of itself that it is not provable – as being synthetic (for it is not analytic, in the sense of not being provable in the system) yet a priori (for it is true by means of considerations which are comprehensible without recourse to empirical data).

Carnap's view, which is usually assimilated to logical positivism (and this even to its most "popular" presentations), is often described as something which defends the reduction of mathematics to logic and the analytic character of mathematical propositions. Carnap himself asserted, for instance in 1931a, that logical and mathematical propositions are tautological, and rejected any possible synthetic a priori knowledge. However, in 1931b Carnap quickly added several interesting nuances. First, he pointed out the difficulty of accepting the logical character of Russell's problematic axioms in his logicism (choice, infinity, reducibility). Second, he suggested lines of solution, for infinity and choice, based on their supposed conditional character, and for reducibility based on F. Ramsey's proposal (1926, 1931), which dispensed with "ramifications" in the theory of types.

Carnap's own detailed solution to all these problems was developed in his *The Logical Syntax of Language* (1934 German, 1937 English; *Logical syntax* in the following), where he originally tried to either eliminate Russell's problematic axioms or make them analytic, while rejecting any Platonism by means of the reduction of every infinitary problem to conventional decisions. This would not have been so difficult had Gödel's results not disturbed so deeply the original programme. In fact, Gödel's incompleteness theorem made it apparently impossible to formally define mathematical truth by separating it radically from provability (while Church's 1936 undecidability theorem made hopeless the search for a general decision procedure for logics higher than sentential logic).

That is why Carnap tried in *Logical syntax* to search for "another completeness" for logic and mathematics, which was able to somehow accomodate some of the outcomes of Gödel's results. This was attempted by means of a syntactic – and also actually semantic – reduction of mathematical propositions, until they were shown to be analytic, by indicating that objective mathematical truth coincides with analyticity in a specific, very sophisticated sense. Thus, the possible temptation to interprete true but non-provable propositions in a given system (like Gödel's "G" statement; see first section of chapter 1 of this part) as being synthetic a priori should vanish. Naturally, the highest price which had to be paid consisted in making analyticity relative to a particular language, namely to a set of premises and rules. On the other hand, the best recourse to do that was, paradoxically, provided by Gödel himself: arithmetization of syntax. This was cleverly used by Carnap to show, in passing, that Wittgenstein's universalism (the belief in a universal, inescapable language) was unjustified. Let us examine, after this general presentation of the main goals of Carnap's program, the main features of his construction of analyticity in the several languages he introduced in that work.

Carnap's "Language I" contains the arithmetic of natural numbers, but is limited in the intuitionistic sense, namely, by admitting only *definite* (finitary) numerical properties, whose possession can be determined in a finite number of steps. However, Language I is not properly a definite language because it contains sentences which are not definite, in other words, resoluble – provable or refutable – ones.

The rules of consequence, which go beyond those of transformation, are justified for Carnap because there are cases in which every sentence of a sign type is provable but the universal corresponding sentence is not (thus, Carnap was obviously thinking of certain technical details of the proof of Gödel's incompleteness theorem). So, *to create this possibility*, the terms "consequence" and "analytic" are introduced by Carnap, who assings meanings which are broader respectively than "derivable" and "provable" (and the same with "contradictory" and "refutable").

To do so, Carnap admits infinite classes of sentences, which for him is equivalent to speak of syntactic "forms" (defined in extension). Two rules are then introduced which define consequence in terms of derivation; one of them refers to infinite classes of sentences, so two methods of deduction are stated. A derivation is defined as an infinite series of sentences, and a series of consequences as a finite series of classes – not necessarily finite – of sentences. In the case of derivation, every step is defined, but not the relation "derivable", which is defined in terms of a complete chain of derivation; in the case of a series of consequences, every step ("the relation of direct consequence") is already indefinite (infinitary), let alone the relation of consequence itself.

Then Carnap gives the following definitions for Language I: an analytic sentence is a sentence which is a consequence of the null set of sentences (i.e., a consequence of every sentence); a contradictory sentence is a sentence such that every sentence is one of its own consequences; an L-determinate (logically determinate) sentence is one which is analytic or contradictory; and a synthetic sentence is one which is neither analytic nor contradictory. Hence every provable sentence is analytic and every refutable one is contradictory, although the reverse does not hold (so Carnap is obviously referring to the proof of Gödel's theorem, where a true but not provable sentence is exhibited, which is supposedly analytic). For the same reason every logical sentence is analytic or contradictory, the descriptive (or non-logical) sentences being the only ones which are synthetic. Fermat's theorem, i.e., that $x^n + y^n = z^n$ is not resoluble for n > 2 by using positive integers, is offered as an instance of an L-determinate sentence, although Carnap admits that we do not know whether it is analytic or contradictory.

Carnap's "Language II" contains not only definite concepts and intuitionistic arithmetic of natural numbers (which were already in I) but also indefinite (non-finitary, or infinitary, as a matter of fact) concepts, classical mathematics (including set theory) and the possibility to formulate physical sentences. In addition, Carnap introduces the following: non-restricted quantifiers ("unlimited operators"); a simple theory of types (Russell used a more complicated, "ramified" one adding also orders to types), in order to avoid any reducibility axiom (with several types and numerical levels according to certain rules); variables of several types, including second-order quantification (i. e., one whose quantifiers range not only over individuals but also over predicates and functions); identity; the equation "0 = 0", which means "the true" and is used to make truth-value assignments; and symbols for sentences (which were absent in Language I, a language of positions), including quantification over them. The rest of elements of Language II are constants.

The paragraph 34 (a–i) is basic for our purposes. Carnap writes there that one of the main tasks in the logical foundation of mathematics consists in stating a formal criterion of validity. In other words, it would amount to determine the necessary and sufficient conditions to determine valid (correct, true) sentences of Language II. In order to obtain completeness for that criterion we should, according to Carnap, go beyond the definite both in itself and in the individual steps of the deduction; i.e., we should adopt the method of consequence, where we operate, not with sentences, as in the method of derivation, but with classes of sentences, which could be infinite.

The rules would be the same as the ones already formulated in Language I (§ 14), and the conclusion is: "In this way a *complete criterion of validity for mathematics* is obtained. We shall define the term 'analytic' in such a way that it is applicable to all those sentences, and only to those sentences, of Language II that are valid (true, correct) on the basis of logic and classical mathematics". The same will be applied to "contradictory", "L-determinate", and "synthetic" (which is defined as that which is not L-determinate). Thus, the method of consequence makes no essential difference between Languages I and II, except as regards quantification, which is not restricted in Language II, and the corresponding changes in the definitions.

Then Carnap criticizes Wittgenstein and Moritz Schlick (one of the founders of the Vienna Circle); the first one for depending upon the method of the truth tables, which is valid only for sentential logic; the second for saying that analytic judgments are immediately seen as being a priori, as soon as we understand their meaning. Carnap argues that, as it is shown by Fermat's theorem and similar instances, it is possible to clearly see the rules of application of a sentence without seeing every possible consequence (a line of thought later followed by Gödel, as we shall see). However, we should say that Carnap's methods are not effective either; namely they are not usable in practice because they are infinitary – this seems to underly his criticisms of Wittgenstein and Schlick. Thus, although Carnap reached more accurate definitions, they do not lead us to a complete criterion of validity, as Carnap himself would later recognize.

In Language II the process is the reverse of that in Language I: first "analytic" and "contradictory" are defined, then the concept of consequence is introduced. Yet the way is first paved by means of reduction rules for sentences; these are able to transform the sentences into a standard form to simplify their handling. Thus, the problem of defining "analytic in II" is reduced to the problem of defining it for the corresponding sentences under a particular standard form: the prenex form, which depends upon arranging the quantifiers at the beginning of every formula by following very precise rules. That is made by following a double process of "valuation" and "evaluation", and by applying Gödel's method of arithmetization of syntax.

The basic idea is explained by Carnap: "We shall not define the term 'analytic' explicitly, but instead we shall lay down rules to the effect that a sentence of a certain form is to be called an analytic sentence when such and such other sentences fulfil certain conditions – for instance, when they are analytic" (§ 34*c*). Carnap's description of the whole process is complex and difficult, but I think it could be summed up this way. Let *F* be a property in a sentence *S*. Then we will examine not the defined sentences of its type, but "all the possible valuations (*Bewertungen*) for *F*", where we understand "possible valuation" as a syntactic assignment, namely the assignment of a value to *F*, in particular of a class of numerals N (which are called "accented expressions"). Now, if *B* is a particular assignment of *F*, and if *F* appears in N as its argument in any place of *S* (for instance, in the partial sentence 'the successor of the successor of 0 has the property *F*'), then "this partial sentence is – so to speak – true on account of *B*, if N is an element of *B*, and otherwise false".

The evaluation of the sentence *S* is then carried out, which will be a transformation according to which the partial sentence is replaced by '0 = 0' if N is an element of *B* and by '0 ≠ 0' if it is not. So *S* "will be called analytic if and only if every sentence is analytic which results from *S* by means of evaluation on the basis of any valuation for '*F*'" (§ 34*c*), and will be contradictory when at least one of the resulting sentences is contradictory. The same will be applied to the rest of types, and a similar procedure is introduced for functions and descriptive predicates. It is then when particular rules of valuation and evaluation are given, and it is clearly pointed out that the evaluation will be applied to sentences, not to partial sentences.

The explanation in the excellent Coffa 1991 may help. According to Coffa, possible valuations are numerals, classes of numerals, functions from numerals to numerals, no matter whether they are definable in the system or not. Once valuations and evaluations have been applied, to say that a matrix is mathematically true or analytic for a given assignment of its "value-bearing signs" means the following (p. 292):

> The idea is simply to find the truth value of the matrix relative to that valuation by evaluating it "from inside out" so to speak, starting from the propositional atoms in the matrix and moving outward. Theorem 34c.1 states that for every quantifier-free reduced matrix $M$ and for every evaluation of its value-bearing signs $v$, the described process of evaluation (in conjunction with the reduction process) leads in a finite number of steps from $M$ and $v$ to either "$0 = 0$" or "$0 \neq 0$" (i.e., the truth values of all matrices are determined, for each appropriate $v$).

But Carnap, says Coffa, by avoiding the "realistic mode" of speaking, does not say that the rules of evaluation determine the truth value of a matrix free of quantifiers in a given valuation. Rather, its rules tell us how to transform one sentence into another, in a way according to which every mathematical sentence of Language II is to be transformed, in the end, either in "the true" ($0 = 0$), or in "the false" ($0 \neq 0$).

Only then does Carnap define "analytic" and "contradictory" in II, by giving rules which define them for classes of sentences, for a given sentence, and for reduced sentences (which are "analytic in respect to certain valuations"). Thus, once again, what is defined is not that which is "analytic" in general, but rather the cases in which a sentence is analytic as a function of partial sentences and certain valuations (see Quine's criticism below). For Carnap this is equivalent to an effective procedure, at least in the sense that it always finishes with a sentence being either true or false. Hence, Carnap finally says that, although a sufficient and necessary criterion of analyticity has been given for every sentence, there is no general method for resolving every individual step, let alone a general criterion. Thus, "analytic" and "contradictory" are indefinite (infinitary) terms.

With this as a starting point, the chain of definitions follows by defining "L-determinate" and by clarifying that "analytic", and similar terms, were introduced in order to overcome the limitations of such terms as "provable". Thus, Carnap's ambition to search for completeness and for Gödel's "objective mathematical truth" is somewhat clarified. Hence, every logical sentence is L-determinate, i. e., either analytic or contradictory (although the absence of a general method of resolution is admitted), and synthetic sentences are always descriptive, namely, not logical. Now: every logical sentence which is defined as analytic is also provable, although there are analytic sentences which are not provable (which becomes clear when one thinks of Gödel's theorem and the true, undecidable sentences involved), and every defined logical sentence is resoiuble, although there is no general method of resolubility.

Part IV of the book is devoted to investigating general syntax, which does not refer to a particular language, but to every language or to every language of a given class. In order to reach my limited goals here I will restrict myself to an attempt to show that Carnap's logical sentences constitute a complete, although irresoluble (undecidable) language.

Rules of formation and transformation are given first, as well as the c-terms ("c" from consequence, as "d" is from demonstration, or derivation), in order to make a complete division into analytic and contradictory sentences possible. What follows from these rules are the definitions of logical content; the distinction between the logical and the descriptive (relative to each language); the distinction between logical and physical rules (which is a matter of convention); and the introduction of the L-terms. Now: every analytic sentence is valid; every logically valid sentence is analytic; and every logical sentence is L-determinate: "there are no synthetic logical sentences".

As for consistency and completeness, Carnap shows again his deep assimilation of Gödel's results: a language is inconsistent if each of its sentences is valid, and consistent otherwise. However, a non-contradictory language may be inconsistent, for although it does not contain any d-contradiction (a contradiction in respect to rules of demonstration), it may still contain a c-contradiction (one which depends on the rules of consequence). This is the reason for introducing the term "consistent", which is applied to languages which do not contain any type of contradiction and corresponds to Gödel's expression "ω-non-contradictory".

Thus, although it is recognized that every consistent language containing a general arithmetic is irresoluble, it is also affirmed that a language sufficiently strong, although it is irresoluble (undecidable), it can be both determinate and complete, provided enough indefinite (infinitary) rules of transformation are stated, at happens with logical sub-languages from I and II. (However, as we shall see, the appropriateness of our speaking here of completeness is rather doubtful, as the analyticity of certain sentences can only be stated in a higher language.)

The following classification by Carnap holds for a language both *irresoluble* and *complete*, and applies to its logical sentences (p. 210).

| | | | |
|---|---|---|---|
| (d-terms) | demonstrable | irresoluble | refutable |
| (c- and L-terms) | valid<br>*analytic* | | contravalid<br>*contradictory* |

The gaps for undecidable (irresoluble) sentences which are analytic (valid), or contradictory (contravalid, or invalid) are clearly evident; there is no place for synthetic logical sentences. Thus, the synthetic would be what is not L-determinate, in other words, neither analytic nor contradictory. In this way the sort of completeness Carnap was looking for could be reached, despite the incompleteness and undecidability of the general system. At any rate, it seems that the reader, on arriving at this point, might be under the impression that the whole construction suggests that "analytic" is somehow equivalent to "mathematically true", although we may not know this to be so in every case.

Carnap continues by explicitly dealing with the relationship of his methods to the Liar paradox, which will lead to the most severe limitation of the work: the clear assimilation of analyticity to truth in Tarski's sense. Thus, Carnap says that our not being able to reconstruct the celebrated paradox with his concepts of "non-demonstrable" or "refutable" is due to the fact that not every analytic sentence is also demonstrable, and that not every contradictory sentence is also refutable, which would had been the case had we used "true" or "analytic". Thus, the paradox can be reconstructed by assuming that "analytic" and "contradictory" are defined in a syntax formulated in the same language.

We thus come to the theorem 60c.1, which is essential for our goals: "If S is consistent, or, at least, non-contradictory, then '*analyt-*

*ic (in* S)*' is indefinable in* S". And the same is true for the rest of the c-terms, as long as they do not coincide with d-terms (valid, consequence, equipollent, etc.). Therefore: "If a syntax of a language $S_1$ is to contain the term 'analytic (in S)' then it must, consequently, be formulated in a language $S_2$ which is richer in modes of expression than $S_1$". Hence "analytic in I" is not definable in Language I, but it is definable in Language II, and we have the same situation for higher languages. On the contrary, although "demonstrable in I" cannot be defined in I, for it is an indefinite concept, "demonstrable in II" can be defined in II, as we can see in Gödel's true but undecidable sentence G affirming of itself that it is not provable (because of the arithmetization of syntax in the same language; see the first section of chapter 1 of this part).

I think this is a clear sign that the whole construction – as Quine told me recently in a letter – does not go beyond Tarski's and Gödel's. In the end, the whole manoeuvre cannot be used to extend analyticity, in the traditional sense, to all mathematics, nor can it therefore be used to offer a genuine criterion of universal validity. It is one which is always limited to a given language. Naturally, the strongest objection to the construction proceeds also from Gödel and Quine: in carrying it out, Carnap built up something very similar to Tarski's definition of truth, in the sense that truth, like analyticity, is indefinable in the same language, so it is always in need of a higher language. Summing up: this illustrates the impossibility of overcoming the Gödelian limitations, by showing that the sort of alternative completeness that Carnap was looking for was rather an illusion, and that mathematics cannot be completely expressed in one language, no matter how rich it may be. However, as Carnap explicitly admitted, that completeness can be reached through an infinite hierarchy of languages, so the logicist program can, at least in a limited sense, be maintained.

In his *Introduction to Semantics* (1942) Carnap renounced the definitions from *Logical syntax* (§39), by admitting that no general definition of logical truth in general semantics is known (§§13, 16). However, it is interesting to see his later controversy with Quine, for it reflects philosophical features which are relevant to understanding contemporary analytic philosophy. Among these features there is his capacity to continue to believe that mathematics is analytic anyway, and that this fact can be formulated with precision by resorting only to the necessary formal apparatus, and a series of conventions, in spite of Gödel's results. As Quine told me, such a belief was always for Carnap as solid as the Rock of Gibraltar, which perhaps is not that surprising when we learn that for Gödel mathematics is analytic too.

## Quine

In his Harvard lectures of 1934 (in Creath 1990), W.V. Quine already clearly detected that the analytic-synthetic distinction is relative, in the sense that what sentences we regard as analytic is a matter of convention. But Quine did not respond to the following questions: Is this equivalent to deny the distinction itself, or rather to seeing it as a matter of degree?; What is the relationship between that relativity and the one which was admitted by Carnap in *Logical syntax* (§ 82), when he said that the distinction between the logical rules and the physical rules is a matter of degree?; What is the relationship between all this and the relativity of analyticity which already had been pointed out by Frege, which depended upon the particular axioms and rules we choose? (see above). Let us see the concrete arguments which have been used by Quine in order to clarify such questions, as well as Carnap's replies. This is relevant to the understanding of the problems involved in Gödel's ideas, for, as we shall see, his notion of analyticity seems to be related to some of these questions. Paradoxically, Gödel's defence of analyticity is close to Carnap's, while he strongly criticizes Carnap's particular construction of analyticity, in spite of the fact that Carnap's analyticity, as we have seen, tried to be a response to the tremendous problems created by Gödel's results.

In Quine's "Two dogmas..." (1951) there are three basic arguments against analyticity (in general, I follow the excellent presentation of Creath 1990). (i) Every attempt to clarify the notion is useless for science, for it finally resorts to more or less obscure notions, such as meaning, synonymy, necessity, semantical rule, etc.. This holds when it is said that a sentence is analytic in virtue of the meaning of its terms, or that it is analytic in case it proceeds from logical truths by replacing synonymous expressions.

(ii) Carnap's attempts to clarify the notion of analyticity for artificial languages fail. In fact, the specification of the corresponding rules already resorts to the term "analytic", so the only thing that many supposedly analytic sentences have in common is that for one reason or another they fall under the title "analytic". In this manner Carnap succeeds in defining "analytic in L" (L being a particular artificial language), but he fails to define analyticity for every L.

(iii) Scientific reasoning is constituted by a body of beliefs, none of which is fundamental, but which can be revised through the principles of simplicity and conservation. Logic and mathematics are in the centre of that body, so we prefer to make changes in the periphery when necessary, but there is nothing really immune. There are conventions, of course, but there is no way to locate that which is conventional as opposed to that which is not. Hence there is no possibility of any analytic-synthetic absolute distinction. For Quine, this whole alternative constitutes a better program than Carnap's.

It is unfortunate that Carnap's reply of 1952 remained unpublished until Creath 1990, as it is much better than his reply to Quine in the Schilpp volume of 1963. According to Carnap, in order to clarify the meaning of certain vague terms in natural language, certain decisions have to be made, for vagueness is unacceptable in science. For instance, it can be said that a certain sentence is a postulate, and that therefore it is analytic. As for semantic rules, they do not make the definition of "analytic in L" arbitrary: "the defined concept embraces what philosophers have meant, intuitively but not exactly, when they speak of 'analytic sentences' or more specifically, of 'sentences whose truth depends on their meanings alone and is thus independent of the contingency of facts'" (Creath 1990, p. 430).

In addition, for Carnap, Quine's requirement that the *explicatum* (*explicans*, I would say) be applicable to every system is unreasonable, and we cannot meet it through semantic or syntactic concepts. The same line of thought is refined when Carnap writes that for Quine the determination of analytic sentences is useless, for first we have to understand the notion of analyticity. The reply is then the same:

we possess a practical comprehension of the notion of analyticity which is enough for many cases, but which is not enough "for other cases or for theoretical purposes. The semantical rules give us an exact concept; we accept it as an explicatum if we find by comparison with the explicandum that it is sufficiently in accord with this" (p. 431).

Carnap clearly accepts Quine's thesis that there is no rule immune to observation, although, unfortunately, without quoting his splendid paragraph 82 of *Logical syntax*. For Carnap, it is true that we have an increasing feeling of doubt as long as we try to revise sentences which are more and more faraway from observation, but this does not imply that there is no definite borderline between logic and physics. So he writes: "the difference between analytic and synthetic is a difference internal to two kinds of statements inside a given language structure; it has nothing to do with the transition from one language to another. 'Analytic' means rather much the same as true in virtue of meanings" (pp. 431–2). But as everything can be changed in the logical structure of a language, the same sentence can be analytic in a system and synthetic in another. As the meaning depends upon the rules and not upon the facts, an analytic sentence is not revisable as long as that rules do not change. So, although the synthetic truth value is changed, this does not affect the analytic sentences. Therefore, the distinction can always be stated in respect to a well-determined, artificial language, although this is not possible for a given historical language.

The correspondence between both authors (in Creath 1990) concerning these matters is extensive and contains certain pages which are really excellent, but it largely is based on the former arguments. In the following, I quote something by Quine which is relative to his rejection of the distinction between natural and artificial languages, that seems to me to be fundamental.

In a letter of 1943, Quine already referred to the fact that speaking of artificial languages changes nothing, for in them we do as if there were people who really spoke them, so a behaviourist definition of "analytic" and of "sentence" remains basic, in the sense that it is a part of the pragmatic substructure of semantics. In a letter of 1954 (thus later than the reply we referred to above), Quine wrote that his arguments were not limited to natural language: "It is indifferent to my purpose whether the notation be traditional or artificial, so long as the artificiality is not made to exceed the scope of 'language' ordinarily so-called, and beg the analyticity question itself" (Creath 1990, pp. 437–8). So, Quine continues, if the definition of analyticity is included in artificial languages, then the distinction itself between natural and artificial is false.

Quine finishes the controversy by saying that artificial languages are not uninterpreted notations, for every predicate has its unique extension, and we have the same situation for logical signs: "But they are not of kind (b) if, as I suspect, 'languages' of kind (b) are conceived as embodying a complement of transformation rules – a ready-made stipulation of a boundary between analytic and synthetic. In view of 'Two dogmas' and our years of discussion, I think the above brief remark will suffice to convey my meaning" (p. 438). Summing up: Quine does not admit that the distinction can be traced, because he believes that the interpretation which already underlies artificial languages makes it blurred anyway, so admitting the distinction natural-artificial amounts to the same as admitting the original analytic-synthetic distinction.

In his contribution to the 1963 Schilpp volume devoted to his old master (the volume for which Gödel's Gibbs lecture was intended to appear) Quine did not accept Carnap's syntactic reduction of *Logical syntax* either. His arguments are somehow similar to the ones we shall see as formulated by Gödel himself, and run as follows. As for Carnap's Language I (see the former section), the formulation of logical truth is closely syntactic, in the usual way of the axiomatic formalizations with inference rules, but for Gödel's results we know that such a procedure is not suitable for general mathematics, set theory, or stronger languages such as Carnap's Language II itself. Thus, for Language II, Carnap really resorts to techniques which are very similar to the ones appearing in Tarski's semantics. On the other hand, Quine adds, Carnap's final result is a more liberal syntax in which we admit (i) names of signs; (ii) an operator expressing the concatenations of expressions; and (iii) the totality of the logical and mathematical vocabulary, as a sort of auxiliary machinery. Then Quine writes (1963, p. 400):

> So construed, however, the thesis that logico-mathematical truth is syntactically specifiable becomes uninteresting. For, what it says is that logico-mathematical truth is specifiable in a notation consisting solely of (a), (b), *and* the whole logico-mathematical vocabulary itself. But *this* thesis would hold equally if "logico-mathematical" were broadened (at *both* places in the thesis) to include physics, economics, and anything else under the sun; Tarski's routine of truth-definition would still carry through just as well. No special trait of logic and mathematics has been singled out after all.

Quine finishes by saying that, in the end, the mathematical we would have to accept would be larger than the apparatus we wanted to explain, as is shown by Tarski's theory of truth.

This anti-reductionist argument by Quine brings to mind his parallel vicious circle argument of 1936 (which is repeated here), according to which logic cannot be explained by conventions, for then we would have to resort to logic to apply these conventions to particular cases. The 1963 version of the argument is brief and interesting, and it recalls also Gödel's argument in our manuscripts against conventions (which to acceptable require a proof of consistency): "logical truths, being infinite in number, must be given by general conventions rather than singly; and logic is needed then to begin with, in the metatheory, in order to apply the general conventions to individual cases" (pp. 391–2).

## Gödel

In order to correctly understand Gödel's views on analyticity in the unpublished manuscripts appearing here, it is advisable to become familiar with his published views on the same subject.

I know of only two published writings by Gödel in which the analytic character of mathematics is explicitly discussed: his article on Russell of 1944, and the remarks on the undecidability results from 1972 (CWII). In 1944 the problem is posed whether the axioms of Russell's and Whitehead's *Principia mathematica* can be regarded as analytic. According to Gödel, analyticity can be understood in two ways, the tautological sense and the properly analytic sense. In the tautological sense (CWII, pp. 139–9)

> it may have the purely formal sense that the terms occurring can be defined (either explicitly or by rules for eliminating them from sentences containing them) in such a way that the axioms and theorems become special cases of the law of identity and disprovable propositions become negations of this law. In this sense even the theory of integers is demonstrably non-analytic, provided that one requires of the rules of elimination that they allow one actually to carry out the elimination in a finite number of steps in each case.

The ultimate reason, as Gödel says in a footnote, is that this would suppose the existence of a decision procedure for arithmetic, which cannot exist. Thus, mathematics cannot be analytic in the first sense, for it is undecidable. (This, by the way, recalls Carnap's similar objection to Wittgenstein's conception of mathematics as tautological.)

Gödel continues to say that if we admit sentences of infinite (and non-denumerable) length in the process of reduction, then every axiom of *Principia mathematica* would be analytic for certain interpretations: "But this observation is of doubtful value, because the whole of mathematics as applied to sentences of infinite length has to be presupposed in order to prove this analyticity, e.g., the axiom of choice can be proved to be analytic only if it is assumed to be true" (p. 139). I think this is also a clear, although implicit, reference to Carnap's "indefinite" rules, and also to the fact that in Carnap's *Logical syntax* the axiom of choice was presented as analytic (although the axioms of infinity and reducibility were eliminated in that work). Certainly Carnap spoke about indefinite rules, and defined them as those which can be defined only in an infinitary way, that is, they are rules which admit quantifiers with no restriction in their domains.

To explain the properly analytic sense of analyticity, Gödel wrote: "a proposition is called analytic if it holds 'owing to the meaning of the concepts occurring in it', where this meaning may perhaps be undefinable (i.e., irreducible to anything more fundamental)" (CWII, p. 139). It would seem then that the axioms of *Principia math-*

*ematica* (except the axiom of infinity) would be analytic for certain interpretations of the primitive terms: "namely if the term 'predicative function' is replaced either by 'class' (in the extensional sense) or (leaving out the axiom of choice) by 'concept', since nothing can express better the meaning of the term 'class' than the axiom of classes and the axiom of choice, and since, on the other hand, the meaning of the term 'concept' seems to imply that every propositional function defines a concept" (*ibid.*). Gödel adds a footnote to say that this does not contradict his former position that mathematics is based upon axioms with a real content, for the existence itself of the concept of "class" (set) constitutes already an axiom of such a kind.

Also, Gödel adds the following illuminating passage: "it is to be noted that this view about analyticity makes it again possible that every mathematical proposition could perhaps be reduced to a special case of $a = a$, namely if the reduction is effected not in virtue of the definitions of the terms occurring, but in virtue of their meaning, which can never be completely expressed in a set of formal rules" (*ibid.*). So he is speaking of something objective, independent from our definitions. Thus, the notion of meaning is used as being relative to reference, i.e., to the objective concepts which are denoted by our terms. However, as objective, but infinite concepts cannot be embraced through finitary rules, they cannot be expressed by a set of such rules (in case these rules are admitted to be infinitary, then the former objection applies). But then the basic problem is that the notion of meaning for concepts is taken to be indefinable, so we have to resort to other passages in which Gödel speaks about meaning and about the nature of concepts.

Yet something relevant is added in the same place: "The difficulty is only that we don't perceive the concepts of 'concept' and of 'class' with sufficient distinctness, as is shown by the paradoxes". According to Gödel, this led Russell to build those concepts up, but in so doing only certain fragments of logic remain, unless their content is reintroduced through infinite propositions or the axiom of reducibility. Gödel, however, says it is preferable to try to make clearer the true meaning of "class" and "concept" by considering them as objectively existing realities and to use the simple theory of types and axiomatic set theory as the basic instruments. With that it seems to me that Gödel is referring to the analysis of concepts in the way Frege and Russell did, that is, by trying to analyze the appearances ("the given") in terms of the fundamental, perhaps both logically and epistemologically. In his "Some remarks on the undecidability results" (1972) he wrote: "there *do* exist unexplored series of axioms which are analytic in the sense that they only explicate the content of the concepts occurring in them, e.g., the axioms of infinity in set theory, which assert the existence of sets of greater and greater cardinality or of higher transfinite types and which only explicate the content of the general concept of set" (CWII, p. 306). However, here the difficult expression "content of the concepts" appears with no explanation, so we can add it to the list of former unexplained ones (meaning, concept, term, etc.).

We come now to the unpublished manuscripts. Useful remarks and arguments can be found in them which contribute to clarify some of the problems which have been pointed out. In particular, the Gibbs lecture (1951) contains two versions of a strong attack against Carnap's syntactic view, which can throw some light on Gödel's view of the analytic and the tautological. The first version of the attack was finally rejected and not actually read, but it is clearer and does not depend upon the analogy between mathematics and physics.

There (see the full text in chapter 2 of part II), Gödel writes that the most simple version of the syntactic conception is the thesis that mathematical propositions express only certain aspects of linguistic conventions, in the sense that they are true only by virtue of the definitions of the terms involved, so that they would be reducible to tautologies. But he adds that this reduction is impossible for the following arguments.

(i) It would entail the existence of a mechanical decision procedure for every mathematical proposition, but this kind of procedure cannot exist, not even in number theory. Curiously enough, this argument seems to me to be the same one that Carnap used against Wittgenstein in *Logical syntax*, so it can hardly be used against Carnap himself. As we have seen, Carnap was fully aware of Gödel's incompleteness results in that book; that is why he resorted there to rather semantic concepts and to the admission of a different type of "completeness", which was to be compatible with the existence of undecidable propositions.

(ii) It is true that the truth of mathematical axioms can be derived from certain semantic rules which are chosen for logical and mathematical systems, but in such a derivation the logical and mathematical concepts themselves have to be used as referring to symbols and their combinations. Therefore, in order to prove the tautological character of mathematical axioms their truth has to be assumed first (as takes place with Ramsey's expression of infinite length and Carnap's infinite sets of finite propositions). Thus, instead of defining their meaning through syntactic conventions, we must first know their meaning to be able to understand those conventions. On the whole, this criticism seems to me very similar to the one which appeared in Gödel's Cantor paper (1947), and should be inserted in the long tradition against reductionism. As Quine wrote in criticizing Carnap's construction of analyticity (see above), the totality of the mathematical vocabulary is already admitted in it as an auxiliary element, so it is already presupposed in the intended elimination of the mathematical content proper, and no reduction is actually reached. Gödel's priority is, however, undeniable for his argumentation was put in writting before this time.

(iii) A proof of the tautological character of mathematical axioms is equivalent to a proof of their consistency, but this proof cannot be reached unless we use stronger means than the ones contained in those axioms. At any rate, to prove the consistency of number theory (or of any other stronger system) certain "abstract" concepts have to be used, i.e., concepts which are not referred to sense objects, such as "set", "function of integers", "derivable", or "there is", and these concepts are not syntactic.

In the other version of the Gibbs lecture, the one which seems to have been actually read by Gödel in Providence, there are several arguments concerning what he calls "relations between concepts", which are very interesting for the notion of analyticity.

(1) There obviously are non-tautological relations between mathematical concepts, for certain primitive terms always have to be assumed in mathematical axioms, and these axioms are not reducible to tautologies, but they follow from the meaning of those primitive

terms. With that Gödel seems to me to follow the Frege-Russell tradition, according to which the mind's eye has to look for the most simple and primary concepts and axioms, with the aim of using them to be able to define and derive the remaining concepts and theorems of mathematics. Gödel actually goes back to the most primary concepts, but with that the problem arises as to whether the axioms "implicitly" define those concepts somehow, or whether they express some of their properties according to our faculty of mathematical intuition. The problem is faced in the following argument.

(2) We can say that the axioms which determine the concept of set are analytic, in the sense that they are valid in virtue of the meaning of the term "set", but this does not mean that they are tautological, for the assertion that there exists a concept of set that satisfies those axioms is so obviously full of content that it cannot be understood without already using the concept of set itself or any other similar abstract concept. With that, Gödel seems to somehow clarify the former problem, for he seems to be saying that it is the axioms which develop the concept of set previously given to us. However, in mentioning the existence of a concept of set which satisfies the axioms, he does not discuss the possibility that there exist several concepts of set, according to the different axioms we choose (although, as we shall see later, he did discuss the question in the Cantor paper). At any rate, in writing that the axioms are analytic in virtue of the meaning of the term "set", Gödel does not clarify whether it is the axioms which define this meaning, or the concept of meaning itself, which by that time was being strongly criticized by Quine through the notions of synonymy, substitution, etc.. Perhaps Gödel assumed the concept to be something fundamental, and therefore "indefinable".

(3) This concept of analyticity is objective, for it depends on the nature of the concepts, and not subjective, which would depend on the definitions, and it is opposed to the synthetic, which depends on the properties and the behaviour of things. However, it has a content, as can be seen by the fact that it is possible that an analytic proposition is undecidible, since our knowledge of the concepts can be so limited and incomplete as our knowledge of the world of things. Also, this can be seen if we explain the paradoxes of set the-

ory as being similar to optical illusions, where, although we do not see what it is real, we are perfectly able to explain why we see what we do.

Here we have a passage which could have been written even by Carnap, as he maintained a very similar position in respect to the analyticity of instances such Gödel's undecidible statement, Goldbach's conjecture and the like. The problem is the way in which Gödel might justify the attribution of analyticity to those examples, which he seems to handle as being simply the same as intuitive "mathematical truth". The reference to the paradoxes doubtlessly means that we are not free to build up the concepts, for their very objectivity constitutes for us an insurmountable barrier. This would be a proof that there exist objective relations between concepts which are independent of our methods and devices, and this seems to me to be what underlies Gödel's interest in defending a certain concept of analyticity: the mere struggle against the subjective, that is to say, the conventional. In the same way, Russell attacked analyticity (see section on Russell above), and he did so for the same reasons: he thought that it unavoidably implied the property of being tautological, i.e., lacking any content. Gödel, as well as Russell, separated the analytic from the tautological and defended the character of mathematical facts as having full content.

In the essay on Carnap, version II (chapter 3 of part II of this book), Gödel goes back to the same problems, but he adds much new material. The first relevant point consists in pointing out the failure of the reductionist view (Ramsey, Carnap), as regards its intended refutation of the thesis according to which mathematics can be replaced by syntax of language. Gödel, after accurately defining what is actually involved in that thesis, returns to his arguments against reductionism. Thus, he writes that the requirements themselves of the syntactic view (with regard both to the syntactic rules and to the derivation from them of the mathematical axioms, and to the consistency proof) call for the use of syntactic concepts only, namely, finitary concepts referred to finite combinations of symbols and evident axioms about them. Otherwise, we would have to resort to "abstract" concepts which cannot be understood without mathematical intuition, and this is what the syntactic view tries to avoid.

However, Gödel adds, the syntactic conception cannot satisfy these requirements. Ramsey admitted propositions of infinite length, and Carnap used infinitary syntactic rules and arguments. Therefore, the syntactic program fails because the replacement of intuition by certain rules on the use of symbols destroys any hope of expecting the existence of consistency, and also because for the proof of consistency an equally strong intuition is required. As I have already pointed out above, this criticism against Carnap is somewhat surprising, as he was already aware of the implications of Gödel's incompleteness results, as well as of the infinitary character of his rules and of the lack of a proof of consistency for them. Perhaps Carnap would have replied that precisely because consistency was undemonstrable, it made no sense to insist on it for his rules, since it makes no sense to insist on it for mathematics. So consistency would have to be "empirical" in both cases, that is, relative to the good results actually obtained.

The second relevant point against the syntactic conception is the intended refutation of the thesis that mathematical propositions are void of content, where interesting remarks about the analytic and the tautological again appear. Gödel writes that mathematics has content, for certain undefined terms and certain axioms about them are always needed; here he clearly goes back to some of the arguments we have seen above. However, he adds that for such axioms there cannot be any rational foundation other than (i) the immediate perception of their truth (according to the meanings of the terms involved, or through an intuition of the objects which fall under them); (ii) the inductive arguments on the basis of the success in the applications.

This essentially is the same as we have seen above. Thus, the problem is also the same as before: Gödel finally depends on a rather vague mathematical intuition, as well as on the belief in primary concepts and axioms whose truth we must perceive immediately. The new shade is the allusion to the empirical success, but it pertains to our other problem: the analogy between mathematics and physics. We can then go on to explicitly consider the similarities between the formal and the empirical sciences.

# 3

# The mathematics-physics analogy

The basic goal of this chapter is to delve deeply into Gödel's second great strategy in the philosophy of mathematics: the analogy between deductive and empirical sciences. Moreover, I shall try to explore the holistic, and even conventionalist, implications of the analogy, such as it appears in some contemporary philosophers of mathematics who have defended the analogy to some extent. To do this, it has been necessary to present an overview of the most important precedents in the use of the analogy, such as Russell, Hilbert, Carnap, Tarski, Quine. After that, I shall present Gödel's views on the analogy, in both his published and unpublished writings. Surprisingly, most of these authors maintained very different philosophical conceptions, in spite of the fact that they all believed that mathematics and physics are very similar, both in their objectivity and methods.

## Russell

B. Russell's *The Principles of Mathematics* (1903) had two basic objectives: (i) technically, to prove that mathematics is only concerned with concepts which are definable in terms of a small number of fundamental logical concepts (the primitive ideas), and that all its propositions are derivable from a small number of fundamental logical principles (the primitive propositions); (ii) philosophically, to explain the concepts which are taken as fundamental and indefinable. This second objective is the one which led Russell, already in 1903, to formulate an analogy between mathematics and empirical science; to my knowledge, it is the first one to appear in his writings. To reach that objective involves, according to Russell, "the endeavour to see clearly... the entities concerned, in order that the mind may have that kind of acquaintance with them it has with redness or the taste of a pineapple" (1903, p. xv).

It is then that Russell resorts to the analogy: "Where, as in the present case, the indefinables are obtained primarily as the necessary residue in a process of analysis, it is often easier to know that there must be such entities than actually to perceive them; there is a process analogous to that which resulted in the discovery of Neptune, with the difference that the final stage – the search with a mental telescope for the entity which has been inferred – is often the most difficult part of the undertaking" (1903, p. xv). Finally, Russell applies the analogy to the concept of class, by recognizing his failure in "perceiving" it, because of the celebrated paradox about whether or not the class of all classes not being members of themselves is a member of itself (if it is a member of itself, then it is not, and vice versa).

In 1906 Russell faced the problem again, now under the form of the nature of mathematical intuition and the sort of reasons we have to maintain the truth of logical propositions, and he solved it by means of the analogy with empirical science. One of the most interesting passages was recently pointed out by Hao Wang (1987) as probably being the one which was referred to by Gödel in his essay on Russell (1944), so this paper by Russell is likely to have somehow influenced Gödel. On the whole, Russell introduces there a series of highly original ideas which could be summed up this way (1906, pp. 194–5).

The method of "logistics" is, for Russell, basically the same as the one applied in other sciences. It is therefore characterized by fallibility and uncertainty, the mixture of induction and deduction, and the confirmation of its principles by means of the agreement with observation. The objective of this method is by no means to eliminate intuition, but to refine it until one discerns the general laws from which we can deduce consistent results, which are then to be confirmed intuitively. The primitive propositions have not always to be intuitively obvious; they have to be also acceptable for inductive reasons. Thus, it is required "that, among their known consequences (including themselves), many appear to intuition to be true, none appear to intuition to be false and those that appear to intuition to be true are not, so far as can be seen, deducible from any system of indemonstrable propositions inconsistent with the system in question". Among the systems which satisfy the former conditions we should, according to Russell, choose that whose primitive propositions are the clearest and in the smallest number, in the same way that Newton's laws are preferable to Kepler's.

As for intuition, it is not, for Russell, infallible, as can be seen by the paradoxes; thus we see that an element of uncertainty is unavoidable, as it happens in empirical science. Finally, the rules of logic should be applied mechanically to primitive propositions, especially to the doubtful cases. If, after the process, they are confirmed, we have inductive evidence of their validity; in the same way the scientist does when trying to verify certain hypothesis. However, an element of imagination is needed in order to see what consequences could be false and to search for the crucial cases: "If, finally, we can arrive at a set of principles which recommend themselves to intuition, and which show exactly how we formerly fell into error, we may have a reasonable assurance that our new principles are at any rate nearer the truth than our old ones".

In 1907 Russell devoted a whole lecture to the analogy, which was read before the Cambridge Mathematical Club. There we can find basically the same arguments, but also several new remarks on the global organization of our thoughts, which exhibit certain holistic nuances. Also, there are other remarks which have to do with realism, pragmatism and the role of intuition in a way that, as it was already noted by Lackey in his introduction, it seems that it is the mathematical theorems (the consequences) that what can justify the axioms (the logical premises), and not the other way round.

Russell's starting point is that the propositions which are easier to learn have only a moderate degree of complexity. In mathematics the premises give us usually the reason for believing in a given propositions, but in the foundations this relation is inverted (1907, pp. 273–4):

> Our propositions are too simple to be easy, and thus their consequences are generally easier than they are. Hence we tend to believe the premises because we can see that their consequences are true, instead of believing the consequences because we know the premises to be true. But the inferring of the premises from consequences is the essence of in-

duction; thus the method of investigating the principles of mathematics is really an inductive method, and is substantially the same as the method of discovering general laws in any other science.

Concerning the method which is common to every science, Russell's ideas are also fundamental. According to him, in every science we start from a body of propositions which we feel secure about; these are our empirical premises, or observable facts. We can ask then about their consequences or about that from which they are consequences, till obtaining simpler propositions (general laws) to deduce them (1907, p. 274):

> The laws only become as certain as the empirical premises if we can show that no other hypotheses would lead to the empirical premises, or if (what may happen in mathematics) the laws, once obtained, are found to be themselves obvious, and thus to be capable of themselves becoming empirical premises. ... But when the general laws are neither themselves obvious, nor demonstrably the only hypotheses to account for the empirical premises, then the general laws remain merely probable; though the degree of probability may be indefinetely heightened as observation and experiment increase the number of empirical premises which they account for.

Curiously enough, through such methodological unification Russell does not distinguish between what we usually call the empirical, which is the object of our perception, and that what we could call the abstract empirical, which would presumably be the object of our intuition, so I think he is clearly anticipating a subject dear to Gödel. However, Russell's comparison between the two types of hypotheses and the way to increase their degree of probability is to be emphasized. In doing that he was almost surely thinking of the problematic axioms he was forced to admit in his attempts to logically found mathematics (especially the axioms of choice, infinity and reducibility).

Also, Russell explicitly mentioned several advantages which can be obtained when we find simple logical premises out: they give us a higher probability of detecting a possible element of falsehood; they organize our knowledge; they have more consequences than the empirical premises, so they lead us to discover many more things. Russell's example are Newton's laws, which for him are opposed to the observable motions of the planets. As for arithmetic, he writes that if we take the usual arithmetic propositions as empirical premises, we arrive at the logical premises, from which we can deduce complex theories, as for instance Cantor's theory of transfinite numbers. I think here we should emphasize at least two elements: (i) the holistic element appearing in the reference to the advantage of a higher organization of our global knowledge, and (ii) the "quasi-empiricist" interpretation of logicism, where the analogy between mathematics and physics is almost perfect.

Russell continues by clarifying the role of intrinsic obviousness in any body of knowledge (that what we sometimes call intuition, in a non-Kantian sense): "In the natural sciences, the obviousness is that of the senses, while in pure mathematics it is an *a priori* obviousness, such as that of the law of contradiction" (1907, p. 279). However, several precautions before the apparently obvious are also pointed out, in the best tradition of the methodology of science: (i) intrinsic obviousness is a matter of degree: in case of a conflict we prefer the most obvious; (ii) even the highest degree does not lead us to infallibility, so consistency with other obvious propositions has pre-eminence, as is clearly seen in the case of hallucinations; (iii) when we can deduce one obvious proposition from another, it has, as a whole, a higher degree of obviousness, which can be applied to whole deductive complex systems: "Thus, although intrinsic obviousness is the basis of every science, it is never, in a fairly advanced science, the whole of our reason for believing any one proposition of the science" (1907, p. 279).

All these elements are immediately applied to symbolic logic, in particular when we arrive at the paradoxes. According to Russell, the more advisable procedure consists in stating a hierarchy of results according to their degree of obviousness, and then in isolating the premises from which the paradoxes seem to depend in order to modify them. But then our procedure would be the same as in empirical science: to avoid the false and to keep what we regard as true, as can be seen in examples taken from physics and mathematics, e.g., both

Boyle's law and Frege's premises would be only approximately true. The ultimate results would be the complete unification of the method of science: "The various sciences are distinguished by their subject matter, but as regards method they seem to differ only in the proportions between the three parts of which every science consists, namely (i) the registration of 'facts', which are what I have called empirical premises; (ii) the inductive discovery of hypotheses, or logical premises, to fit the facts; (iii) the deduction of new propositions from the facts and hypotheses" (1907, p. 282).

Once again, one should point out Russell's emphasis on the holistic elements when describing the way to organize our knowledge as a whole. It is not the intrinsic obviousness in itself that is used, but rather the global consistency of the system. This is not yet explicit holism, for the impossibility of verifying isolated hypotheses and similar conditions is still lacking, but if we add the methodological unification which he tried to reach, it is really difficult to deny that in these passages there is not only an obvious similarity to Gödel's analogy between mathematics and physics, but also something as a sort of precedent to Quine's later holistic ideas.

In *Principia mathematica* (1910–13) such principles are first of all applied to the very problematic reducibility axiom, the one needed to avoid the paradoxes according to the theory of logical types. This axiom is recognized as being hardly obvious, but obviousness is, for Russell, only a part of our justification to accept an axiom, and it is never indispensable: "The reason for accepting an axiom, as for accepting any other proposition, is always largely inductive, namely that many propositions which are nearly indubitable can be deduced from it, and that no equally plausible way is known by which these propositions could be true if the axiom were false, and nothing which is probably false can be deduced from it" (1910, p. 59). Russell adds that indubitability is not infallible, for it can lead us to error, while if an axiom is indubitable this is simply added to the inductive evidence which is derived from the fact that its consequences are in fact indubitable. Summing up: infallibility is never reached, for an element of doubt is always present in every axiom, as well as in every of its consequences. As for the axiom of reducibility itself, although Russell insists that the inductive evidence in favour of it is very strong, he also recognizes that perhaps it would be possible to deduce it from another more fundamental and obvious axiom. Thus we see that Russell was hardly happy with fallibilism, which seems to have been merely a provisory and anyway undesirable stage for him.

Russell's "Sur les axiomes de l'infini et du transfini" (1911a) is an important lecture that he read in Paris and devoted to the other two problematic axioms which were used in the *Principia* construction infinity and choice. He begins by admitting that these axioms cannot be proved, but that in lacking intuitive obviousness it is necessary to analyze their nature and consequences (1911a, p. 163). After showing that both axioms are necessary and sufficient to prove important theorems about infinite progressions, Russell admits that they are epistemologically very different. So, while the axiom of choice has the form of a logical axiom, in the sense that we do not know how to prove it by resorting to empirical data, and its truth or falsehood depends upon a priori considerations, the infinity axiom is purely empirical: the empirical evidence about the divisibility of objects seems to lead us to reject that the universe has only a finite number of individuals.

It is then when Russell speaks about "hypothesis" and brings to mind once again Gödel's way to approach the problem of the status of certain problematic mathematical propositions. Empirical data are not enough to *prove* that the number of individuals is not finite, but they are enough to show that the finitistic hypothesis is less simple than the other alternative. This leads Russell to admit the infinity axiom as a sort of scientific hypothesis, namely something fallible, by applying considerations similar to the ones we already have seen concerning the axiom of choice.

Some of these ideas appear again in his paper 1911b, as well as in the *Introduction to mathematical philosophy* (1919), although with less extension and precision. It is "Logical atomism" (1924) where it is shown that such fallibilist and quasi-empiricist a conception was really lasting, at least until Russell's later more positivistic stage. Russell begins by saying that in a deductive system the premises are less obvious than some of their consequences. Then he writes (1924, pp. 325–6):

It is not the logically simplest propositions of the system that are the most obvious, or that provide the chief part of our reasons for believing in the system. With the empirical sciences this is evident. Electrodynamics, for example, can be concentrated into Maxwell's equations, but these equations are believed because of the observed truth of certain of their logical consequences. Exactly the same thing happens in the pure realm of logic; the logically first principles of logic – at least some of them – are to be believed, not on their own account, but on account of their consequences. The epistemological question: 'Why should I believe this set of propositions?' is quite different from the logical question: 'What is the smallest and logically simplest group of propositions from which this set of propositions can be deduced?' Our reasons for believing logic and pure mathematics are, in part, only inductive and probable, in spite of the fact that, in their *logical* order, the propositions of logic and pure mathematics follow from premises of logic by pure deduction. I think this point important, since errors are liable to arise from assimilating the logical to the epistemological order, and also, conversely, from assimilating the epistemological to the logical order. The only way in which work on mathematical logic throws light on the truth or falsehood of mathematics is by disproving the supposed antinomies. This shows that mathematics *may* be true. But to show that mathematics *is* true would require other methods and other considerations.

Finally, Russell insists on the same ideas, although more briefly, in the introduction to the second edition of *Principia* (1925).

Some remarks can be made about Russell's conception of the analogy. First, the constancy of the analogy is to be noted, as well as its difficult relationship to the analytic-synthetic distinction. Thus, while the analogy was always maintained, the distinction seemed to vary in different periods, and even to have been understood in a rather contradictory manner in the same period. Second, some degree of skepticism can be perfectly recognized, and even a certain degree of "if-thenism", in Russell's explanation of formal sciences. This was probably due to Russell's sensation of fragility after the discovery of the paradoxes and the need for resorting to highly controversial axioms. It is in this connection that we should insert Russell's final renunciation of the original dream to identify the logical and the epistemological orders, which seemed to have tempted him in some stages of his development.

At this moment the similarity between logic and zoology, so dear to Russell some years before, is no longer valid, and that for at least two reasons: first, because our "perception" of the logical and mathematical objects have already been shown to be dangerous and even illusory. Second, because the element of consistency is increasingly important as long as the logical and the epistemological orders are more and more separated, and the analogy between mathematics and physics is much more than a mere comparison. The holistic implications of the whole situation were unfortunately not drawn by Russell himself, at least in this field of his philosophy. Yet as we shall see later some sort of conventionalism can be avoided at this stage only by means of a conception based upon the intrinsic admissibility of certain "hard" data. Where we do not have these data available, the dangerous alliance between our analogy and the conventionalist conception (through holism) seems to be unavoidable.

At any rate, it has been necessary to start this chapter with a presentation of Russell's views, both because of their intrinsic interest and because of the fact that, although they are not very well-known, they were certainly known to Gödel (see above the reference to Hao Wang). To my knowledge they have been taken into some consideration only by Lakatos (1967, who quoted only from Russell 1924), and more recently by Irvine (1989). Yet in this last paper by Irvine the connection with the paradoxes, the similarities with Gödel, and the nexus with the analytic-synthetic distinction are missing, in the sense that it is necessary to show that the evolution of the analogy was parallel to the progressive admission of the synthetic character of some mathematical axioms: precisely those whose logical nature was frankly doubtful.

## Hilbert

Even D. Hilbert can be said to have based a great part of his philosophy of mathematics on its analogy with physics, in spite of years and years of propaganda telling us that his formalist approach had nothing to do with reality. To Hilbert mathematical propositions can be divided into two classes: *real* propositions, dealing with the finite, and *ideal* propositions, which, in admitting non-restricted quantification, refer to the infinite. Now, Hilbert said in his article 1927 that ideal objects are accepted only as long as they are governed by the rules that we introduce to handle them, but that they are really nothing by themselves. Therefore, they would be only a sort of "theoretical concept" – to use a later terminology – whose admissibility would depend upon a proof of their consistency. Only that proof would be a guarantee that the extension from real to ideal elements is a legitimate one, for it would mean that their acceptance would not involve the introduction of any contradiction in the original dominion. Thus, the relationships which result for the real objects would continue to be valid after the ideal objects had been introduced, used and eliminated (1927, p. 471). In this way, the comparison with the objects of physical theories, and the corresponding physical holism, cannot be avoided.

In fact, clearly against the usual argument that this sort of procedure would transform mathematics in a mere game of formulas, Hilbert writes (1927, p. 475):

> This formula game enables us to express the entire thought-content of the science of mathematics in a uniform manner and develop it in such way that, at the same time, the interconnections between the individual propositions and facts become clear. To make it a universal requirement that each individual formula then be interpretable by itself is by no means reasonable; on the contrary, a theory by its very nature is such that we do not need to fall back upon intuition or meaning in the midst of some argument. What the physicist demands precisely of a theory is that particular propositions be derived from laws of nature or hypotheses solely by inferences, hence on the basis of a pure formula game, without extraneous considerations being adduced. Only certain combinations and consequences of the physical laws can be checked by experiment – just as in my proof theory only the real propositions are directly capable of verification. The value of pure existence proofs consists precisely in that the individual construction is eliminat-

> ed by them and that many different constructions are subsumed under one fundamental idea, so that only what is essential to the proof stands out clearly; brevity and economy of thought are the *raison d'être* of existence proofs.

Let me point out, at least, the clear holism resulting from the analogy and emphasize the "contact" with experience, where the parallelism with the empirical is particularly clear. Also, it is important to notice that the distinction between theory and observation, which underlies the one existing between the ideal and the real, is closely related to the non-distinction between perception and mathematical intuition. As we shall see, most of these elements can also be seen in Gödel's overview of mathematics, in spite of his frontal opposition to any kind of formalism.

Curiously enough, it was precisely Weyl who, in his remarks of the same year to Hilbert's 1927 paper, seemed to accept Hilbert's holism, and even to formulate it more clearly. According to Weyl, no physical proposition can be compared to experience in an isolated way; what we reach with physical theories is not a description or copy of reality, but a symbolic construction of the world. With that, we have not only a clear precedent of Quine but also an instance that the holism of Duhem and Poincaré was fruitful also in the philosophy of mathematics, even independently from the philosophy of physics (Popper, etc.). At any rate I think Weyl's conception can be seen as an instance that something – as Quine indicated to me in a letter – was, by that time, in the air.

Weyl even emphasizes that our theoretical interest is not exhausted in what is observable, as is the case with Hilbert's "real" propositions; it also extends to the entities assumed by the theory, such as ideal objects in mathematics or electrons in physics: "according to Hilbert, already pure mathematics goes beyond the bounds of intuitively ascertainable states of affairs through such ideal assumptions" (Weyl 1927, p. 480). It is then unavoidable to liken these ideal assumptions more closely to "theoretical concepts" than to "physical objects", as Gödel used to write.

Very briefly, in his incompleteness theorems Gödel proved the existence of true, although non-provable sentences in a formal system strong enough to try to formulate every arithmetical truth; also he proved that a sentence asserting the consistency of that system is undecidable as well (see the first section of chapter 1 of this part for a more accurate presentation and more details). In the following are some of the implications of these results which I think are philosophically more relevant for the mathematics-physics analogy. To analyze them in every detail would require a more detailed discussion that the one which follows, but although there is no room here to do so, these implicatons have to be at least clearly indicated in order that the material contained in the following sections can be fully understandood.

In having proved that mathematical truth and demonstrability are very different things, in the sense that the latter can no longer be presented as an analysis of the first, the truth of certain sentences which cannot be proved in the same system becomes something more intuitive, to be decided by resorting to means other that a direct proof. In the case of empirical science, that would take place with the concepts which go beyond observation. As Gödel wrote in his letters to Hao Wang (see chapter 1 of this part), this distinction itself between mathematical truth and provability already led him to his incompleteness result, as a sort of heuristic principle. In those letters the heuristic value of the belief in an objective transfinite truth is shown. However, as we saw in chapter 1 (this part) it seems that Gödel was already explicitly searching for something similar, so those results could be viewed as something like "proofs" that certain philosophical hypotheses are true, rather than as a sort of helpful guide which assists us in a "neutral" investigation.

In any event, after Gödel's results the ideal of a secure, consistent, decidable, formalizable and complete mathematics is no longer tenable, and Hilbert's program with it (see, however, Detlefsen 1987). Yet what should be emphasized now is that, as a consequence of that, the parallelism between mathematics and the humble empirical science is again appropriate, as it was before mathematics was seen as something capable of being presented as perfect. Quine's following passage (taken from a letter to me of May 20, 1992) excellently sums up the new situation: "Before Gödel proved the contrary, I had supposed that mathematics was deductively completable, and that it was probably already complete, in sharp contrast to natural science, for which I entertained no such hope. Such was the sharp contrast, to my mind, between mathematics and natural science, and was this that was blurred by Gödel's discovery".

I think the impossibility to prove consistency, because of its implications for Hilbert's program, is especially important for our analogy. According to Hilbert, consistency is somehow equivalent to existence (see chapter 1 of this part for details), which would mean that in mathematics we would have a special criterion that would allow us to dispense with intuition. Thus, our belief in the existence of mathematical entities which go beyond the mathematically "observable" would be just a matter of formal consistency, while in empirical science the existence of problematic entities would always depend upon both observation and theory. After Gödel, who proved consistency of interesting mathematical theories to be undecidable, the situation is similar for both kinds of sciences: now mathematics has to resort to intuition (the abstract equivalent of observation) and also theory for sources of our "knowledge" of mathematical objects.

## Carnap

As we have seen in chapter 2 of this part, R. Carnap's reaction to Gödel's results is better understood by following his technical attempt to defend the thesis that, in spite of those results, mathematics can be shown to be analytic. However, it is also clear that Carnap was convinced that Gödel's results had destroyed any precise distinction between mathematics and physics, at least in the sense of the formerly understood distinction. In this connection, *Logical syntax* (1934–37) contains a paragraph (§ 82), which can be regarded as a genuine Quinean treatise of philosophy *avant la lettre*. The following is a summary of some of its ideas, especially of the ones which directly lead to holism and conventionalism. To make the exposition more systematic I will precede each paragraph with a notation of its main subject.

(All the materials and quotations proceed from the mentioned § 82, unless otherwise indicated.)

*The language of physics.* According to Carnap, the logical analysis of physics, as a part of the logic of science, is the syntax of the physical language. Such logical analysis is constituted by means of rules of the formation for sentences, which are classified into concrete sentences (without non-restricted variables) and general laws (without constants). Both L-rules (logical rules) and P-rules (physical rules) can be used as rules of transformation from sentences to sentences, which are called P-primitive sentences. There will be also primitive laws, although synthetic descriptive sentences, taken as P-primitive sentences, will be admitted too.

*Observation and verification.* Protocol-sentences express, according to Carnap, the results of observation. A particular physical sentence will be verified by drawing consequences from it through the application of the transformation rules. Once new sentences have been obtained under the form of protocol-sentences, these will be compared with protocol-sentences previously admitted, thus to be confirmed or refuted by them. This idea, which already involves the impossibility of comparing directly language and reality, can be seen as the common basis of the underlying idealism, and perhaps also as the seed to any possible holistic conception.

*Induction and the statement of new primitive laws.* According to Carnap, there is no inductive rule, hence in order to state new laws it is possible to apply only general criteria of convenience and fruitfulness. There is no rule because the content of a law is completely universal, so it always goes beyond the content of every finite class of protocol-sentences. Therefore, such laws will exhibit the character of hypotheses in connection to the protocol-sentences. They cannot be inferred from the protocol-sentences, but only selected and stated on the basis of existent protocol-sentences, which are always re-examined with the help of new protocol-sentences, which are always emerging.

*The impossibility of refutation or confirmation of an hypothesis.* Even when an hypothesis is incompatible with certain protocol-sentences we can always keep the hypothesis and drop the sentence as being a protocol one. There can be only degrees of confirmation according to the number of concordances with the protocol-sentences. In general, Carnap concludes, even the verification of an isolated hypothesis is impossible, for such an hypothesis has no consequences of the form of protocol-sentences: to carry out the deduction it is always necessary to resort to additional hypotheses. Thus, what is really tested is the total system of physics as a global system of hypotheses. It is really difficult to find out a better, more precise exposition of the Duhem-Quine thesis, as it is now called.

*The difference between mathematics and physics.* For Carnap there is no definitive rule of the physical language. This is immediately applied to the logical rules, which include those of mathematics: "In this respect there are only differences in degree; certain rules are more difficult to renounce than others". This seems to me to be fundamental, for it shows that Carnap's former holism was not only a question of physics, as it is usually presented, but it embraced mathematics as well. The result is a conception of science where holism and conventionalism come together: science depends *globally* upon protocol-sentences, but these are admitted by *convenience.* In this respect, Quine told me in a recent letter that it was doubtless this Carnapian *published* holistic conception, which he had completely forgotten, what assuaged Quine's first doubts concerning the distinction between formal and empirical sciences for a while. (By the way, Quine is proud of the recent discovery of a passage in Carnap's diary of 1933 in which he wrote that the attack by the young Quine of the distinction was essentially justified.)

*Conventionalism and indetermination as results.* According to Carnap, the construction of the physical system is not carried out through fixed rules, but by means of conventions which are inspired in practical methodological considerations. Moreover, the hypotheses always contain an element of convention which proceeds from the fact that the system of hypotheses is never univocally determined for the empirical material, no matter how rich it is. This is basic to me, for it seems to somehow embrace well known theses which are usually attributed to Quine: the empirical subdetermination of theories; the empirical equivalency of systems; the indetermination of translation, etc. (See my 1991f for more details.)

Unfortunately, the application of all this to mathematics is not very explicit in Carnap. However, in § 84 of the same work he says that the way to satisfy the best elements of logicism and formalism consists in admitting that the application of mathematics to real science is possible only by means of the inclusion of the mathematical calculuses in the global language of science. In this way the apparent contradictions between logicism and formalism vanish. The fact that we later admit primitive logical symbols (as in Russell), or only mathematical ones (as in Hilbert), is also a matter of convention. Yet with that conception it seems to me that Carnap is now clearly admitting that it is the global system of science which, without making any difference between mathematics and physics, is tested and improved by means of its consequences. It is difficult to determine whether it was Quine who influenced Carnap or vice versa.

At any rate we cannot finish without saying something on the analytic-synthetic distinction in connection with our present analogy. The analytic is conventional in the sense that we can choose what propositions we regard as analytic (in other words, not subject to observation), then there is no strict separation between the logical and the factual which proceeds from our data. This is also interesting because it shows that § 82 of *Logical syntax* is not only a mere addendum, but it underlies the whole construction, which is, therefore, ruled by conventionalism.

## Tarski

It is not know whether or not Gödel's results influenced A. Tarski concerning his philosophy of mathematics, but Tarski took a stand against Carnap, concurring with Quine's conception of the 1940's that the analytic-synthetic distinction is hopeless. However, apart of that piece of information, Tarski's explicit view on the relationship between the deductive and the empirical sciences was a mystery to all those outside a small circle of friends and close colleagues. Fortunately, White published in 1987 a letter received from Tarski in 1944, in which Tarski maintains a view which seems to follow Stuart Mill's empiricist philosophy of mathematics. The basic ideas of that letter are the following (all the references are to White 1987, unless otherwise indicated).

Tarski wrote that he was inclined to believe that "logical and mathematical truths don't differ in their origin from empirical truths – both are results of accumulated experience". Tarski's example was "*p* or not *p*", which according to him must have been a generalization from numerous particular cases in which it is true. However, he finally says that this problem lacks any philosophical nature and belongs rather to the history of science.

Concerning the supposed difference between logical and empirical premises, Tarski writes: "I am ready to reject certain logical premises (axioms) of our science in exactly the same circumstances in which I am ready to reject empirical premises (e.g., physical hypotheses); and I do not think that I am an exception in this respect". He adds that we reject certain hypotheses or scientific theories because of their internal inconsistency, their lack of agreement with observation, or with individual empirical sentences, but we could also modify many auxiliary hypotheses or add others in order to save them. As for the axioms of logic, they

> are of so general a nature that they are rarely affected by such experiences in special domains. However, I don't see here any difference 'of principle'; I can imagine that certain new experiences of a very fundamental nature may make us inclined to change just some axioms of logic. And certain new developments in quantum mechanics seem clearly to indicate this possibility. That we are reluctant to do so is beyond any doubts; after all, 'logical truths' are not only more general, but also much older than physical theories or even geometrical axioms. And perhaps we single out these logically true sentences, combine them in a class, just to express our reluctance to reject them.

Carnap's possible objection to Tarski could be that the difference between logical and physical truths would suppose a change of language (with which we would use the terms with new meanings), while a change of physical theory would not lead to such changes. Tarski's actual reply was that this would follow simply from the fact that Carnap defines the notion of meaning on the basis of logical terms and the notion of logical truth, so it is obvious that we cannot change the logical axioms without changing our language as well, for such axioms will embrace in their definition the fact that they cannot be modified without affecting the language itself.

As in every conception which assimilates logical truths to empirical ones, also here it is possible to point out the holism and conventionalism. The first is rather explicit by means of defending internal consistency as a criterion of truth, as well as rejecting isolated observations as being able to be used to falsify certain hypotheses. In addition, Tarski insists that the difference between logical and empirical truths is a matter of degree, and that both ultimately depend upon global experience. Hence the conventionalist character of any precise borderline between the two types of truths appears clearly.

## Quine

We already saw above that Quine's reaction to Carnap's attempts to clearly distinguish between mathematics and physics was partially based on the impact of Gödel's results, which made the basic difference – completeness – irrelevant. However, as Quine himself told me in a recent letter, he already had understood, by reflecting on the paradoxes, that there a complete formalization of mathematics in the strong classical sense was not possible, and that, at any rate, any formalization should refer to set theory, regarded as an extended logic (in this sense Quine says he is still a logicist).

The canonical place of Quine's attack against any precise distinction between mathematics and physics is also his famous paper of 1951. Yet we should not forget that the analogy between mathematics and physics, in the sense that both are different layers of a global conceptual vision of the world, can be found already in the original preface to *Methods of Logic* (1950). Thus, it clearly appears that the analogy can be formulated independent of the criticism to the analytic-synthetic distinction, i.e., it results from a complete pragmatic epistemological alternative to the positivist conception. That is why in the following I will emphasize those aspects of the celebrated paper of 1951 which seem to be less affected by the criticism to the analytic-synthetic distinction proper.

We can start by pointing out the nexus between the two dogmas, namely the famous distinction and reductionism. According to Quine, both dogmas are closely connected: "as long as it is taken to be significant in general to speak of the confirmation and infirmation of a statement, it seems significant to speak also of a limited kind of statement which is vacuously confirmed, *ipso facto,* come what may; and such a statement is analytic" (1951b, p. 41). In fact, Quine adds, science presents both the linguistic and the factual faces, but such a duality cannot be maintained until isolating scientific sentences one by one. Summing up: admitting sentences as units is better than admitting words, but the genuine unit of empirical significance is the whole science. (The obvious reference to P. Duhem is, however, acknowledged only in a footnote which was added later. Quine told me that he did not know about Duhem in writing the passage, so by that time he must have forgotten about § 82 of Carnap's *Logical syntax,* where, as we have seen, the Duhem-Quine thesis, as it is now known, was clearly introduced, and credited to Duhem.)

It is then that the global scheme of experience appears. To Quine, the totality of our knowledge, including logic and mathematics as well, is an artifact built up by us which is not in contact with experience but along the borderlines. A conflict with experience is the way only to internal changes, as the total field is insufficiently determined for that partial contact. Thus, no particular observation is especially connected to a particular sentence: any sentence can hold if we make certain changes in the system. Hence there is no possible sentence which is really immune to revision, as for instance takes place with the law of the excluded middle when it is seen under the light of quantum mechanics. Considerations of convenience and simplicity are then relevant, so we are reluctant to modify the most theoretical sentences of physics, logic and ontology, while the ultimate aim of the global scheme of our knowledge is simply to predict and handle experience.

Unfortunately, Quine did not say at that time that the role of the second dogma (reductionism) is no less important than the fuzziness of the analytic-synthetic distinction itself. It is to be noted that when we abandon reductionism our mathematics is closer to our physics, because if physical propositions are not reducible to sense perceptions, then they depend on a theory, as do mathematical propositions. I think this contributes to the destructive implications of Gödel's results from the other extreme. Quine has referred to the point, too, by writing that in abandoning the second dogma, logic and mathematics are mixed with physics and other sciences (1986, p. 207).

On the whole, Quine's view clearly shows the different elements we have seen before, together with the close connection between the analogy mathematics-physics, holism and conventionalism. Perhaps the pragmatic nuance is more properly Quinean, although I think it could have been defended by Carnap, too, had he not been so obsessed by strict formal methods. If my interpretation of the ideas of Russell, Hilbert, Weyl, Carnap, Tarski and Gödel is right, then Quine's role seems to have been mainly the one of the constructor of a new conglomerate by using a series of ideas already available in the scholarly atmosphere of the time. These ideas were certainly in need of a convincing global presentation, a philosophically clever practical application, and doubtless a popularization by an author prestigious enough, in the field of logic, to be able to be accurately heard by the analytic philosophers of the period, who were rather deaf to doctrines coming from any other origin.

## Gödel

Gödel discussed the analogy between mathematics and physics in two places in his publications: the Russell article (1944) and the paper on Cantor (1947–64). Before proceeding, however, I think it important to first describe what seems to me to be the essentials of what he wrote there, which was obviously the context to the unpublished essays appearing here. Then we shall come to some analysis of the relevant ideas which can be found in the unpublished materials themselves.

In 1944 Gödel devotes a couple of paragraphs to the analogy. The first of them established a link with Russell's former ideas according to which mathematical axioms have to be regarded as hypotheses to be evaluated for their consequences. The second paragraph tries to justify our acceptance of classes and concepts as real objects, in the sense of pluralities or structures of things (classes), and properties and relations of things which exist independent of our definitions and constructions (concepts). It is then that the "indispensability argument" appears: "It seems to me that the assumption of such objects is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions and in both cases it is impossible to interpret the propositions one wants to assert about these entities as propositions about the 'data', i.e., in the latter case the actually occurring sense perceptions" (CWII, 128).

First of all, there is an underlying non-distinction between physical objects and theoretical concepts, for they both make sense only within a theory, although the "theory" of physical objects is a theory only in a loose, indefinite sense. I think it is precisely this non-distinction which makes it possible to link the argument to Quine's, at least in the sense that Gödel explicitly mentions a "satisfactory system", which could refer to a conceptual scheme which, as a whole, is successful in the applications. Also, Gödel's argument can be inserted in the Russellian line that certain necessary assumptions, although lacking any observable basis, can be given the status of successful "inferences" (as opposed to "constructions").

The holistic nuance is present only in an indirect, but I think clear, way. The use of the expression "a satisfactory system", both for physics and mathematics, seems to allow us to take a further step and speak of a common system for both sciences, or at least of certain basic traits in common. However, perhaps Gödel wanted to avoid the presentation of physical objects as mere entities depending upon the theoretical support ("theoretical concepts"). This would certainly suppose the mixture of very different levels of hypothetical assumption, precisely in order not to come too close to explicit holism. The final lines of the passage doubtless constitute a clear rejection of Quine's second dogma of empiricism: reductionism, which seems to me to be a sign of the role that this rejection naturally plays in our analogy. But what seems to me really amazing is Gödel's clarity and boldness in rejecting reductionism earlier to any explicit separation between both kinds of sciences. With that, I think we can point out some link to his rejection of Hilbert's finitism, which can be seen, in the philosophy of mathematics, as a version of phenomenalism in the philosophy of physics.

As for the Cantor paper, the more philosophical version is the one from 1964, where some paragraphs were rewritten in a more philosophical manner, and to which a philosophical supplement was added. It is there where the analogy between mathematics and physics appears in its more pragmatic form: the decision about the truth of mathematical axioms, in case they lack intrinsic necessity, could possibly be made inductively, that is, by studying their success (CWII, p. 261):

> Success here means fruitfulness in consequences, in particular in 'verifiable' consequences, i.e., consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to contract into one proof many different proofs. ... There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems... that, no matter whether or not they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory.

I think this passage is useful for delving deeply into the analogy, especially because the non-distinction between physical objects and theoretical concepts appears again. In this case Gödel is more explicit for he speaks of "physical theory". Thus the interpretation in terms of theoretical concepts could be the correct one, but we should remember that for physical objects Gödel used to speak of the theory of our sense perceptions, which is also a theory. As for the parallelism between empirical verification and the sort of verification appearing in mathematics, it seems to refer to the consequences of set theoretical axioms in arithmetic. Yet there are other passages which allow us to make different interpretations, in particular the ones which refer rather to physical applications, or at least to applications in conjunction with well established physical laws. Where this is so, the holistic nuance is of course emphasized.

I come now to some of the interesting ideas which can be found in Gödel's supplement added in 1964. This supplement was partially motivated by the nominalistic attempts to interpret the continuum hypothesis, especially after Cohen's result from 1963, in a similar way to Euclid's fifth postulate. In the same way that Euclid's postulate can be said to be true in certain geometries and false in others, the continuum hypothesis, once shown to be undecidable, can be seen as true in certain set theories and false in others. Thus, it could perhaps be said that the undecidability of the hypothesis might be interpreted as being equivalent to saying that the question of its truth loses its meaning. Let me point out what seems to me to be the philosophically more interesting elements in these ideas.

Gödel writes that the undecidability argument is sound only if the terms of the axioms system under consideration are left undetermined. But Euclid's postulates refer to physical entities, while our access to the objects of transfinite set theory is mathematical intuition, for they do not belong to the physical world, "and even their indirect connection with physical experience is very loose (owing primarily to the fact that set-theoretical concepts play only a minor role in the physical theories of today)" (CWII, p. 267). Here Gödel clearly admits that the physical connection of a mathematical axiom is due to its role in physical theories, so that this physical connection may change in so far the role changes; but this depends upon the whole system of theories, so it seems to me we have again some justification to speak of holism.

Then Gödel continues by defending the existence of mathematical intuition, which is presented as the faculty, similar to sense perception, that makes our access to the objects of set theory possible, despite their remoteness from sense experience. The argument is simply the assertion that "the axioms force themselves upon us as being true" (CWII, p. 268). According to Gödel, there is no reason to be more suspicious about mathematical intuition than about sense perception, which allows us to build up theories and expect that future perceptions agree with them. The point to be emphasized here seems to me to be the deep parallelism between mathematical intuition and sense perception. Yet this parallelism makes sense only *within a theory*, the theory of sense perception, the theory of physical objects, or the one which embraces certain specific theoretical concepts which are not perceivable. As we shall see later, this could be a reason to make some criticisms of the concept of science that Gödel was in fact maintaining.

According to Gödel, mathematical intuition does not provide us with immediate knowledge, although we form our ideas in this field by means of something immediately given to us. This takes place through a process similar to that by which we form our ideas of physical objects by means of the syntheses of our sensations provided by the idea of object itself: "Evidently the 'given' underlying mathematics is closely related to the abstract elements contained in our empirical ideas. It by no means follows, however, that the data of this second kind, because they cannot be associated with actions of certain things upon our sense organs, are something purely subjective... their presence in us may be due to another kind of relationship between ourselves and reality" (CWII, p. 268). Yet some causal theory, or any other similar theory about constant connections, seems to be underlying here, without any explanation. Also, the role of synthesis attributed in a footnote to the concept of set seems to lead us to a rather Kantian theory of knowledge, which may hardly be compatible with the view of science which Gödel seems to maintain at times, in which hypotheses can be immediately verified by experience. As we shall see later, the problem of our relation to mathematical objects appears again in the manuscripts, and in an even clearer way.

Finally, Gödel insists that the question of the objective existence of the objects of mathematical intuition is exactly like the question of the objective existence of the external world. Thus, we have another criterion for the truth of mathematical axioms, besides that of mathematical intuition: their fruitfulness in mathematics and possibly also in physics. In fact, in the manuscripts he speaks more frankly about fruitfulness in physics, but this is so when he speaks about another kind of application of mathematical axioms: that which takes place in combination with well established physical theories. I think Gödel is likely to have been concerned about the danger of stating a clear connection between this new sense of fruitfulness and explicit holism. Let us therefore have a look at the relevant passages in the unpublished materials.

In the Gibbs lecture a very original argument appears in connection with Gödel's attack on Carnap's syntactic conception of mathematics, according to which there are no mathematical objects and facts, and mathematical propositions are void of content, and therefore reducible to conventions. Gödel counters this by saying, first, that to deny the existence of mathematical facts we have to use similar mathematical facts, for we have to start from the consistency of the conventions used, and also to accept abstract concepts for the proof of this consistency. Then the new argument takes place that we can also deny the content of empirical facts. To do that, we can divide these facts into two parts A, B, such that B does not imply anything in A. Then we could build up a language in which propositions expressing B would be void of content. And in case it is objected that in so doing certain observable facts B had been neglected, then we might reply that the same takes place in the field of mathematical objects, for instance by saying that the law of complete induction is perceivable by our understanding.

With that I think that Gödel was saying that the same situation takes place when we compare mathematical and empirical facts. Thus, we would be allowed to assert that although it is true that mathematical propositions say nothing about empirical reality, they have an objective content, as they make assertions about relations between concepts. I have no room here to make remarks about the different arguments which follow in the manuscript, but I would like at least to say that the presentation of the "division" argument is still a little audacious, as becomes apparent as soon as we compare it with the way in which it appears in later manuscripts.

Actually, in the essay on Carnap, version II, Gödel says that although pure mathematics can be replaced by syntax under certain requirements, this is due to the fact that (i) pure mathematics implies nothing about the truth-value of those propositions which do not contain logical or mathematical symbols; (ii) mathematics follows from a finite number of axioms and formal rules which are known when the relevant language is built up. But such conditions, Gödel adds, can be also satisfied by some part of empirical science in relation to the rest of it. Thus, we are obviously facing an improved version of the "division" argument which goes as follows.

We could possess an additional sense, Gödel writes, which would show us a second reality, so separated from the space-time reality, that it would not be possible for us to draw any conclusion about the empirical facts, and so regular that it could be described through a finite number of laws. We could then arbitrarily recognize only the first reality, and say that the propositions concerning the second one are void of content, i.e., true by convention, and choose them with the aim to have them agree with those which were true according to the additional sense. Needless to say, the additional sense which Gödel is speaking about is precisely "reason", in the old rationalistic sense. Unfortunately, he does not devote more room to explain it or to reply to some possible objections against so audacious an argument. For instance, it could be said that we cannot even imagine an empirical sense which could be totally independent from the rest of senses, so I think the actual device is again based upon the mere analogy between mathematical intuition and sense perception.

In the end, all this leads Gödel to deny once again that the syntactic program may be equivalent to what we can reach by means of mathematical intuition, unless we admit we need a genuine mathematical fact, i.e., the consistency of the syntactic system used. However, this considerations are useful to us in realizing that what seems to underlie the new form of the analogy is precisely a further link with holism in Quine's sense, i.e., with the vision of human knowledge as proceeding from a unique conceptual scheme, where formal and empirical sciences are adapted to a unique pattern of working, and for which they should have a common foundation. I think this is the way we have to interpret Gödel's allusions to his theory of objective concepts, according to which that which defines a science is only the set of primitive concepts chosen, as well as the conceptual relationships which are determined by its axioms, but not the perceptive or intuitive support of this primary apparatus. Besides, if we remember that the acceptability of this apparatus depends upon its fruitfulness, and this upon its successful applicability, then the connection is confirmed.

The problem would then consist of explaining whether Gödel, in spite of the many times he resorts to the analogy, is able to convincingly avoid explicit holism, i.e., the thesis that since both sciences embody a common conceptual pattern, they differ from one another only gradually, as Carnap and Quine had argued before. I think that is why Gödel, in the same manuscript, accuses the syntactic conception of making it difficult to clearly distinguish between mathematics and natural science. The argument was that it is precisely this conception, which denies mathematical evidence and does not differentiate between causation and other constant connections, that makes it impossible to make a distinction between the two kinds of science. For, in so far as we take into consideration the verifiable consequences of theories, as Gödel adds, mathematical axioms are actually as necessary to obtain those consequences as are natural laws. Thus, he is able to insist that even from the syntactic view mathematical axioms can be seen as part of physical theory, which can be well defined only when they have been given, and therefore to conclude that such axioms are irreducible, and are hypotheses necessary for the scientific description of reality. This is very much in the Quinean style, but then Gödel did not succeed in overcoming the holistic conception of science.

To formulate it paradoxically: in so far as Gödel wishes to emphasize the objectivity of mathematics, he needs to significantly extend the analogy with natural science; but this leads him directly to holism, where the distinction works only in a pragmatic way, and where the expected conclusion is that the distinction is only a gradual one, so that we can protect from revision certain logical or mathematical statements only in a conventional way. On the contrary: in so far as Gödel wishes to defend the distinction precisely to escape from the serious holistic consequences of the non-distinction, he has to emphasize the importance of the pure relationships between primitive concepts. Yet not all primitive concepts are equally relevant for, as we have seen, every science has a set of them, but only those which are given to us through mathematical intuition. However, mathematical intuition is only explained in terms of the usual metaphorical parallelism with sense perception. This could perhaps explain why Gödel insisted again and again that mathematics is analytic: this could be the only way to escape from the conventionalist consequences implicit in his overall conception.

I think the root of the problem lies in the conception of science that Gödel was presupposing. According to this conception, the argument of the "empirical" application of set-theoretical axioms to arithmetic is holistic, for it really supposes not a class of pure facts (the numerical series in itself), but a whole theory (number theory). Thus, the success in the applications should be evaluated by taking into consideration not only the particular hypothesis we are trying to verify, but also other auxiliary hypotheses, or certain laws supposedly established. All that, as a whole, would make an actual theory possible. Therefore Gödel presupposes "the given" in a field even more difficult than natural science proper, where nobody admits it today in order to avoid to be accused of foundationalism. To sum up: Gödel seems to accept pure facts, so he seems to accept facts which are independent of theories, while some of his arguments can be holistically interpreted. Thus, his analogy between mathematics and empirical science seems to have a basic flaw: to depend upon a concept of science rather typical of certain conceptions of logical positivism.

# Bibliography


**A**spray, W. and Kitcher, P.

1988 (eds.) *History and philosophy of mathematics,* Minnesota Studies in the Philosophy of Science, vol. XI, Minneapolis: University of Minnesota Press.

Ayer, A.

1959 (ed.) *Logical positivism.* Glencoe, Ill.: Free Press.

**B**enacerraf, P.

1965 "What numbers could not be". In Benacerraf/Putnam 1983.

1973 "Mathematical truth". In Benacerraf/Putnam 1983.

Benacerraf, P. and Putnam, H.

1983 (eds.) *Philosophy of mathematics,* Cambridge: Cambridge University Press.

**C**arnap, R.

1931a "The old and the new logic". In Ayer 1959.

1931b "The logicist foundations of mathematics". In Benacerraf/Putnam 1983.

1937 *The logical syntax of language,* London: Routledge & Kegan Paul.

1942 *Introduction to semantics.* Cambridge, Mass.: Harvard Univ. Press.

1952 "Quine on analyticity". In Creath 1990.

1963 "Intellectual autobiography". In Schilpp 1963.

Cavaillès, J.

1938 *Métode axiomatique et formalisme.* Paris: Hermann, 1981.

Coffa, A.

1987 "Carnap, Tarski and the search for truth". *Noûs* 21, pp. 547–572.

1991 *The semantic tradition from Kant to Carnap. To the Vienna station.* Cambridge: Cambridge University Press.

Copi, I.

1949 "Modern logic and the synthetic *a priori*". *Jrn. Phil.* 46, pp. 243–5.

Creath, R.

1990 (ed.) *Dear Carnap, Dear Van. The Quine-Carnap correspondence and related work,* Berkeley: University of California Press.

Dawson, J. W.

1984a "The reception of Gödel's incompleteness theorems". *PSA 1984,* vol. 2. Also in Shanker 1988.

1984b "Cataloging the Gödel 'Nachlass'". *Philosophia Naturalis* 21, pp. 538–545.

1991 "*Nachlass* Inventory". Unpublished.

Detlefsen, M.

1987 *Hilbert's program.* Dordrecht: Reidel.

1990 "On an alleged refutation of Hilbert's program using Gödel's first incompleteness theorem". *Journal of Philosophical Logic* 19, pp. 343–77.

Dummett, M.

1963 "The philosophical significance of Gödel's theorem". *Ratio* 5, pp. 140–155. Also in 1978a.

1978a *Truth and other enigmas,* Londres: Duckworth.

1978b "Platonism". En 1978a.

1991 *Frege. Philosophy of mathematics.* Londres: Duckworth.

Feferman, S.

1984 "Kurt Gödel: conviction and caution". *Philosophia naturalis* 21, pp. 546–562. Also in Shanker 1988.

Frege, G.

1884 *Die Grundlagen der Arithmetik.* Breslau: Koebner. English trans. by J.L. Austin, Oxford: Blackwell, 1950; 2nd. rev. ed., 1953.

Friedman, M.

1988 "Logical truth and analyticity in Carnap's *Logical syntax of language*". In Aspray/Kitcher 1988.

Gödel, K.

1944 "Russell's mathematical logic". In Schilpp 1944, pp. 123–154 and CWII.

1947 "What is Cantor's continuum problem?". *Amer. Math. Monthly* 54, pp. 515–525. Rep. in CWII, together with the edition of 1964; also in Benacerraf/Putnam 1983.

1951 "Some basic theorems on the foundations of mathematics and their philosophical implications" (the "Gibbs lecture"). Unpublished manuscript reconstructed and edited in this volume.

1953–8 "Is mathematics syntax of language?". Unpublished manuscript, whose versions II and VI are reconstructed and edited in this volume.

1981 *Obras Completas*, Jesús Mosterín (ed.), Madrid: Alianza, 1981. 2nd. rev. ed., 1987.

1986 *Collected Works. Volume I. Publications 1929–1936* (CWI), S. Feferman *et al.* (eds.), New York: Oxford University Press.

1990 *Collected Works. Volume II. Publications 1938–1974* (CWII), S. Feferman *et al.* (eds.), New York: Oxford University Press.

Grattan-Guinness, I.

1977 *Dear Russell – Dear Jourdain*, London: Duckworth.

1979 "In memoriam Kurt Gödel: his 1931 correspondence with Zermelo on his incompletability theorem". *Hist. Math.* 6, pp. 294–304.

Haaparanta, L. and Hintikka, J.

1986 (eds.) *Frege synthesized*, Dordrecht: Reidel.

Hahn, L.E. and Schilpp, P.A.

1986 (eds.) *The philosophy of W.V. Quine*, La Salle, Ill.: Open Court.

Henze, H. W.

"Concierto de violín n° 2", based on the text of the poem "Hommage à Gödel", by H. M. Enzensberger (see Schimanovich *et al.* 1993). CD, DECCA 430 347–2 DM.

Hilbert, D.

1927 "The foundations of mathematics". In van Heijenoort 1967.

Hofstadter, D. R.

1979 *Gödel, Escher, and Bach: an eternal golden braid.* New York: Basic Books.

Irvine, A.D.

1989 "Epistemic logicism and Russell's regressive method". *Synthese* 55, pp. 303–327.

**K**itcher, P.
1986 "Frege, Dedekind and the philosophy of mathematics". In Haaparanta e Hintikka 1986.
Köhler, E.
1989 "Gödel and Carnap in Vienna". Paper read at the First Kurt Gödel Symposium, Salzburg. Extended version in Köhler 1991.
1991 "Gödel und der Wiener Kreis". In Kruntorad 1991.
1993 "Gödels platonismus". In Schimanovich *et al.* 1993.
Körner, S.
1965 "On the relevance of post-Gödelian mathematics to philosophy". *Proc. of the Intern. Coll. in the Phil. of Sci.* (Londres). Amsterdam: North-Holland, 1967.
Kotarbinski, T.
1955 "The fundamental ideas of pansomatism". *Mind* 64, pp. 488–500.
Kruntorad, P. von
1991 (ed.) *Jour fixe der Vernunft. Der Wiener Kreis und die Folgen.* Vienna: Hödel-Pichler-Tempsky.
**L**akatos, I.
1967 "A renaissance of empiricism in the recent philosophy of mathematics". In *Philosophical Papers,* vol. 2, Cambridge University Press, 1978.
Lorenzo, J. de
1979 "Lógica y matemática en K. Gödel". *Estudios Filosóficos,* vol. 28, núm. 79, pp. 391–453.
**M**yhill, J.
1952 "Some philosophical implications of mathematical logic". *Rev. of Metaphysics* 6, pp.165–198.
**N**agel, E. and Newman, J. R.
1958 *Gödel's Proof,* New York: New York University Press.
**Q**uine, W. V.
1934 "Lectures on Carnap". In Creath 1990.
1936 "Truth by convention". Rep. in Benacerraf/Putnam 1983.
1951 "Two dogmas of empiricism". Rep. in Quine 1953.
1953 *From a logical point of view.* Cambridge, Mass.: Harvard University Press.

1963 "Carnap and logical truth". In Schilpp 1963; rep. in Benacerraf/Putnam 1983.

1986 "Reply to Geoffrey Hellman". In Hanhn/Schilpp 1986.

**R**amsey, F.P.

1931 *The foundations of mathematics and other logical essays.* London: Kegan Paul.

Rodríguez-Consuegra, F.

1990 "La primera filosofía de Moore, I". *Agora* 9, pp. 153–170. Part II in vol. 10 (1991), pp. 91–105.

1991a *The mathematical philosophy of Bertrand Russell: origins and development,* Basel, Boston and Berlín: Birkhäuser. Rep. in 1993.

1991b "Gödel's unpublished objections against the linguistic accounts of mathematics". In Z.W. Wolkowski (ed.), *First international symposium on Gödel's theorems.* Singapur: World Scientific, 1993, pp. 174–194.

1991c "Bradley y la filosofía análitica, I: significado, lógica y ontología". *Análisis filosófico* 11, pp. 153–175. Part II in 12 (1992), pp. 51–78.

1991d "Números, objetos y estructuras". *Crítica* 68, pp. 7–86.

1991e "La filosofía de la matemática inédita de Gödel". In Garma, S. *et al.* (eds.), *Contra los titanes de la rutina. Encuentro en Madrid de investigadores Hispano-franceses sobre la historia y la filosofía de la matemática.* Comunidad de Madrid/ CSIC, 1994, pp. 447–458.

1991f "La reducción ontológica y sus problemas". *Crítica* 69 (1992), pp. 17–64.

1992a "Un inédito de Gödel contra el convencionalismo: historia, análisis, contexto y traducción". *Arbor,* CXLII, 558-559-560, pp. 323–348.

1992b "Gödel's first works, 1929–1936: mathematics without philosophy?". *Modern Logic,* 3, pp. 58–74.

1992c "El convencionalismo y la analogía entre matemática y física". *Mathesis,* 8, pp. 441–458.

1992d "Carnap, Quine, Gödel y la distinción analítico-sintético". *Mathesis,* 9 (1993), pp. 241–260.

1992e "Russell, Gödel and logicism". In J. Zcermak (ed.), *Philosophy of mathematics,* Proceedings of the

Fifteenth International Wittgenstein Symposium (part 1), Vienna: Hölder-Pichler-Tempsky, 1993, pp. 233–242.

1992f "Matemática y filosofía: el caso de Kurt Gödel". In C. Martín Vide (ed.), *Lenguajes naturales y leguajes formales, VIII.* Barcelona: PPU, pp. 119–130.

1992g "Some issues on Gödel's unpublished philosophical manuscripts". Paper read in the Vienna University, invited by the Kurt Gödel Society. *Collegium Logicum.* Annals of the Kurt-Gödel-Society, vol. 1, Wien, New York: Springer-Verlag, 1995, pp. 8–25.

1993a "Gödel's last works, 1938–1974: the emerging philosophy". *Modern Logic,* 4 (1994), pp. 318–27.

Russell, B.

1903 *The principles of mathematics,* Cambridge: Cambridge University Press.

1906 "On 'insolubilia' and their solution by symbolic logic". Rep. in Russell 1973.

1907 "The regressive method of discovering the premises of mathematics". Rep. in Russell 1973.

1910 (with A.N. Whitehead) *Principia mathematica,* vol. I. Cambridge: Cambridge University Press, vols. II and III in 1912 and 1913. 2nd. ed., 1925.

1911a "Sur les axiomes de l'infini et du transfini". English trans. in Grattan-Guinness 1977.

1911b "The philosophical implications of mathematical logic". Rep. in Russell 1973.

1912 *The problems of philosophy.* Londres: William & Norgate; rep. in Oxford University Press, 1980.

1919 *Introduction to mathematical philosophy,* London: Allen & Unwin.

1924 "Logical atomism". Rep. in Russell 1956.

1956 *Logic and knowledge,* R.C. Marsh (ed.), Londons: Allen & Unwin.

1973 *Essays in analysis,* D. Lackey (ed.), London: Allen & Unwin.

Sacristán, M.

1964 *Introducción a la lógica y al análisis formal,* Barcelona: Ariel.

Schilpp, P. A.
1963 (ed.) *The philosophy of Rudolf Carnap*, La Salle, Ill.: Open Court.
Schimanovich, W., Buldt, B., Köhler, E. and Weibel, P.
1993 (eds.) *Wahrheit und Beweisbarkeit. Leben und Werk Kurt Gödels*, Vienna: Hölder-Pichler-Tempsky.
Shanker, S. G.
1988 (ed.) *Gödel's theorem in focus*, London: Croom Helm.
Skolem, T.
1922 "Some remarks on axiomatized set theory". In van Heijenoort 1967.
Taylor, G.G.
1981 "The analytic and the synthetic in Russell's philosophy of mathematics". *Synthese* 39, pp. 51–59.
van Heijenoort, J.
1967 (ed.) *From Frege to Gödel: a source book in mathematical logic, 1879–1931*. Cambridge, Mass.: Harvard University Press.
Wang, Hao
1974 *From mathematics to philosophy*, New York: Humanities Press.
1985 *Beyond Analytic Philosophy*, Cambridge, Mass.: The MIT Press.
1987 *Reflections on Kurt Gödel*, Cambridge, Mass.: The MIT Press.
1991 "To and from philosophy – Discussions with Gödel and Wittgenstein". *Synthese* 88, pp. 229–277.
199? *Conversations with Gödel*, unpublished.
Weibel, P. and Schimanovich, W.
1986 "Kurt Gödel – A mathematical mystery". NTSC videotape, English version, 80 minutes.
Weyl, H.
1927 "Comments on Hilbert's second lecture on the foundations of mathematics". In van Heijenoort 1967.
1949 *Philosophy of mathematics and natural science*, Princeton, N.J.: Princeton University Press.
White, M.
1950 "The analytic and the synthetic: an untenable dualism".

In S. Hook (ed.), *John Dewey: Philosopher of Science and Freedom,* New York: Dial Press.
1987 "A philosophical letter of Alfred Tarski". *Jrn. Phil.* 84, pp. 28–32.

Wittgenstein, L.
1922 *Tractatus logico-philosophicus.* English trans., London: Routledge & Kegan Paul. Rep. with new translation in 1961.

**Y**ourgrau, P.
1991 *The disappearance of time. Kurt Gödel and the idealistic tradition in philosophy,* Cambridge: Cambridge University Press.

# Part II

# The unpublished essays

Zwangvolle Plage!
Müh ohne Zweck!
Mime

# 1 The character and origin of the manuscripts in the present edition

In the following I try to provide the reader with all the details needed to insert the Gödel essays appearing here in the broader context of the rest of his unpublished work, as well as everything about my personal work on them. I start with a short description of what can be found in the Gödel *Nachlass* in Princeton, USA, by referring to the catalogue prepared by John Dawson. Then I attempt to justify the particular selection I made of the manuscripts which I finally decided to study, reconstruct and publish. Such an explanation seems to be advisable given the great amount of Gödel's material unpublished but still extant. Also, I describe some of the historical details relevant to understanding Gödel's reason for writing the manuscripts, and also for withholding them from publication. The rest of the chapter concerns my own work related to these essays. I begin by explaining the state in which I found the original manuscripts; then I give the criteria which I have applied to their reconstruction and editing.

## Gödel's Nachlass

Already during Gödel's lifetime certain rumours were heard that his personal archive constituted a very valuable scientific resource, both because of the materials which he had been writing but left unpublished, and because of certain technical works which had been informally circulating. This made some people think that Gödel was working much more than his publications allowed one to suppose. Moreover, some people rather close to him knew that Gödel often devoted himself to writing on philosophy and theology. After his death it became known that many of these materials were written in Gabelsberger shorthand, an already obsolete system which at present is understood by only very few people – among them by Cheryl Dawson, who is transcribing those materials.

Gödel even received a letter from the Library of the Congress, Washington, asking him to donate his archive to the Library, but he never replied, as was typical of his habits. (Yet, on the other hand, he often wrote replies which were never actually sent out.) The truth is that when Gödel died, in 1978, his archive was simply inherited by his wife Adele. This was a hardly fortunate fact, since she did not have the capacity to understand its real value. (For example, there is reason to believe that she probably destroyed some letters between Gödel and his mother, perhaps because of the hard feelings she still had from the time Gödel's family opposed their marriage.)

Fortunately, Adele soon came to understand the great value of the archive, as well as the importance of donating it to the Institute for Advanced Study, in Princeton. Thus, some time before she died, in 1981, the archive was deposited in that institution in which Gödel had worked since 1940, when he came definitively to the USA. The next step was the cataloguing of the documents, carried out by John Dawson, a logician and historian of mathematics, and former student of Solomon Feferman. From 1982 to 1984 the original sixty boxes of documents were catalogued and ordered into a number of folders. Finally, the Institute for Advanced Study deposited the *Nachlass* in the Firestone Library, Princeton University, where after some time it was opened to scholars for study. It is from this source that this book and the third and further volumes of Gödel's *Collected Works* have, and are being, prepared.

The content of the *Nachlass* is extensive and varied, and it is perfectly listed and classified in Dawson's "*Nachlass* Inventory" (1991), at the disposal of the scholars who arrive at the Firestone Library, or who ask him directly for a copy at the Pennsylvania State University at York. In its present state the documents are located in 12 series of folders, of which the most interesting are the ones devoted to scientific and personal correspondence (01), the working notebooks (03), and the drafts and offprints of lectures and papers (04). The rest contain virtually *everything*, from small pieces of paper containing notes concerning books to bills from the Viennese cafés in the 1930's, together with materials of academic records, photographs and medical documents (for instance a daily record of the amount of milk of magnesium Gödel consumed for over thirty years).

The personal and scientific correspondence is really extensive (comprising about 3,500 letters, most of them received) and particularly interesting, given the quality of many of his correspondents: Bernays, Cohen, Heyting, Kreisel, Menger, Morgenstern, von Neumann, and Hao Wang, as well as Church, A. Robinson, Schilpp, D. Scott, Takeuti, and van Heijenoort. The letters which Gödel wrote but he never actually sent out are very interesting as well. Among them, there are two, to Mr. Balas and Mr. Gradjean, which give us many details concerning Gödel's ideas which have not been found in other sources of information. The enormous task of reconstructing and publishing Gödel's correspondence is one of the main goals of the team of editors of the *Collected Works,* who at present are already working on the fourth volume, entirely devoted to some of this correspondence.

As for the drafts, most of them are of a technical nature, although it seems that they contain no essentially new results in any of the areas in which Gödel had been working. However, it is to be noted that he often obtained certain rather minor results which were later independently obtained and published by others, which can be explained by Gödel's preference to devote himself strictly to really fundamental work. Also, the notes written for lectures and courses are interesting, although they do not seem to contain anything not known by means of other sources. As for the material of a philosophical character, doubtless the star is the Gibbs lecture, the six versions of the Carnap essay, and the extended version of the paper on time and relativity. In addition, there is of course the great amount of material in Gabelsberger, among which a short paper on Kant and Husserl can be mentioned. The content of this material can be revealed by no other means than deciphering, although John and Cheryl Dawson told me in 1990 that, according to the material already decoded, they did not expect those materials to contain spectacular revelations on Gödel's already known philosophical views.

## The present selection

Among the several unpublished materials in Gödel's *Nachlass*, those which were destined to be published, whether as lectures or papers, are doubtless the most interesting ones. Gödel was really a perfectionist, so they were written – and sometimes revised – more accurately than the rest. Among them, the Gibbs lecture and the several versions of the Carnap essay are the ones with substantial philosophical content. In addition, there are many philosophical materials among the several hundreds of sheets written in Gabelsberger, but it seems that those materials are written in a more dogmatic style, namely, with less argumentation.

The Gibbs lecture is the most philosophical of all the lectures which have been preserved, and it is also very interesting because it was mainly devoted to examining the philosophical implications of Gödel's celebrated metamathematical results. The rest of lectures which are extant are much more technical, and are mainly devoted to spreading new results rather than to analyzing their philosophical consequences. Gödel himself was entirely aware of the philosophical importance of the text, which was read in Providence, Rhode Island, in 1951, as is shown by the fact that he was working on the lecture for one year. Also, he openly referred to the content of the lecture as being clearly philosophical in his conversations with Hao Wang and others. Yet he never showed the text of the lecture to anyone, probably because of the state of the manuscript, which was in need of revision and rewriting (which Gödel himself never found the time to do), and perhaps also because of his caution in communicating his philosophical ideas to others. Sometimes, during his conversations with Hao Wang, Gödel told him that he was going to show the manuscript to him, but this never took actually place. All these factors justify, I think, the inclusion of the Gibbs lecture in this volume.

Concerning the critical series of versions of the essay on Carnap's logical syntax of language, it doubtless constitutes an exception among the unpublished manuscripts, which are technical almost in their entirety (apart from the extended version of the Einstein essay and some of the Gabelsberger materials). It is genuinely philosophical, from beginning to end, and it was destined to develop the ideas which Gödel had already outlined in his published papers on Russell (1944) and Cantor (1947). The only places where some of these ideas were published, in a cautious and very fragmentary way, are the 1958 paper on finitism and the philosophical postscript to the 1964 edition of the Cantor paper. Therefore, it was clear to me from the beginning that *some* version of the Carnap essay had to be published here. The problem was, which one?

None of the six extant versions is dated, and there does not seem to be any indication in correspondence or conversations about their particular dates of composition, although generally on the title page we have the number of the version ("I Fassung", etc.). Also, we can divide with certainty the series into two sets, one containing the four first versions, the other the two last.

The fifth and sixth versions are the final result of Gödel's enormous effort to sum up the best arguments in the least controversial possible way, so a great part of the philosophical complexity and interest of the first versions is inevitably lost. However, apart from their interest as useful summaries, there is also the fact that they constitute the versions closest to being in publishable form, according to Gödel's standards, in spite of the fact that he finally decided against publication, probably because he found them insufficiently convincing. Therefore, it seemed to me that the sixth version – the last one – should be included here, together with a series of footnotes written by myself in which I offer to the reader a systematic comparison with the arguments contained in the fifth version, which is more philosophical and less reticent in tone. In this way I try to do justice to the final stages of Gödel's efforts, roughly spanning the years from 1953 to 1959, with intermediate periods devoted to other tasks.

As for the versions I–IV, I pondered between including II or III. Version I is mainly the handwritten version of II, which is mostly typewritten, and IV constitutes only a new starting point of a very different character, doubtless due to Gödel's dissatisfaction with his former attempts. Finally I opted for version II, which seemed to me to be the most finished one. Apart from a few finishing touches, it appeared almost ready for the print (although, of course, a fair copy would first have to be obtained from the mass of addenda, footnotes and interpolations). Also, it seemed to me that version III included a certain development of some of the points of II, but that the state of the manuscript was not so good (although it included more folios in type). On the other hand, some of its pages were missing, and probably they were added to IV. By taking into consideration all these factors, I finally decided to include version II here. Finally, it is also interesting the fact that II constitutes a version in which Gödel's increasing imposed self-censure had not yet begun, while it is possible to interpret the later evolution of the essay as an attempt to somehow "smooth" many of the often controversial arguments against Carnap's position and the positivist philosophy. This process finally transformed the one hundred folios of version I into the eight folios of VI.

Some time after having made these decisions, and even after having finished the reconstruction and editing of the manuscripts which appear here, I learned from Hao Wang that, among the unpublished materials to appear in the third volume of Gödel's *Collected Works*, the Gibbs lecture was to be included, as well as some version of the Carnap essay. However, as Hao Wang told me too, my reconstruction of the Gibbs lecture, which he compared with the reconstruction to appear in the official edition, is longer. This is probably due to the fact that my main criterion of reconstruction and transcription was to save as many fragments as it was possible, even in cases when it was difficult – or impossible – to determine precisely where these fragments were originally located, or even in cases they were finally crossed out by Gödel. Later, I learnt that the versions chosen by the editors were III and V. Thus, I think that this edition and the third volume of the *Collected Works* can be regarded as complementary. At any rate, I am convinced that it is good that more and more materials written by such a genius are published.

## The origin of the present manuscripts

Concerning the details of the Gibbs lecture, there is not much to say, and is mostly due to the information given to me by Hao Wang and John Dawson, both personally and in correspondence. In particular, it is not known when or by what means Gödel received the original invitation. The lecture was read on December 26, 1951, at 8 pm, during the annual meeting of the American Mathematical Society, which took place at Brown University, Providence, Rhode Island. It seems that Gödel devoted most of the year to writing it, at a time when he was very interested in certain philosophical ideas once his public declaration of Platonism was made in the 1944 and 1947 papers. However, certain difficulties related to his weak state of health made the task more difficult. In particular, bleeding caused by a duodenal ulcer forced him to receive medical treatment in a hospital.

According to Hao Wang, who attended the lecture, Gödel limited himself to reading the manuscript – obviously the same text which appears here – very quickly, including the final quotation by Hermite. The audience was large, consisting mostly of mathematicians. At the end there were no questions (perhaps due to Gödel's directions), although an enthusiastic applause followed, which is very understandable given the rarity of the opportunity to see and hear so eminent a speaker.

There is no sign that Gödel revised the manuscript after reading it in Providence. Yet its appears in a list of writings he regarded as suitable for publication, which has been found among his documents. This is not strange given the fine quality of the essay, and the great ease with which Gödel was able to prepare it. However, the task of preparing and improving essays was precisely Gödel's big difficulty because of his almost pathological perfectionism, together with his great and almost paranoid fear of controversy. For example, in response to some people who were interested in its publication, Gödel said (around 1953–4) that he was trying to publish the lecture in the *Bulletin of the American Mathematical Society.* Yet it is virtually certain that not only did he never submit the paper to that journal, but that he never made any attempt in that direction. As for the content of the lecture, only certain allusions were made by Gödel in a few conversations and pieces of correspondence with Hao Wang and others, but it seems that Gödel never showed the paper to anybody.

Concerning the series of versions of the essay on Carnap, the story is considerably longer, although fortunately I have been able to reconstruct it in detail. In May, 1953, Paul Schilpp, the editor of the *The*

*Library of Living Philosophers* series, wrote to Gödel inviting him to contribute to a volume devoted to Carnap, which was then in preparation. Gödel had formerly accepted a similar invitation for the Russell volume, which appeared in 1944, although his contribution had some deadline problems. Also, Gödel was by then devoting nearly all of his time to philosophical studies at the Institute for Advanced Study, in Princeton, so the invitation was kindly accepted. Moreover, in 1951 he had devoted most of the Gibbs lecture to criticizing the philosophy of mathematics of logical positivism on the basis of his famous metamathematical results and his Platonist philosophy, so it can be said that he should have had the way well paved with only some improvement necessary to the main critical arguments.

Between 1953 and 1959 Gödel wrote up to six different versions of a manuscript entitled "Is mathematics syntax of language?", which, as we have seen above, constitute some of the more interesting philosophical materials in the *Nachlass*. Probably most of the work was carried out between 1953 and 1955 or 1956, when Gödel accepted another invitation to write a paper on the occasion of the 70 birthday of his friend Paul Bernays (finally published in 1958). Curiously enough, Carnap was also at the Institute for Advanced Study between 1952 and 1954, although it seems that no philosophical interaction between these old friends took place then, in spite of the fact that they had collaborated in the late 1920s, and especially in the early 1930s, after Gödel's spectacular results, when Carnap was working on the first versions of his *Logische Syntax der Sprache*.

After 1943 Gödel's philosophical interests were clearly increasing, and in the 1950s was reading and thinking about philosophy even more seriously, as can be seen in some of the philosophical notebooks written (in Gabelsberger's shorthand) of that time and by entries in library books at Princeton. Such writings were probably destined to face the general philosophical problems involved in his criticism of Carnap. It seems that among his favourite readings there were works by Leibniz and Husserl; probably he was in search for ideas which allowed him to make some advances in the problem of the nature of concepts, their combinations, and our knowledge of them.

The only explicit testimonies concerning the decision not to publish any of the versions proceed from Gödel's conversations with Hao

Wang in the 1970s. The problem is that, although Hao Wang has referred to the point several times in print, he hardly did so in a consistent way. In his 1986 (p. 19) he writes that Gödel told him several times that he did not publish the text because he did not have an adequate response to the question about the nature of mathematics, that is to say, to the positive features which characterize mathematics as a specific science different from other sciences. I think we should interpret this as meaning that it is very different to say what mathematics *is not*, as for instance when Gödel said that it is not a set of propositions without content and which are reducible to a few linguistic conventions, than saying what mathematics actually *is*. But later Hao Wang, although he insists on the same point (1987, pp. 23, 28), he also writes that Gödel did not publish the paper because he believed that it would be possible to give a more convincing response to Carnap (p. 46). However, nothing about Gödel's particular dissatisfaction with the work can be found in Wang's book, except a reference, in passing, when he writes (p. 119) that Gödel's kind of precision, compared to Carnap's, was so different that he was probably unable to find out a common language satisfactory enough to allow him to present his arguments in a entirely convincing way.

On the other hand, Wang also referred to Gödel's three alleged reasons for not publishing any of the six versions, which appear in his letter to Schilpp of January 1959. According to Wang, the reasons were that: (i) Gödel was unsatisfied; (ii) the final manuscript was overly critical; (iii) Carnap was already unable to reply. I think that even a superficial comparison between the different versions of the manuscript shows that, although Gödel tried again and again to improve his general position, he succeeded only in eliminating a series of substantial arguments which, in spite of their great interest, present many openings for frontal attack. Thus, his habit in mathematics to present his arguments in a concise, precise and definitive way, led him in philosophy to disaster, for in philosophy (fortunately) there are no absolutely conclusive arguments, in the sense of leaving the interlocutor without any possible reply.

Therefore, an analysis of Gödel's letter to Schilpp of 1959 is interesting, for it may add new relevant elements to Wang's treatment of the problem. These elements seem to me to point out to another explanation, which we can sum up this way: Gödel was simply very afraid of provoking strong adverse criticism. The text of this letter, which can be found in Gödel's *Nachlass*, is as follows:

> I am extremely sorry I cannot give an affirmative answer to your inquiry of Jan. 24. In view of the fact that my article would severely criticize some of Carnap's statements, it does not seem fair to publish it without a reply by Carnap. Nor would this be conducive to an elucidation of the situation.
>
> However, I feel I owe you an explanation why I did not send my paper earlier. The fact is that I have completed several different versions, but none of them satisfies me. It is easy to allege very weighty and striking arguments in favor of my views, but a complete elucidation of the situation turned out to be more difficult than I had anticipated, doubtless in consequence of the fact that the subject matter is closely related to, and in part identical with, one of the basic problems of philosophy, namely the question of the objective reality of concepts and their relations. On the other hand, in view of widely held prejudices, it may do more harm than good to publish half done work.
>
> I hope that in view of the reasons stated, and also in view of the fact that I was considerably hampered in my work by illness and other difficulties, you will kindly excuse it that I could not carry out my original intention.

The circumstances were, therefore, similar to the ones surrounding Gödel's contribution to Russell's 1944 volume in the same series. Then, Gödel finally did send the paper, which was the result of several earlier versions, although so late that Russell was unable to reply. Schilpp could have sent a former version to Russell for him to prepare the reply, but Gödel had given Schilpp permission to send Russell only the final version, as can be seen by reading the unpublished correspondence between Schilpp and Gödel. This is another instance demonstrating how Gödel often felt uncomfortable in waiting for the reactions of others to his writings, especially when there were of a philosophical character.

There are two more similarities. Once again, Gödel had problems in deciding about an absolutely *final* version, which was probably due to his constant obsession with making more and more improvements. This resulted in several letters by Schilpp in which he referred to successively later deadlines. And also, again, Gödel refers to his own state of health, which seems to me to proceed rather from a hypochondriac personality, especially as he was working on his contribution for a very long time.

As for the "widely held prejudices", there is no doubt that Gödel is referring to nominalism and positivism. Clear evidence of this can be seen in the difficulties Benacerraf and Putnam had in obtaining Gödel's permission for reprinting his essay on Cantor's continuum problem in the first edition of their celebrated anthology (1964); that is, only four or five years later. According to Moore's account (CWII, p.166), "Gödel hesitated to grant permission, fearing that the introduction to their book would subject his article to positivistic attacks", and he finally gave permission only when Benacerraf assured him that the editors wanted only to outline the essay, and that at any rate he would be able to see the introduction before it was printed.

Against such a background it is not strange that Gödel's philosophical stage, the longest period of his whole lifetime (from 1943 to 1978), was so unfruitful in terms of publications: the Russell paper, the philosophical fragments contained in the Cantor paper, some of the ideas from the articles on Einstein and Bernays, and some other reprints were the only results. This is not completely consistent with Feferman's explanation of Gödel's Platonism as a reaction to nominalism. Feferman's thesis is that, before the 1940s, "Gödel was understandably cautious about making public his platonistic ideas, contrary as they were to the 'dominant philosophical prejudices' of the time. With his reputation solidly established and with the security provided by the Institute, Gödel felt freer to pursue and publicly elaborate his philosophical vision" (CWI, p. 34; see also Feferman 1984). And yet, had Gödel really "felt freer", he should have been able to face any criticism *after* his appointment as full professor in the Institute for Advanced Study, which took place on July 1, 1953, where he would have been completely safe from any kind of "dangerous" consequences, either actual or simply imagined.

It is true that Gödel did not accept the Schilpp invitation (of May 15, 1953) until July 2, namely, just *one day after* his appointment, but at any rate the professional security was never enough for him to face what perhaps had implied a series of replies and counter-replies (probably written by others) in the best journals of the field. I think that such a fear of criticism was due more to the probable personal disturbances caused by the expected reaction than to the "technical" difficulty in replying to actual criticisms, all the more so for the expectation which doubtless had caused any philosophical publication written by such a genius, who was really much more inclined to the usually quiet and predictable way of life at the Institute.

At any rate, it is also true that the whole question seems to be related to the unsatisfactory nature of Gödel's general philosophical position at the time, which, according to Wang 1987, remained unsettled despite many efforts in search for a definitive clarification. As we have seen, when we analyze Gödel's efforts in these manuscripts, and when we see them as a whole, numerous difficulties come to light, and the main ones are precisely related to what he calls in the letter above "one of the basic problems of philosophy", namely, the problem of the objectivity of concepts and the truth of the corresponding propositions, particularly as we try to know them by means of some kind of immediate intuition.

## The original manuscripts: their reconstruction and editing

The Gibbs lecture is doubtless what John Dawson described to me in a letter as "an editor's nightmare". The original manuscript, without a title and in Gödel's hand, was written in English in pencil, and with unequivocal signs that had been erased again and again, then written again in the same places. Besides, Gödel often changed his mind regarding the actual text which should be read in Princeton, so there are many fragments which were crossed out. One very undesirable consequence of the script having been written in pencil – a very soft one – is that many parts of the text have, over the years, become extremely fuzzy, often simply unreadable. Also, I have to mention what seems to have been Gödel's obsession to save paper, which can be seen in the exhaustive use of every sheet, where there is no blank margin left on any side. In addition, many words are abbreviated; I do not know whether this, too, is due to his economy or a desire to write faster.

To make matters worse, the manuscript is composed of four parts: the main text of the lecture (number 040293 of catalogue; 43 folios); the footnotes (040295; 26 folios, and 040296; 5 folios); and the interpolations (040294, 18 folios). The problem with the interpolations is that it is not only necessary to insert them in the main text by means of a very complex system of keys, but also in the text of the footnotes, and even in the text of other interpolations. This leads to a system of cross references of a degree of complexity almost intolerable, where sometimes we have to deal with five or six levels at a time (e.g., interpolation to another interpolation of a footnote which has been divided into several fragments which belong to different folios of a passage belonging to the main text which has been fragmented too...), and where everything is often written, erased, rewritten, and with some fragments simply crossed out, while others are crossed out but with a note (which is often almost unreadable) saying that the original text really holds.

My main criterion in reconstructing the manuscript concealed within such a nightmare has been to preserve the greatest amount of material possible, even in cases when the original location of the reconstructed fragments could not be precisely ascertained. In such cases this material has been included in the final appendix, which contains also some loose footnotes and interpolations. In addition, I have sought to preserve the paragraphs, footnotes and interpolations which were finally crossed out, and to do so preferably in the original positions; these appear in the final edition between double square brackets ([[ ]]). This, however, has sometimes proven impossible, so there are a few crossed out fragments which do not appear in these texts. In cases where a safe reading has been impossible, or seemed at best very doubtful to me, I have proposed what I found to be the best possible reading. These passages are followed by a question mark between simple square brackets ([?]), indicating that my rendition is not absolutely certain. When it happens not only that the reading is doubtful, although complete, but that certain sentences or paragraphs become unreadable or doubtful in a particular place, then I offer the readable part, followed by suspension points and the corresponding question mark as in the the former case (... [?]). Finally, my own few interpolations always appear between simple square brackets, unless otherwise indicated.

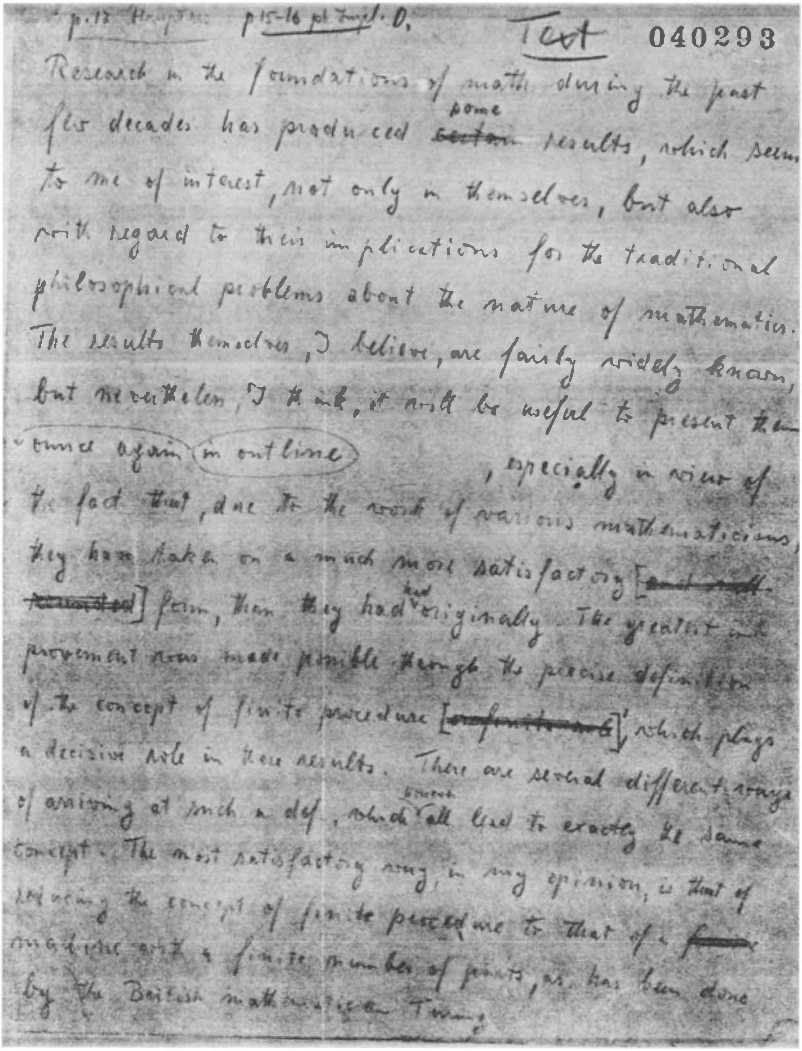
Concerning the final presentation, I have been forced to introduce new paragraphing. Except for the deletions and interpolations, Gödel's original script seems often to consist of *one enormous paragraph,* perhaps again with view to saving paper (which, by the way, was generally already recycled, often having been used before for other things, then erased), perhaps in order to obtain a manuscript consisting of only a few sheets, which was more convenient to handle in the moment of reading the actual lecture. Finally, I transcribe in italics, not only what has been originally underlined by Gödel, but also the titles of books and journals, as well as some letters used as symbols and a few Latin terms.

As for the version II of the Carnap essay, which was written originally in English as well, it is comparatively delightful, although the cross reference system is as complex as the one used in the Gibbs lecture. The original manuscript (catalogue number 040434) has no title, although it doubtless belongs to the series with a common title. Gödel tried originally to typewrite it entirely, but finally he was forced to introduce numerous handwriting interpolations, which were often written in separate folios. It includes 63 folios as a whole, from which 33 are devoted to the main text (among which there are 10 containing handwritten materials), and 33 more folios devoted to the footnotes (among which there are 10 additional ones of handwritten interpolations), plus 2 more folios containing attempts to state a correspondence between the several series of numbers corresponding to the footnotes. Such attempts have been useful to me when trying to reconstruct the original text, but the numbers of the footnotes which appear in this edition do not coincide with any of them. On the one hand, Gödel finally added new footnotes; on the other, I have preserved certain footnotes which Gödel himself finally crossed out (in which case they appear between double square brackets).

The main principle here has been the same as before, namely, to preserve the greatest amount of material. In this case the application of the criterion has led me, after much effort, to an almost complete reconstruction, that strives nonetheless to respect Gödel's original intentions concerning the "final" version. Yet I have included also the fragments which were ultimately crossed out, inserting them where they appeared originally (always between double square brackets). This should be useful in helping to understand Gödel's development when trying to express an idea or argument in alternative ways. I have not included discarded material when the old version differs from the new one only in an obvious stylistic way, or when, in two or three cases, the fragments are extremely short (two or three lines) and I have been unable to determine their original location. I felt that creating an appendix only for these insignificant fragments was unnecessary. The rest of the criteria coincide strictly with the ones which have been applied to the editing of the Gibbs lecture.

Concerning version VI of the Carnap essay, the original English text which has been transcribed here is the one which can be found under the catalogue number 040446. It appears to be a carbon copy of a typewritten manuscript of eight pages, numbered from 1 to 8. (The fact that the original version is not extant suggests that Gödel gave it to someone.) On the title page we read "VI Fassung", and I have used the "V Fassung" to introduce a systematic comparison in order to take advantage of the fact that the fifth version is more philosophical in character. There are only a few corrections and a few modifications in Gödel's hand. Finally, I have eliminated Gödel's projects for cross references, which were obviously incomplete ("see number ..."), in part because this could have introduced mistakes, in part because the manuscript is very short and such references can be easily imagined by the reader.

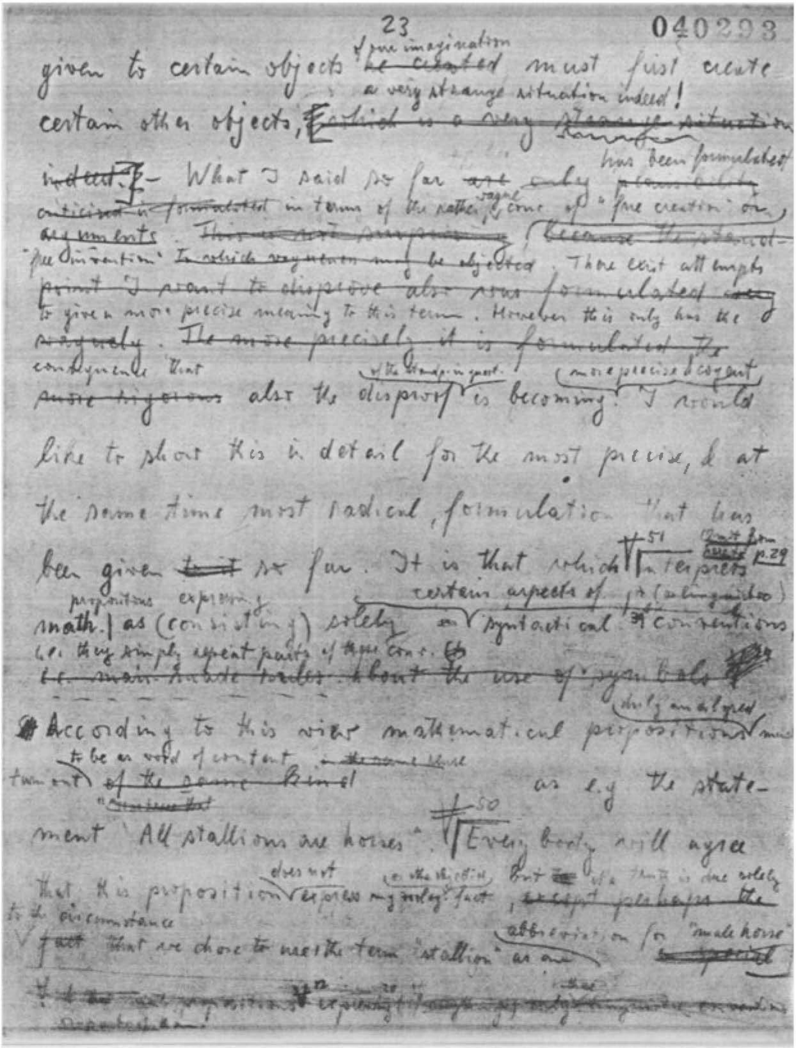
I have incorporated, without comment, all of Gödel's corrections and interpolations in order to keep the text free of footnotes (except where comparison is made between versions VI and V). The main criterion has always been to offer a text as uninterrupted as possible, with the only exception of including fragments which, while eventually crossed out, contribute significantly to the understanding of the final version. An alternative would have been to offer the textual variants in footnotes or in a final appendix, but I think the present system does not disrupt the reading. As for Gödel's English, no attempt has been made to improve it, even in the few cases where some sentences are not grammatical.



p. 18 Hempton p. 15-16 pt Engl. 0. Text 040293

Research in the foundations of math during the past few decades has produced ~~certain~~ some results, which seem to me of interest, not only in themselves, but also with regard to their implications for the traditional philosophical problems about the nature of mathematics. The results themselves, I believe, are fairly widely known, but nevertheless, I think, it will be useful to present them once again (in outline), especially in view of the fact that, due to the work of various mathematicians, they have taken on a much more satisfactory [~~and well rounded~~] form, than they had had originally. The greatest improvement was made possible through the precise definition of the concept of finite procedure [~~[illegible]~~], which plays a decisive role in these results. There are several different ways of arriving at such a def., which however all lead to exactly the same concept. The most satisfactory way, in my opinion, is that of reducing the concept of finite procedure to that of a ~~[illegible]~~ machine with a finite number of parts, as has been done by the British mathematician Turing.

1. Gibbs lecture. First folio of the main text.

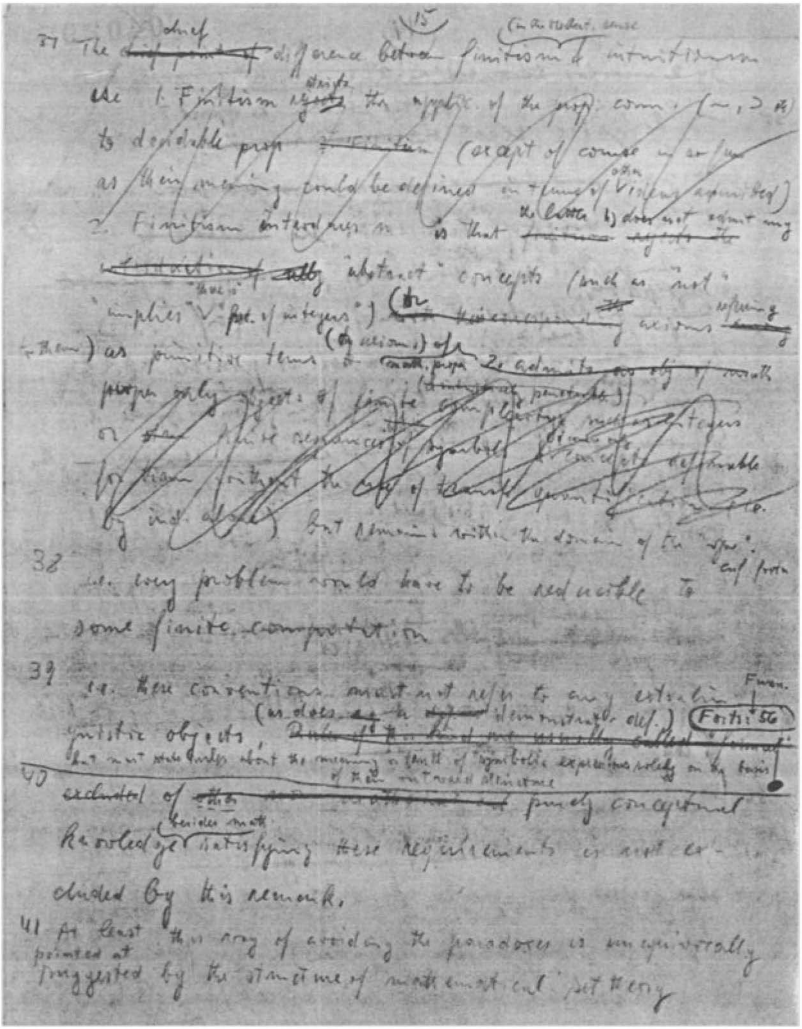


23 040293

given to certain objects of pure imagination must first create certain other objects, a very strange situation indeed! — What I said so far has been formulated only in terms of the rather vague concept of "free creation" or "free invention". There exist attempts to give a more precise meaning to this term. However this only has the consequence that the disproof of the standpoint is becoming more precise & cogent. I would like to show this in detail for the most precise, & at the same time most radical, formulation that has been given so far. It is that which interprets certain aspects of propositions expressing math. as (consisting) solely of syntactical (or linguistic) conventions, i.e. they simply repeat parts of these conventions.

According to this view mathematical propositions, duly analyzed, must turn out to be as void of content as e.g. the statement "All stallions are horses". Everybody will agree that this proposition does not express any zoological fact, but that its truth is due solely to the circumstance that we chose to use the term "stallion" as an abbreviation for "male horse".

2. Gibbs lecture. Folio 23 of the main text. It is to be noted the several keys, in the form of an inverted A plus a number, pointing to interpolations.



37 The chief difference between finitism & intuitionism (in the strictest sense) is 1. Finitism rejects the application of the prop. connectives (~, ⊃ etc) to decidable prop. [illegible] (except of course in so far as their meaning could be defined in terms of other [illegible]) 2. [illegible] is that the latter does not admit any "abstract" concepts (such as "not" "implies" "for all integers") (or axioms referring to them) as primitive terms [illegible] But remains within the domain of the [illegible]

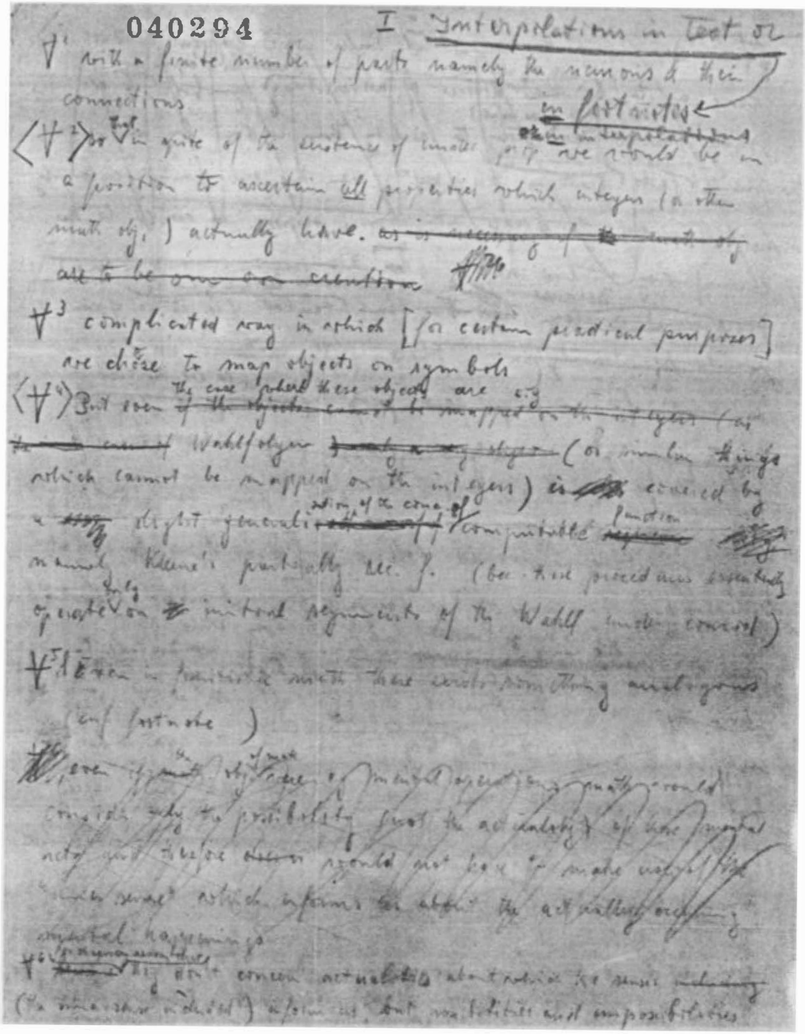
38 [illegible] every problem would have to be reducible to some finite computation

39 i.e. these conventions must not refer to any extralinguistic objects, (as does e.g. [illegible]) but must state rules about the meaning or truth of symbolic expressions solely on the basis of their outward structure

40 [illegible] purely conceptual knowledge (besides math) satisfying these requirements is not excluded by this remark.

41 At least this way of avoiding the paradoxes is unequivocally suggested by the structure of mathematical set theory

3. Gibbs lecture. Folio 15 of the footnotes. There are folios which are much more difficult to be read.



040294 I Interpolations in Text of

∀¹ with a finite number of parts namely the neurons & their connections — in footnotes

⟨∀²⟩ or ∀ in spite of the existence of innate [other interpolations] we would be in a position to ascertain all properties which integers (or other math obj.) actually have. ~~as is necessary of all to be own creation~~

∀³ complicated way in which [for certain practical purposes] we choose to map objects on symbols

⟨∀⁴⟩ But even the case where these objects are e.g. ~~of the objects cannot be mapped on the integers (as the case of~~ Wahlfolgen (or similar things which cannot be mapped on the integers) is covered by a slight generalization of the concept of computable function namely Kleene's partially rec. f. (bec. these procedures essentially operate only on initial segments of the Wahlf. under concept)

∀⁵ Even in praxis math there exists something analogous (cf footnote )

~~∀⁶ even if math objects were of mental operations math would consider only the possibility [illegible] the actuality of these mental acts and therefore would not have to make use of the "inner sense" which informs us about the actually occurring mental happenings~~

∀⁶ they don't concern actualities about which the senses inform us, but possibilities and impossibilities

4. Gibbs lecture. First folio of the interpolations.

2

# Some basic theorems on the foundations of mathematics and their philosophical implications

Research in the foundations of mathematics during the past few decades has produced some results, which seem to me of interest, not only in themselves, but also with regard to their implications for the traditional philosophical problems about the nature of mathematics. The results themselves, I believe, are fairly widely known, but nevertheless, I think, it will be useful to present them in outline once again, especially in view of the fact that, due to the work of various mathematicians, they have taken on a much more satisfactory form, than they had had originally. The greatest improvement was made possible through the precise definition of the concept of finite procedure, which plays a decisive role in these results. There are several different ways of arriving at such a definition, which however all lead to exactly the same concept. The most satisfactory way, in my opinion, is that of reducing the concept of finite procedure to that of a machine with a finite number of parts, as has been done by the British mathematician Turing. As for the philosophical consequences of the results under consideration, I don't think they have ever been adequately discussed or only taken notice of.

The metamathematical results I have in mind are all centered around, or, one may even say, are only different aspects of one basic fact, which might be called the incompletability or inexhaustibility of mathematics. This fact is encountered in its simplest form, when the axiomatic method is applied, not to some hypothetico-deductive system as geometry (where the mathematician can assert only the conditional truth of the theorems), but mathematics proper, i.e. to the body of those mathematical propositions, which hold in an absolute sense, without any further hypothesis. There must exist propositions of this kind, because otherwise there could not exist any hypothetical theorems either. E.g. *some* implications of the form: If such and such axioms are assumed then such and such a theorem holds, must necessarily be true in an absolute sense. Similarly any theorem of finitistic number theory such as 2 + 2 = 4 is, no doubt, of this kind. Of course the task of axiomatising mathematics proper differs from the usual conception of axiomatics in so far, as the axioms are not arbitrary, but must be correct mathematical propositions and moreover evident without proof, there is no escaping the necessity of assuming some axioms or rules of inference as evident without proof because the proofs must have some starting point. However there are widely divergent views as to the extension of mathematics proper, as I defined it. The intuitionists and finitists e.g. reject some of its axioms and concepts, which others acknowledge, such as the law of excluded middle or the general concept of set.

The phenomenon of the inexhaustibility of mathematics,[1] however, always is present in some form, no matter what standpoint is taken. So I might as well explain it for the simplest and most natural standpoint, which takes mathematics as it is, without curtailing it by any criticism. From this standpoint all of mathematics is reducible to abstract set theory. E.g. the statement that the axioms of projective geometry imply a certain theorem means, that if a set $M$ of elements called points and a set $N$ of subsets of $M$ called straight lines satisfies the axioms, then the theorem holds for $N$, $M$. Or, to mention another example, a theorem of number theory can be interpreted to be an assertion about finite sets. So the problem at stake is that of axiomatising set theory. Now, if one attacks this problem, the result is quite different from what one would have expected. Instead of ending up with a finite number of axioms, as in geometry, one is faced with an infinite series of axioms, which can be extended further and further, without any end being visible and, apparently, without any possibility of comprising all these axioms in a finite rule producing them.[2] This comes about through the circumstance that, if one wants to avoid the paradoxes of set theory without bringing in something entirely extraneous to actual mathematical procedure, the concept of set must be axiomatised in a step-wise manner.[3]

If e.g. we begin with the integers, i.e. the finite sets of a special kind, we have at first the sets of integers and the axioms referring to them (axioms of the I[rst] level), then the sets of sets of integers with their axioms (axioms of the second level), etc. for any finite iteration of the operation "set of".[4] Next we have the set of all these sets of finite order. But now we can deal with this set in exactly the same manner, as we dealt with the set of integers before, i.e. consider the subsets of it (i.e. the sets of order $\omega$) and formulate axioms about their existence. Evidently this procedure can be iterated beyond $\omega$, in fact up to any transfinite ordinal number. So it may be required as the next axiom, that the iteration is possible for *any* ordinal, i.e. for any order type belonging to some well ordered set. But are we at an end now? By no means, for we have now a new operation of forming sets, namely forming a set of some initial set *A* and some well ordered set *B* by applying the operation "set of" to *A* as many times as the well ordered set *B* indicates.[5] And, setting *B* equal to some well ordering of *A* now we can iterate this new operation, and again iterate it into the transfinite. This will give rise to a new operation again, which we can treat in the same way etc. So the next step will be, to require that *any* operation producing sets out of sets can be iterated up to any ordinal number (i.e. order type of a well-ordered set). But are we at an end now? No, because we can require, that not only the procedure just described can be carried out with any operation, but that moreover there should exist a set closed with respect to it, i.e. one which has the property that, if this procedure (with any operation) is applied to elements of this set, it again yields elements of this set.

You will realise, I think, that we are still not at an end, nor can there ever be an end to *this* procedure of forming the axioms, because the very formulation of the axioms up to a certain stage gives rise to the next axiom. It is true that in the mathematics of today the higher levels of this hierarchy are practically never used. It is safe to say that 99.9% of present day mathematics are contained in the first 3 levels of this hierarchy. So for all practical purposes all of mathematics *can* be reduced to a finite number of axioms. However this is a mere historical accident, which is of no importance for questions of principle. Moreover it is not altogether unlikely that this character of present day mathematics may have something to do with another character of it, namely its inability to prove certain fundamental theorems, such as e.g. Riemann's hypothesis, in spite of many years of effort. For, it can be shown that the axioms for sets of higher levels, in their relevance, are by no means confined to these sets, but, on the contrary, have consequences even for the 0-level, i.e. the theory of integers.

To be more exact, each of these set theoretical axioms entails the solution of certain Diophantine problems, which had been undecidable on the basis of the preceding axioms.[6] *The Diophantine problems in question are of the following type: Let* $P(x_1...x_n, y_1...y_m)$ *be a polynomial with given integral coefficients and* $n + m$ *variables* $x_1...x_n, y_1...y_m$ *and consider the variables* $x_i$ *as the unknown and the variables* $y_i$ *as parameters, then the problem is: Has the equation* $P = 0$ *integral solutions for any integral values of the parameters, or are there integral values of the parameters for which this equation has no integral solutions? To each of the set theoretical axioms a certain polynomial P can be assigned, for which the problem just formulated becomes decidable owing to this axiom. It even can always be achieved that the degree of P is not higher than 4.* Mathematics of today has not yet learned to make use of the set theoretical axioms for the solution of number theoretical problems, except for the axioms of the first level. These are actually used in analytic number theory. But for mastering number theory this is demonstrably insufficient. Some kind of set theoretical, number theory, still to be discovered, would certainly reach much farther.

I have tried so far to explain the fact I call incompletability of mathematics for one particular approach to the foundations of mathematics, namely axiomatics of set theory. That however this fact is entirely independent of the particular approach and standpoint chosen appears from certain very general theorems. The first of these theorems simply states that, *whatever well defined system of axioms and rules of inference may be chosen, there always exist Diophantine problems of the type described*[7] *which are undecidable by these axioms and rules, provided only that no false propositions of this type are derivable.* If I speak of a well defined system of axioms and rules here, this only means that it must be possible actually to write the axioms down in some precise formalism or, if their number is infinite, a finite procedure for writing them down one after the other must be given. Likewise the rules of inference are to be such that, given any premisses, either the conclusions by any one of the rules of inference can be written down, or it can be ascertained that there exists no immediate conclusion by the rule of inference under consideration. This requirement for the rules and axioms is equivalent to the requirement that it should be possible to build a finite machine in the precise sense of a "Turing machine" which will write down all the consequences of the axioms one after the other. For this reason the theorem under consideration is equivalent to the fact that there exists no finite procedure for the systematic decision of all Diophantine problems of the type specified.

The second theorem has to do with the concept of freedom from contradiction. For a well defined system of axioms and rules the question of their consistency is, of course, itself a well defined mathematical question. Moreover, since the symbols and propositions of one formalism are always at most enumerable, everything can be mapped on the integers and it is plausible and in fact demonstrable that the question of consistency can always be transformed into a number theoretical question (to be more exact into one of the type described above). Now the theorem says, that *for any well defined system of axioms and rules in particular the proposition stating their consistency*[8] *(or rather the equivalent number theoretical proposition) is undemonstrable from these axioms and rules, provided these axioms and rules are consistent and suffice to derive a certain portion*[9] *of the finitistic arithmetic of integers.* It is *this* theorem which makes the incompletability of mathematics particularly evident. For, *it makes it impossible that someone should set up a certain well defined system of axioms and rules and consistently make the following assertion about it: All of the axioms and rules I perceive (with mathematical certitude) to be correct*[10] *and moreover I believe that they contain all of mathematics.* If someone makes such a statement he contradicts himself. For if he perceives the axioms under consideration to be correct, he also perceives (with the same certainty) that they are consistent. Hence he has a mathematical insight not derivable from his axioms. However, one has to be careful in order to understand clearly the meaning of this state of affairs. Does it mean that no well defined system of correct axioms can contain all of mathematics proper? It does, if by mathematics proper is understood the system of all true mathematical propositions, it does not however, if one understands by it the system of all demonstrable mathematical propositions.

I shall distinguish these two meanings of mathematics as mathematics in the objective and in the subjective sense. Evidently no well defined system of correct axioms can comprise all objective mathematics, since the proposition which states the consistency of the system is true, but not demonstrable in the system. However as to subjective mathematics, it is not precluded, that there should exist a finite rule producing all its evident axioms. However, if such a rule exists, we with our human understanding could certainly never know it to be such, i.e. we could never know with mathematical certainty that all propositions it produces are correct;[11] or in other terms we could perceive to be true only one proposition after the other for any finite number of them. The assertion however that they are all true could at most be known with empirical certainty on the basis of a sufficient number of instances or by other inductive inferences.[12] If it were so this would mean that the human mind (in the realm of pure mathematics) *is* equivalent to a finite machine, that however he is unable to understand completely[13] his own functioning. This inability to understand himself would then wrongly appear to him as its boundlessness or inexhaustibility. But, please, note that if it were so this would in no way derogate from the incompletability of objective mathematics. On the contrary, it would only make it particularly striking. For if the human mind were equivalent to a finite machine, then objective mathematics not only would be incompletable in the sense of not being contained in any well defined axiomatic system, but moreover there would exist *absolutely* Diophantine problems of the type described above, where the epithet "absolutely" means that they would be undecidable not just within some particular axiomatic system but by *any* mathematical proof the human mind can conceive.

So the following disjunctive conclusion is inevitable: *Either mathematics is incompletable in this sense that its evident axioms can never be comprised in a finite rule, i.e. to say the human mind (even within the realm of pure mathematics) infinitely surpasses the powers of any finite machine, or else there exist absolutely unsolvable Diophantine problems of the type specified* (where the case that both terms of the disjunction are true is not excluded so that there are, strictly speaking, three alternatives). It is this mathematically established fact which seems to me of great philosophical interest. Of course in this connection it is of great importance that at least this fact is entirely independent of the special standpoint taken toward the foundations of mathematics.[14] There is however one restriction to this independence, namely the standpoint taken must be liberal enough to admit propositions about all integers as meaningful. If someone were so strict a finitist that he would maintain that only particular propositions of the type $2 + 2 = 4$ belong to mathematics proper,[15] then the incompletability theorem would not apply. But I don't think that such an attitude could be maintained consistently, because it is by exactly the same kind of evidence that we judge that $2 + 2 = 4$ and that $a + b = b + a$ for any two integers $a, b$. Moreover this standpoint in order to be consistent would have to exclude also *concepts* that refer to *all* integers such as "+" (or to all formulas such as "correct proof by such and such rules") and replace them with others that apply only within some finite domain of integers (or formulas). It is to be noted however that, although the truth of the disjunctive theorem is independent of the standpoint taken, the question as to which alternative holds need not be independent of it.

I think I now have explained sufficiently the mathematical aspect of the situation and can turn to the philosophical implications. Of course, in consequence of the undeveloped state of philosophy in our days, you must not expect these inferences to be drawn with mathematical rigour.

Corresponding to the disjunctive form of the main theorem about the incompletability of mathematics the philosophical implications *prima facie* will be disjunctive too, however under either alternative they are very decidedly opposed to materialistic philosophy. Namely if the first alternative holds, this seems to imply that the working of the human mind cannot be reduced to the working of the brain, which to all appearances is a finite machine with a finite number of parts namely the neurons and their connections. So apparently one is driven to take some vitalistic viewpoint. On the other hand the second alternative, where there exist absolutely undecidable mathematical propositions, seems to disprove the view, that mathematics (in any sense) is only our own creation. For the creator necessarily knows all properties of his creatures, because they can't have any others except those he has given to them. So this alternative seems to imply that mathematical objects and facts or at least *something* in them exist objectively and independently of our mental acts and decisions, i.e. to say some form or other of Platonism or "Realism" as to the mathematical objects.[16] For, the empirical interpretation of mathematics,[17] i.e., the view that mathematical facts are a special kind of physical or psychological facts, is too absurd to be maintained (see below).

[[Of course in these brief formulations I have oversimplified matters. There are in both cases certain objections which however, in my opinion, do not withstand a thorough examination. In case of the first alternative one may object that the fact that the human mind is more effective than any finite machine does not necessarily imply that some non-materialistic entity such as an entelechy exists besides the brains, but only that the laws governing the behaviour of living matter are much more complicated than had been anticipated, and in particular do not allow one to deduce the behaviour of the whole from the behaviour of the isolated parts.[18] (This view incidentally seems to be supported also by quantum mechanics where the state of a compound system in general cannot be described as composed of the states of the partial systems.) There actually exists a school of psychologists, which defends this view, namely the so called wholists. However it seems clear to me that this theory in effect also abandons materialism, because it ascribes from the beginning to matter all the mysterious properties of mind and life, whereas originally it was the very essence of materialism to explain these properties from the structure of the organism and the relatively simple laws of interaction of the parts.]]

It is not known whether the first alternative holds, but at any rate it is in good agreement with the opinions of some of the leading men in brain and nerve physiology, who very decidedly deny the possibility of a purely mechanistic explanation of psychical and neurons processes. As far as the second alternative is concerned, one might object that the constructor need not necessarily know *every* property of what he constructs. E.g. we build machines and still cannot predict their behaviour in every detail. But this objection is very poor. For we don't create the machines out of nothing, but build them out of some given material. If the situation were similar in mathematics, then this material or basis for our constructions would be something objective and would force some realistic viewpoint upon us even if certain other ingredients of mathematics were our own creation. The same would be true if in our creations we were *to use* some instrument in us but different from our ego (such as "reason" interpreted as something like a thinking machine). For mathematical facts would then (at least in part) express properties of this instrument, which would have an objective existence.

One may thirdly object, that the meaning of a proposition about all integers, since it is impossible to verify it for all integers one by one, can consist only in the existence of a general proof. Therefore, in the case of an undecidable proposition about all integers neither itself not its negation is true, hence neither expresses an objectively existing but unknown property of the integers. I am not in a position now to discuss the epistemological question as to whether this opinion is at all consistent. It certainly looks as if one must *first* understand the meaning of a proposition, *before* he can understand a proof of it so that the meaning of "all" could not be defined in terms of the meaning of "proof". But independently of this epistemological investigations I wish to point out that one may conjecture the truth of a ... [?] proposition (e.g. that I shall be able to verify a certain property for *any* integer given to me) and, at the same time conjecture that no general proof for this fact exists. It is easy to imagine situations in which both these conjectures would be very well founded. For the first half of it this would ... [?] be the case if the proposition in question were some equation $F(n) = G(n)$ of two number theoretical functions which could be verified up to very great numbers $n$.[19]

Moreover, exactly as in the natural sciences, this *inductio per enumerationem simplicem* by no means is the only, inductive method conceivable in mathematics. I admit that every mathematician has an inborn abhorrence to giving more than heuristic significance to such inductive arguments. I think however that this is due to the very prejudice that mathematical objects somehow have no real existence. If mathematics describes an objective world just like physics, there is no reason why inductive methods should not be applied in mathematics just the same as in physics. The fact is that in mathematics we still have the same attitude today that in former times one had toward all science, namely we try to derive everything by cogent proofs from the definitions (i.e. in ontological terminology from the essences of things). Perhaps this method, if it claims monopoly, is as wrong in mathematics as it was in physics. It is true that only the second alternative points in this direction. This whole consideration incidentally shows that the philosophical implications of the mathematical facts explained do not be entirely on the side of rationalistic or idealistic philosophy, but that in one respect they favor the empiristic viewpoint.[20]

However, *and this is the item I would like to discuss now,* it seems to me that the philosophical conclusions drawn under the second alternative, in particular concerning Realism (Platonism), are supported by modern developments in the foundations of mathematics also irrespectively of which alternative holds. The main argument pointing in this direction seems to me the following. First of all, if mathematics were our free creation, ignorance as to the objects we created, it is true, might still occur, but only through lack of a clear realisation as to what we really have created (or ... [?] due to the practical difficulty of too complicated computations). Therefore it would have to disappear (at least in principle although perhaps not in practice[21]) as soon as we attain perfect clearness. However modern developments in the foundations of mathematics have accomplished an insurmountable degree of exactness, but this has helped practically nothing for the solution of mathematical problems.

Secondly the activity of the mathematician shows very little of the freedom a creator should enjoy. Even if e.g. the axioms about integers were a free invention, still it must be admitted that the mathematician, after he has imagined the first few properties of his objects, is at an end with his creative ability, and he is not in a position also to create the validity of the theorems at his will. If anything like creation exists at all in mathematics, then what any theorem does is exactly to restrict the freedom of creation. That however which restricts it must evidently exist independently of the creation.[22]

Thirdly: If mathematical objects are our creation, then evidently integers and sets of integers will have to be two different creations.

The first of which does not necessitate the second. However, in order to prove certain propositions about integers the concept of set of integer is necessary. So here, in order to find out what properties *we* have given to certain objects of pure imagination must be first create certain other objects, a very strange situation indeed!

What I said so far has been formulated in terms of the rather vague concept of "free creation" or "free invention". There exist attempts to give more precise meaning to this term. However this only has the consequence that also the disproof of the standpoint in question is becoming more precise and cogent. I would like to show this in detail for the most precise, and at the same time most radical, formulation that has been given so far. It is that which asserts mathematical propositions to be true solely due to certain arbitrary rules about the use of symbols.

[Gödel's note:] Omit from here to p. 29 [p. 143 of this edition]

[[[It is that which] interprets mathematical propositions as expressing solely certain aspects of syntactical (or linguistic)[23] conventions, i.e. they simply repeat parts of these conventions. According to this view mathematical propositions duly analysed must turn out to be void of content as e.g. the statement "All stallions are horses".

Everybody will agree that this proposition does not express any zoological or other objective fact, but its truth is due solely to the circumstance that we chose to use the term "stallion" as an abbreviation for "male horse", since the simplest rules about the use of symbols are definitions. Now by far the most common type of symbolic conventions are definitions (either explicit or contextual where the latter however must be such as to make it possible to eliminate the term defined in any context it occurs). Therefore the simplest version of the view in question would consist in the assertion that mathematical propositions are true solely owing to the definitions of the terms occurring in them, i.e. that by successively replacing all terms by their *definientia*, any theorem can be reduced to $a = a$ (note that $a = a$ must be admitted as true if definitions are admitted, for one may define $b$ by $b = a$ and then owing to this definition replace $b$ by $a$ in this equality). [An alternative version deleted after: "any theorem can be reduced to":] [[an explicit tautology, such as $a = a$ or $p \supset p$ or $pq \supset p$ or something like it; (it is immaterial in *this* connection what is considered to be an explicit tautology, except that, in order to justify the term "explicit", it must be possible and even easy to find out about a given proposition whether or not it is an *explicit* tautology).]]

But now it follows directly from the theorems mentioned before that such a reduction to explicit tautologies is impossible. For it would immediately yield a mechanical procedure for deciding about the truth or falsehood of every mathematical proposition. Such a procedure however cannot exist, not even for number theory. The disproof, it is true, refers only to the simplest version of this (nominalistic) standpoint. But the more refined ones do not fare any better. The weakest statement that at least would have to be demonstrable in order that the view concerning the tautological character of mathematics be tenable, is the following: Every demonstrable mathematical proposition can be deduced from the semantical rules about the truth and falsehood of sentences alone (i.e. without using or knowing anything else except these rules),[24] and that the negations of demonstrable mathematical propositions cannot be so derived (cnf. Footnote 23). (In precisely formulated languages such rules – i.e. rules which stipulate under which conditions a given sentence is true – occur as a means for determining the meaning of sentences. Moreover in all known languages there *are* propositions which seem to be true owing to these rules alone.) E.g. of disjunction and negation are introduced by the rules: 1.) $p \vee q$ is true, if at least one of its terms is true, and 2.) $\sim p$ is true if $p$ is not true. Then it clearly follows from these rules that $p \vee \sim p$ is always true whatever $p$ may be. (Propositions so derivable are called tautologies.).

Now it is actually so, that for the symbolisms of mathematical logic, with suitable chosen semantical rules, the truth of the mathematical axioms *is* derivable from these rules;[25] however (and this is the great stumbling block) in this derivation the mathematical and logical concepts and axioms themselves must be used in a special application, namely as referring to symbols, combinations of symbols, sets of combinations, etc. Hence, this theory if it wants to prove the tautological character of the mathematical axioms, must first assume these axioms to be true. So while the original idea of this viewpoint was to make the truth of the mathematical axioms understandable by showing that they are tautologies, it ends up with just the opposite, i.e. the truth of the axioms must *first* be assumed and *then* it can be shown that, in a suitably chosen language, they are tautologies [[That this can be done is of course not surprising. It could be done for any axioms whatsoever.[26]]]. Moreover a similar statement holds good for the mathematical concepts, i.e.: Instead of being able to define their meaning by means of syntactical conventions one must first know their meaning in order to understand the syntactical conventions in question or the proof that they imply the mathematical axioms but not their negation.

Now of course it is clear that the elaboration of the nominalistic view does not satisfy the requirement set up on p. [?] because not the syntactical rules alone but all of mathematics in addition is used in the derivations. But moreover this elaboration of nominalism would yield an outright disproof of it (I must confess I can't picture any better *disproof* of this view than this proof of it), provided that one thing could be added, namely that the outcome described is unavoidable (i.e. independent of the particular symbolic language and interpretation of mathematics chosen). Now it is not exactly this that can be proved but something so close to it that it also suffices to disprove the view in question. This however can be done, namely it follows that a proof for the tautological character (in a suitable language) of the mathematical axioms is at the same time a proof for their consistency, and therefore by the metatheorems mentioned cannot be achieved with any weaker means of proof than are contained in these axioms themselves. This does not mean that all the axioms of a given system must be used in its consistency proof. On the contrary usually the axioms lying outside the system which are necessary make it possible to dispense with some of the axioms of the system (although they do not imply these latter).

However what follows with practical certainty is this: In order to prove the consistency of classical number theory (and *a fortiori* of all stronger systems) certain abstract concepts (and the directly evident axioms referring to them) must be used where "abstract" means concepts which do not refer to sense objects,[27] of which symbols are a special kind. These abstract concepts however are certainly not syntactical (but rather those whose justification by syntactical considerations should be the main task of nominalism). Hence it follows that *there exists no rational justification of our precritical beliefs concerning the applicability and consistency of classical mathematics (nor even it's undermost level, number theory) on the basis of a syntactical interpretation.* It is true that this statement does not apply to certain subsystems of classical mathematics which may even contain some *part* of the theory of the abstract concepts referred to. In this sense nominalism can point to some partial successes. For it is actually possible to base the axioms of these systems on purely syntactical considerations [[(without any reference using any "abstract" concepts).]]. In this manner the use of the concepts of "all" and "there is" referring to integers can be justified (i.e. proved consistent) by means of syntactical considerations. However for the most essential number theoretical axiom, complete induction, such a syntactical foundation, even within the limits in which it is possible, gives no justification of our precritical belief in it, since this axiom itself has to be used in the syntactical considerations.[28]

The fact that, the more modest you are in the axioms for which you want to set up a tautological interpretation, the less of mathematics you need in order to do it has the consequence that if finally you become so modest as to confine yourself to some finite domain, e.g. to the integers up to 1000, then the mathematical propositions valid in this field can be so interpreted as to be tautological even in the strictest sense, i.e. reducible to explicit tautologies by means of the explicit definitions of the terms. No wonder because the section of mathematics necessary for the proof of the consistency of this finite mathematics is contained already in the theory of the finite combinatorial processes which are necessary in order to reduce a formula to an explicit tautology by substitutions. This explains the well known, but misleading, fact that formulas like $5 + 7 = 12$ can, by means of certain definitions, be reduced to explicit tautologies. This fact, incidentally, is misleading also for this reason that in these reductions (if they are to be interpreted as simple substitutions of the *definiens* for the *definiendum* on the basis of explicit definitions) the + is not identical with the ordinary + because it can be defined only for a finite number of arguments (by enumeration of this finite number of cases). (If on the other hand + is defined contextually then one has to use the concept of finite manifold already in the proof of 2 + 2 = 4.) A similar circularity [[similar to that I have just pointed out in the reduction of 5 + 7 = 12 to an explicit identity[29]]] also occurs in the proof that $p \vee \sim p$ is a tautology, because disjunction and negation in their intuitive meaning evidently occur in it.]]

[Gödel's note, after these omitted pages:] Begin

[[My considerations about Platonism so far have been chiefly *apagoge* i.e. I have tried to disprove the opposite view in its various forms. In conclusion of this lecture I would like to describe positively in some more detail the view about the nature of mathematics to which in my opinion one is driven by modern developments in the foundations. I think this can be best done ... [?] the view I was criticizing.]]

The essence of this view is that there exists no such thing as a mathematical fact, that the truth of propositions by which we believe to express mathematical facts only means that (due to the rather complicated rules which define the meaning of propositions i.e. which determine under what circumstances a given proposition is true) an idle running of language occurs in these propositions in that the said rules make them true no matter what the facts ... [?]. Such propositions can rightly be called void of content. Now it is actually possible to build up a language in which mathematical propositions are void of content in this sense. The trouble only is 1. that one has to use the very same mathematical facts (or equally complicated mathematical facts) in order to show that they don't exist, 2. that by this method if a division of the empirical facts in 2 parts *A, B* is given, such that *B* implies nothing in *A*, a language can be constructed in which the propositions expressing *B* would be void of content. And if your opponent were to say: You are arbitrary disregarding certain observable facts *B*, one may answer you are doing the same thing e.g. with the law of complete induction which I perceive to be true on the basis of my understanding (i.e. perception) of the concept of integer. Moreover it is easily seen that for any division of the *empirical* facts in two classes *A, B* such that the facts of *B* imply nothing about those of *A*, using the facts of *B* one could construct a language in which the propositions expressing the facts of *B* would be "void of content" and true solely due to semantical rules.

However, it seems to me that nevertheless one ingredient of this wrong theory of mathematical truth is perfectly correct and really discloses the true nature of mathematics. Namely it is correct that a mathematical proposition says nothing about the physical or psychical existing in space and time, because it is true already owing to the meaning of the terms occurring in it, irrespectively of the world of real things. What is wrong however is, that the meaning of the terms (i.e. the concepts they denote) is asserted to be something man-made and consisting merely in semantical conventions. The truth I believe is that these concepts form an objective reality of their own, which we cannot create or change, but only perceive and describe.[30] Therefore a mathematical proposition, although it does not say anything about space-time reality, still may have a very sound objective content, in so far as it says something about relations of concepts. The existence of non "tautological" relations between the concepts of mathematics, appears [[not so much in the trivial fact that necessarily certain primitive, i.e. indefinable ideas, must be assumed both for mathematics and syntax, but]] above all in the circumstance, that for the primitive terms of mathematical axioms must be assumed, which are by no means tautologies, in the sense of being in any way reducible to $a = a$, but still do follow from the meaning of the primitive terms under consideration.

E.g. the basic axiom or rather axiom-schema for the concept of set of integers says that, given a well defined property of integers (i.e. a propositional expression $\varphi(n)$ with an integer-variable $n$), there exists the set $M$ of those integers which have the property $\varphi$). Now considering the circumstance that $\varphi$ may itself contain the term "set of integers", we have here a series of rather involved axioms about the concept of set. Nevertheless these axioms (as the aforementioned results show) cannot be reduced to anything substantially simpler, let alone to explicit tautologies. It is true that these axioms are valid owing to the meaning of the term "set", one might even say they express the very meaning of the term set, and therefore they might fittingly be called analytic; however the term tautological, i.e. devoid of content, is for them entirely out of place, because even the assertion of the existence of a concept of set satisfying these axioms (or of the consistency of these axioms) is so far from being empty that it cannot be perceived [?] without again using the concept of set itself or some other abstract concepts of similar nature.

Of course this particular argument is addressed only to mathematicians who admit the general concept of set in mathematics proper. For finitists, however, literally the same argument could be alleged for the concept of integer and the axiom of complete induction. For, if the general concept of set is *not* admitted in mathematics proper, then complete induction must be assumed as an axiom. [[I don't think it can be objected to this viewpoint, concerning the analyticity of mathematics, that an undecidable mathematical proposition, whose truth could be recognized at most with probability, cannot be analytic. For I am using the term]] I wish to repeat that analytic [[not in]] here does not mean [[the subjectivistic sense of]] "true owing to our definitions", but rather [[in the objectivistic sense of]] "true owing to the nature of the concepts occurring"; in contradistinction to [[synthetic, which would mean]] "true owing to the properties and the behaviour of things."

This concept of analytic is so far from meaning "void of content" that it is a perfectly possible that an analytic proposition might be undecidable (or decidable only with probability). For our knowledge of the world of concepts may be as limited and incomplete as that about the world of things. It is certain and undeniable that this knowledge (in certain cases) not only is incomplete, but even indistinct. This occurs in the paradoxes of set theory, which are frequently alleged as a disproof of Platonism, but, I think, quite unjustly. Our visual perceptions sometimes contradict our tactile perceptions, e.g. in the case of a rod immersed in water, but nobody in his right mind will conclude from this fact that the outer world does not exist.

Of course I do not claim that the foregoing considerations amount to a real proof of this view about the nature of mathematics. The most I could assert would be to have disproved the nominalistic view, which considers mathematics to consist solely in syntactical conventions and their consequences. Moreover I have adduced some strong arguments against the more general view that mathematics is our own creation. There are however other alternatives to

Platonism, in particular Psychologism and Aristotelian realism. In order to establish Platonistic realism, these theories have to be disproved one after the other, and then it would have to be shown that they exhaust all possibilities. I am not in a position to do this [[conclusively]] now; however, I would like to give some indications along these lines.

One possible form of psychologism admits that mathematics investigates relations of concepts and that concepts cannot be created at our will, but are given to us as a reality, which we cannot change, however it contends that these concepts are only psychological [[structures or]] dispositions [[in our minds]], i.e. that they are nothing, but so to speak the wheels of our our thinking machine. To be more exact a concept would consists in the disposition 1. to have a certain mental experience, when we think of it and 2. to pass certain judgements (or have certain experiences of direct knowledge) about its relations to other concepts and to empirical objects. The essence of this psychologistic view is that the object of mathematics is nothing but the psychological laws by which thoughts, convictions, etc. occur in us, in the same sense as the object of another part of psychology is the laws by which emotions occur in us. The chief objection to this view I can see at the present moment is that if it were correct, we would have no mathematical knowledge whatsoever. We would not know e.g. that $2 + 2 = 4$, but only that our mind is so constituted as to hold this to be true and there would then be no reason whatsoever why, by some other train of thought, we should not arrive at the opposite conclusion with the same degree of certainty. Hence, whoever assumes that there is some domain, however small, of *mathematical* propositions which we *know* to be true, cannot accept this view.

[[Another form of psychologism says that, not the mathematical concepts, but the objects to which they refer, are something purely subjective or mental e.g. operations of the mind, such as going over to the next integer in counting. If under this view it is maintained that the propositions about these mental entities are analytic (in whatever sense of this term), then he [[he also is a Platonist[31]]] must affirm that our knowledge of analytic prop. is confined to propositions referring to mental phenomena which [[if one once accepts Platonism]] seems quite unnatural and unacceptable to me. If, on the other hand, it is maintained that the propositions about these mental entities are synthetic, it is hard to see how any universal mathematical proposition can be known, except by inductive generalization.[32]

As to the view corresponding to Aristotelian realism [[(which asserts the concepts to be parts or "aspects" of space-time things) it seems to me it will hardly be able to give a satisfactory account of concepts of higher level than the first (and all mathematical concepts are such)]] it will hardly be maintained that the objects of mathematics are single objects in nature (such as heaps of pebbles). If however the objects in nature with which mathematics deals are assumed to be qualities (and relations) then one is faced with all difficulties connected with the Aristotelian view that qualities and relations are (abstract) parts of the things. In particular the transitivity of the relation of part seems to imply that qualities of qualities are qualities of the things. Moreover it is very hard to think of all possible worlds as parts of the real world. I have not yet clarified every aspect of these questions to my own satisfaction. All these of course are rather loose considerations.]]

I am under the impression that after sufficient clarification of the concepts in question it will be possible to conduct these discussions with mathematical rigour and that the result then will be that (under certain assumptions which can hardly be denied – in particular the assumption that there exists at all some thing like mathematical knowledge) the Platonistic view is the only one tenable. Thereby I mean the view that mathematics describes a non-sensual reality, which exists independently both of the acts and the dispositions of the human mind and is only perceived, and probably perceived very incompletely, by the human mind. This view is rather unpopular among mathematicians, there exist however some great mathematicians who have adhered to it. E.g. Hermite once wrote the following sentence:

> Il existe, si je ne me trompe, tout un monde qui est l'ensemble des verités mathematiques, dans lequel nous n'avons accès que par l'intelligence, comme existe le monde des reálités physiques; l'un et l'autre indepéndant de nous, tous deux de création divine.[33]

# Gödel's footnotes

1 This concept for the applications to be considered in this lecture is equivalent to the concept of a "computable function of integers" (i.e. one whose definition makes it possible actually to compute $f(n)$ for each integer $n$ to be considered). The procedures do not operate on integers but on formulas, but because of these ... [?] of the formulas in question they can always be reduced to procedures operating on integers.

2 In the axiomatisations of non-mathematical disciplines such as physical geometry what I call mathematics proper is presupposed; and the axiomatisation refers to the content of the discipline under consideration only in so far as it goes beyond mathematics proper. ... [?]

3 This circumstance in the usual presentation of the axioms is not directly apparent but shows itself on closer examination of the meaning of the axioms.

4 The operation "set of" is substantially the same as the operation "Power set" where the power set of $M$ is by definition the set of all subsets of $M$.

5 In order to carry out the iteration one may put $A = B$ and assume that a special well ordering has been assigned to any set.

6 This theorem in order to hold also if the intuitionistic or finitistic standpoint is assumed requires as a hypothesis the consistency of the axioms of set theory which of course is self-evident (and therefore can be dropped as a hypothesis) if set theory is considered to be mathematics proper. However for finitistic mathematics a similar theorem without any questionable hypothesis of consistency holds.

7 This last hypothesis can be replaced by consistency (as shown by Rosser in ["Extensions of some theorems of Gödel and Church", *Jrn. Symb. Logic* I, pp. 87–91]) but the undecidable propositions then have a slightly more complicated structure. Moreover the hypothesis must be added that the axioms imply the primitive propositions ... [?] addition and multiplication and <.

8 It is one of the propositions which are undecidable provided that no false number theoretical [propositions] are derivable (cnf. the precedent theorem).

9 Namely Peano's axioms plus the rule of definition by ordinary induction with a ... [?] satisfying the strictest finitistic requirements.

10 If he only says "I believe I shall be able to perceive one after the other to be true" (where their number is supposed to be infinite) he does not contradict himself (see below).

11 For this (or the consequence concerning the consistency of the axioms) would constitute a mathematical insight not derivable from the axiom [and?] rule under consideration contrary to the assumption.

12 E.g. it is conceivable (although far outside the limits of present day science) that brain physiology would advance so far that it would be known with empirical certainty 1. that the brain suffices for the explanation of all mental phenomena and is a machine in the sense of Turing; 2. that such and such is the precise mathematical structure and physiological functioning of the part of the brain which performs mathematical thinking. Furthermore in case the finitistic (or intuitionistic) standpoint is taken such an inductive inference might be based on a (more or less empirical) belief that non finitistic (or non intuitionistic) mathematics is consistent.

13 Of course the physical working of the thinking mechanism could very well be completely understandable. The insight that this

particular mechanism must always lead to correct (or only consistent) results would surpass the powers of human reason.

14 For Intuitionists and Finitists the theorem holds as an implication (instead of a disjunction). It is to be noticed that Intuitionists have always asserted the first term of the disjunction and negated the second term in the sense that no dem. [onstrably?] undecidable proposition can exist (cnf. above p. [?]). As for finitism it seems very likely that the first disjunctive term is false.

15 K. Menger's "... [?]" (cnf. *Blatter f. d. Phil.* 4 (1930) p. 323) if taken in the strictest sense would lead to such an attitude since according to it the only meaningful mathematical proposition (i.e. in any term the only ones belonging to mathematics proper) would be those that assert that such and such a conclusion can be drawn from such and such axioms and rules of inference in such and such manner. This however is a proposition of exactly the same logical character as 2 + 2 = 4. Some of the undesirable consequences of this standpoint are the following: A negative proposition to the effect that the conclusion *B* cannot be drawn from the axioms and rule *A* would not belong to mathematics proper. Hence nothing could be known about it except perhaps that it follows from certain other axioms and rules. However a proof that it does so follow (since these other axioms and rules again are arbitrary) would in no way exclude the possibility that (in spite of the formal proof to the contrary) a derivation of *B* from *A* might some day be accomplished. For the same reason also the usual inductive proof for $a + b = b + a$ would not exclude the possibility of discovering two integers not satisfying this equation.

16 There exists no term of sufficient generality to express exactly the conclusion drawn here which only says that the objects and theorems of mathematics are as objective and independent of our free choice and our creative acts as is the physical world. It determines however in no way what these objective entities are, whether they are located in nature or in the human mind or in neither of the two. These three views about the nature of math-

ematics correspond exactly to the three views about the nature of concepts which ... [?] by the names of psychologism, Aristotelian conceptualism, and Platonism.

17 I.e. the view that mathematical objects and the way in which we know them are not essentially different from physical or psychical objects and laws of nature. The true on the contrary is that if the objectivity of mathematics is assumed it follows at once that its objects must be totally different from sensual objects because they *can be* known (in principle) without using senses (i.e. by means of reason alone, for ... [?] they don't concern actualities about which the senses (the ... [?] sense included) inform us, but possibilities and impossibilities). The mathematical objects are general... [?]. Mathematics ...[?] assert nothing about the actualities of the space-time world. In physics e.g. nothing is known except by probability but nobody thinks of denying for this reason the status of an exact science to physics. That our attitude toward mathematics is different to my mind is [?]

18 [[The other possibility namely to ascribe "reason" already to the behaviour of the elementary parts (i.e. the neurons or all ... [?]) seems utterly unlikely (both in itself and in view of the success of physics in explaining the behaviour of non structured wholes in terms of "computable" laws).]]

19 Such a verification of an *equality* (not an inequality) between two number theoretical functions of not too complicated or artificial structure would certainly give a great probability to their complete equality although its numerical value could not be estimated in the present state of science. However it is easy to give examples of general propositions about integers where the probability can be estimated even now. E.g. the probability of the proposition which states that for each *n* there is at least one digit $\neq 0$ between the $n^{th}$ and $n^{2th}$ digit of the decimal expression of $\pi$ converges toward 1 as one goes on verifying it for greater and greater *n*.. A similar situation is also [?] for Goldbach and Fermat theorems.

20 To be more precise it suggests that the situation in mathematics is not so very different from that in the natural sciences. As to whether in the last analysis apriorism or empiricism is correct, is a different question.

21 I.e. every problem would have to be reducible to some finite computation.

22 It is of no avail to say that these restrictions are brought about by the requirement of consistency which itself is our free choice, because one might choose to bring about consistency *and* certain theorems. Nor does it help to say that the theorems only repeat (wholly or in part) the properties first invented because then the exact realization of what was first assumed would have to be sufficient for deciding any question of the theory which is disproved by the first and the third argument. As to the question of whether undecidable propositions can be decided arbitrarily by a new act of creation cnf. footn. [?].

23 [[I.e. these conventions must not refer to any extralinguistic objects (as does a demonstrat. [ive?] definition), but must state rules about the meaning or ... [?] of symbolic expressions solely on the basis of their outward structure. Moreover of course these rules must be such that they do not imply the truth or falsehood of any factual proposition (since in that case they could certainly not be called void of content nor syntactical). It is to be noted that if the term "syntactical rule" is understood in this generality the view under consideration includes as a special elaboration of it the formalistic foundation of mathematics. Since according to the latter mathematics is based solely on certain syntactical rules of the form: propositions of such and such structure are true (the axioms) and: if propositions of ... structure are true, then such and such other propositions are also true. And moreover the consistency proof, as can easily be seen, gives the consequence that these rules are void of content in so far as they imply no factual propositions. On the other hand also vice versa it will turn out below that the feasibility of the nominalistic program implies the feasibility of the for-

malistic program. It may be doubted whether this (nominalistic) view should at all be subsumed under the view considering mathematics to be a free creation of the mind, because it denies altogether the existence of mathematical objects. However the relatedness between the two is extremely close since also under the other view the so called existence of mathematical objects consists solely in their being constructed in thought and nominalists would not deny that we actually imagine (non existent) objects behind the mathematical symbols and that these subjective ideas might even furnish the guiding principle in the choice of the syntactical rules. For very lucid expositions of the philosophical aspect of this nominalistic view cnf. H. Hahn, *Act. Sci. et ind.* 226 (1935) or R. Carnap, *Act. Sci.* 291 (1935), *Erk.* 5 (1935) p. 30.

24 [[As to the requirement of consistency cnf. footnote [?] [[Otherwise of course the solution would be trivial. The requirement of consistency also follows directly from the concept of syntactical rule (as explained in footnote [?]) since an inconsistent system of syntactical rules would imply the truth of every factual proposition, while lack of content means that no factual proposition should follow and hence conflict with the criteria of truth following from dem. [onstrative?] definitions.

25 [[Cnf. Ramsey F. P. *Proc. Lond. Math. Soc.* II ser 25 (1926) p. 368 p. 382, Carnap R. *Log. Synt. of Lang.* 1937 p 39 and 110 and 182. It is worth mentioning that Ramsey even succeeds in reducing them to explicit tautologies $a = a$ by means of explicit definitions but at the expense of admitting propositions of infinite (and even transfinite) length which of course entails the necessity of presupposing transfinite set theory in order to be able deal with these infinite entities. Carnap confines himself to propositions of finite length but instead has to consider infinite sets, sets of sets, etc. of these finite propositions.

26 [[Let us assume e.g. someone had a sixth sense which would give him only a few perceptions and these in no causal connection with the perception of the other senses. Then he could incorpo-

rate these perceptions in a few syntactical rules which he could prove tautological (i.e. of no consequence for the other perceptions) using in this proof the perceived prop. [erties?] of the perceptions of his sixth sense. This simile in my opinion expresses very well both the relationship of reason to the senses and the truth value of the theories which undertake to prove reason to be tautological.

27 [[Examples for such abstract concepts are e.g. "set", "function of integers" or "demonstrable" (the latter in the non formalistic sense of "knowable to be true") or "derivable" etc. or finally "there is" referring to ... [?] *possible* combinations of symbols. The necessity of such concepts for the consistency proof of classical mathematics results from the fact that symbols can be mapped on the integers and therefore finitistic (and *a fortiori* classical) number theory contains all proofs based solely upon them. The evidence for this fact so far is not absolutely conclusive because the evident axioms referring to the non-abstract concepts under consideration have not been investigated thoroughly enough. However the fact itself is acknowledged even by leading formalists.

28 [[The objection raised here against a syntactical foundation of number theory is substantially the same which Poincaré leveled against both Frege's and Hilbert's foundation of number theory. However this objection is not justified against Frege because the logical concepts and axioms he has to presuppose do not explicitly contain the concept of "finite manifold" with its axioms while the grammatical concepts and considerations [?] necessary to set up the syntactical rules and establish their tautological character do.

29 [[This circularity does not imply that (as Poincaré was ... [?]) Frege's derivation of such equations from the logical or set theoretical axioms contains a vicious circle (cnf. footnote [?]). [[because for Frege, in contradistinction to nominalists, an inference is not a combinatorial operation on certain finite combinations of *symbols* (which involves the concept of a finite manifold) but an insight about the logical *concepts* occurring in it.]] ]]

30 This holds good also for those parts of mathematics which *can* be reduced to syntactical rules (cnf. above). For these rules are based on the idea of a finite manifold (namely of a finite sequence of symbols) and this idea and its properties entirely independent of our free choice. In fact its theory is equivalent to the theory of integers. The possibility of so constructing a language that this theory is incorporated into it in the form of syntactical rules proves nothing, cnf. Footn. [?].

31 [[As remarked in footn [?], the mere assumption that concepts are something objective (i.e. extramental) does not yet mean Platonistic realism but rather a disjunction of this view and Aristotelian conceptualism [[that concepts are elements (or "abstract parts") of the space time world which come to our knowledge by applying the analysing (or abstracting) faculty of our mind to the material furnished by the senses]]. However, under this theory no other a priori propositions about concepts seem to be possible except those that state relations of part and whole between these constituents i.e. such as can be reduced to explicit tautologies. Hence in consequence of the non-tautological nature of the mathematical axioms (see above) Aristotelian conceptualism [[seems to imply the synthetic nature of mathematics cannot be maintained]] is inapplicable to mathematics.

32 [[Kant has maintained the possibility of it owing to his "pure" intuition whose function is to present to us a totality of *single* objects (i.e. points, lines, etc.) in such manner that, in contradistinction to sense perception one can directly read the general propositions off this perception without any extrapolation or induction.... [?].]]

33 Cnf. G. Darboux, *Eloges academ. et discours*, 1912, p. 142. The passage quoted continues as follows: "qui ne semblent distincts qu'à cause de la faiblesse de notre esprit qui ne sont pour une pensée plus puissante qu'une seule et même chose et dont la synthèse se revèle partiellement dans cette merveilleuse correspondance entre la mathématique abstraite d'une part, l'Astronomie et

toutes les branches de la physique de l'autre." So here Hermite seems to turn toward Aristotelian realism. However he does so only figuratively since Platonism remains the only conception understandable for the human mind.

# Appendix

## Loose fragments and footnotes

What follows is a series of interpolations and footnotes which Gödel ultimately deleted from the original manuscript – including deleted (crossed out) fragments as usual – and whose position within the text cannot, at this stage, be determined with certainty. They are nonetheless interesting because of their content, and often can be understood independently, or, at any rate, can be related of to other ideas in the text.

## Interpolations

15 [[There is a milder (not quite so absurd) form of empiricism (advocated by Aristotle) according to which the concepts (i.e. properties) are parts of the things (not so very different from their spatial ...[?]) which come to our knowledge by the senses ...[?]. Mathematical or logical propositions however are not empirically true but only state this relation of part and whole.]]

17 One might however say that in order to carry through the nominalistic view a mathematical *proof* for this fact is not necessary but empirical evidence (obtained by trying out the consequences of the syntactical rules) is sufficient. [[In this restricted sense the nominalistic standpoint actually can be upheld (by taking as one of the semantical rules that everything derivable from the – arbitrarily chosen – mathematical axioms is to be true).]] But to this suggestion it must be objected that [[with this sense not only mathematics but every science (also physics) can be made tautological]] the very fact in question (or better the proposition expressing it) namely that the semantical rules imply no empirical propositions on the one hand is not empirical according to the nominalistic own interpretation of mathematical propositions (it

says nothing about the space time world ... [?]), on the other hand is not tautological because in that case it would have to be demonstrable by analysis of the content of the syntactical rules (while by the axioms under which we ... [?] indemonstrable). So the semantical standpoint (in this formulation) precisely presupposes one of the mathematical facts whose non-existence it wants to prove [?].

21 [[The reason why (in my opinion) the exclusion of empiricism, together with the objectivity of mathematics leads to something like Platonism is that we have the two categories "thing", "concept", both taken in the widest sense (i.e. actuality and possibility)... [?]]]

24 To be more exact it is asserted that the meaning of mathematical symbols is completely contained in the man-made rules governing their use and that mathematical theorems are those propositions which are true owing to the linguistic conventions concerning the use of the symbols occurring in them.

26 Secondly nominalists might say that under the assumption that mathematical objects and facts are free creations the existence of an undecidable proposition is something to be expected. It only means that by our creative acts we have not determined the objects in every respect and therefore have to supplement these acts by new ones determining e.g. whether *p* or not-*p* is true (in case *p* is undecidable). This argument stated generally sounds very convincing. However applied to certain ... [?] situation it turns out to be ... [?]. Namely since in particular the question of the consistency of the mathematical system created is one of the undecidable propositions of the system the argument here says that one may decide on the consistency of an arbitrary system by a new arbitrary assumption.

28 In direct contradiction to the nominalistic standpoint that wants to confine himself in its presuppositions to a definite (and very small) subsystem of mathematics namely the one dealing (in a

finitistic manner) with finite combinations of discrete objects (the symbols). Now this basis is demonstrably insufficient even for the consistency proof of number theory (because it is contained in mathematics). [Footnote:] The decisive point in this connection is that for this proposition (and *a fortiori* for all more far [?] reaching consistency proof) axioms about certain abstract concepts must be used (i.e. concepts that do not directly refer to sense objects such as symbols). Now it is the essence of nominalism that they do not accept such abstract concepts in themselves but only in so far as they can be interpreted in terms of symbols and sense-objects. But such an interpretation proves impossible except for a very small part of mathematics provided it is required of an interpretation that it should give a rational foundation for our precritical belief (which is the very purpose of any such interpretation.)

## Footnotes

11 This argument is not valid for finitists because this standpoint rejects explicitly any general concept of set or function of integers even in the restricted intuitionistic sense of constructible or computable function. However in finitistic mathematics a similar situation prevails in so far as in order to prove certain propositions about certain functions (such as + and .) other (recursively defined) functions (such as exponentiation) must be introduced and in finitistic mathematics the definitions by induction cannot be considered to be mere abbreviations but each of them constitutes an ... [?].

12 For in order to have a tautological interpretation of mathematics it must be required that it should follow from the semantical rules not only that the mathematical axioms are true but also that their negations are not true, or at least some similar ... [?] must be made. [[Because if the semantical rules concerning logical and mathematical concepts are to be nothing else but mere devices of associating new kinds of expressions in a more complicated but more useful way to reality than is done by stating the single em-

pirical facts by atomic propositions (such as "this is red") then these rules must certainly not allow one to deduce new atomic propositions as would be the case if they implied a contradiction.]] Because the axioms in question would certainly not be tautological (i.e. devoid of content) if they implied any empirical propositions as would be the case if they implied a contradiction, because then every empirical proposition whatsoever would follow. [[Also it is clear that if mathematics consists merely in linguistic conventions it must cannot imply any empirical proposition.]] Hence in order to prove the tautological character of the mathematical axioms it is not sufficient to show that in a suitably constructed language they follow from the syntactical rules, but moreover these rules must be proved to be such that they imply no empirical proposition such as "this is red".

14 Even in this case subjective mathematics also would be incompletable in the following sense: If some of the mathematical undecidable propositions were decided by probability arguments (see below) and adjoined as new axioms, other propositions of the same type would remain unsoluble so that the process of adjoining new axioms in this manner could never come to an end.

21 [[To assume that the concepts are more than just dispositions in our minds is not yet Platonism because the concepts might be considered to be something in the things or parts of the things (not much different from their spatial parts) which come to our knowledge by the senses (the outer or the inner sense). This (Aristotelian) theory however (not to mention the difficulties encountered in connection with relations), seems to entail the hardly tenable consequence (which is in evident contradiction with inner observation namely) that each concept (the primitives included) should have as many parts as there are different assertions about it and moreover also parts involving this concept itself.]]

22 This may be identified with Kant's view except that according to Kant the mental entities concerned are not operations but permanent structures in the mind, namely space and time, the indi-

vidual elements or other constituents of which (not the concepts referring to them) are perceived by "pure intuition". Moreover according to the *Kritik* of pure reason the mathematical concepts too are subjective, since they are obtained by applying the purely subjective categories of thinking to the objects of intuition. Not so according to Kant's earlier writing *"De mundi [sensibilis atque intelligibilis forma et principiis,* 1770]" where only the world of the senses (including its forms, space and time) is considered to be subjective phenomenon which abstract thinking conveys knowledge of the things in themselves. The writing quoted is interesting also for this reason that it avoids the faulty analogy: "arithmetics : time = geometry : space", but instead holds that the intuition of time gives rise to the science of kinematics, while the concept of number is considered to belong to the sphere of abstract thinking and to require pure intuition (of either time or space) only for its *"actuatio in concreto"* (cnf. §12 of the writing quoted).

23 The wording of Kant's definition of "analytic" given in *Proleg.*, §2a agrees better with this conception of analytic than the conception of tautological, since Kant defines a proposition to be analytic "if in the predicate it says nothing except that which in the concept of the subject had actually been thought although not so clear and consciously". In order to approximate the modern conception of tautology he would have had to require that the predicate should be contained in the *definition* of the subject. Evidently we think something under the indefinable concepts. Hence there should exist non-void analytic propositions also about them. But there can exist no tautologies about them (except explicit ones such as $a = a$), since they have no definition. The Kantian definition quoted fits literally to the two examples given in the text if the totality of sets of integers (or the totality of integers) is made the subject of the sentences expressing the axioms. All this however applies only to the wording of Kant's definition quoted while from other passages of his writings (cnf. in part. *Logik*, §36, 37) it clearly appears that the concept of "analytical" he really had in mind agrees in essence with the modern concept of "tautological".

26 [[One might ask: isn't it sufficient that the tautological character of mathematics follows from the truth of mathematics? For this seems [?] to mean that mathematics must either be rejected [or?] considered to be tautological. But to this objection two things are to be replied. 1. For those parts of mathematics which need abstracts concepts for their consistency proof (see footnote [?]) the condition stated on p. [?] (which is necessary in order that the semantical system [?] be tenable) is *not* satisfied at all, since the truth of mathematical axioms does *not* follow the semantical rules *alone*, but only from these rules plus certain properties of abstract concepts which have nothing to do with semantics (on the contrary, if the semantical view were correct it is precisely these concepts which in the first place would have to be reduced to semantics. But if one tries to do this then the semantical substitutes always fall short of the abstract concepts they are to represent and necessarily so according to what has been explained).

2. As for subsystems of number theory (cnf. footnote [?]) it is to be noted that it is not the tautological character which can be proved but only the existence of a tautological interpretation which does not preclude the existence of other interpretations. Hence another necessary requirement for the semantical view in this case would be that the tautological interpretation at least is sufficient for all purposes. But this precisely is not the case since for the very setting up of the tautological interpretation the intuitive interpretation (referred to the symbols) is presupposed and the tautological interpretation is by no means only a precization [?] of the intuitive one, because although the two agree with each other extensionally, i.e. make the same propositions true still in the latter we doubtless imagine mathematics to be something as objective as physics.]]

One might ask: isn't it sufficient at least [?] a refutation of realism that the tautological character of mathematics can be concluded from mathematics itself? For this inference although not binding for nominalists who have to leave the validity of mathematics in abeyance until they succeed to derive it on the basis of their philosophical presuppositions will have to be acknowledged at least by Realists and hence implicate them into self contradiction. This conclusion would be correct if "tautology" in this connection really meant "void of content". However what (by definition) it means is [[two things. 1. That mathematics in a suitable language follows from the syntactical rules of this language and 2. that these rules have no consequences in the realm of space time reality. But the second circumstance means voidness of content only if (by use of a *petitio principii*) fact is identified with empirical fact. But the first (if "syntactical rule" is understood in the generality explained in footnote [?]) can easily be accomplished for any theory (or system of propositions) which is known (or assumed) to be true at the time the language is defined (except that if there are too many independent assertions in this theory their incorporation as syntactical rules would make the language intolerably complicated)]] that *there exists a language* in which mathematics is void of content *in so far* as follows from the rules of syntax.

This however amounts to very little since the division of the true propositions into those that are expressed by syntactical rules and those that are arrived at by means of dem. [onstrative?] definitions is quite arbitrary except that the former 1. must be known (at least through the principles from which they follow) at the time the language is constructed, and 2. must be sufficiently disconnected from the latter to avoid conflict between the two classes of rules. Hence if e.g. all astronomical truth were to follow from a few axioms, and there were moreover no correlation between directions on the sky and on the earth, then the axioms of astronomy could be incorporated as syntactical rules and demonstr. [ative?] definitions be restricted to earthly objects. This would make astronomy tautological. Nevertheless it would imply no loss in astronomical knowledge but only a change in the interpretation of astronomy. This procedure would succeed even better for astronomy than for mathematics because astronomy itself would not be necessary in order to justify the syntactical rules by their consistency. It is true that in the case of mathematics this representation by syntactical rules (if it is done along the lines of Ramsey, cnf. [?]) is particularly easy and so to speak "natural". Therefore it is justified to conclude that there is a close relationship between mathematics and language. Moreover in the light of the facts presented in this lecture, one should conclude not that mathematics is an outgrow of language, but rather that language is possible only by mathematics.

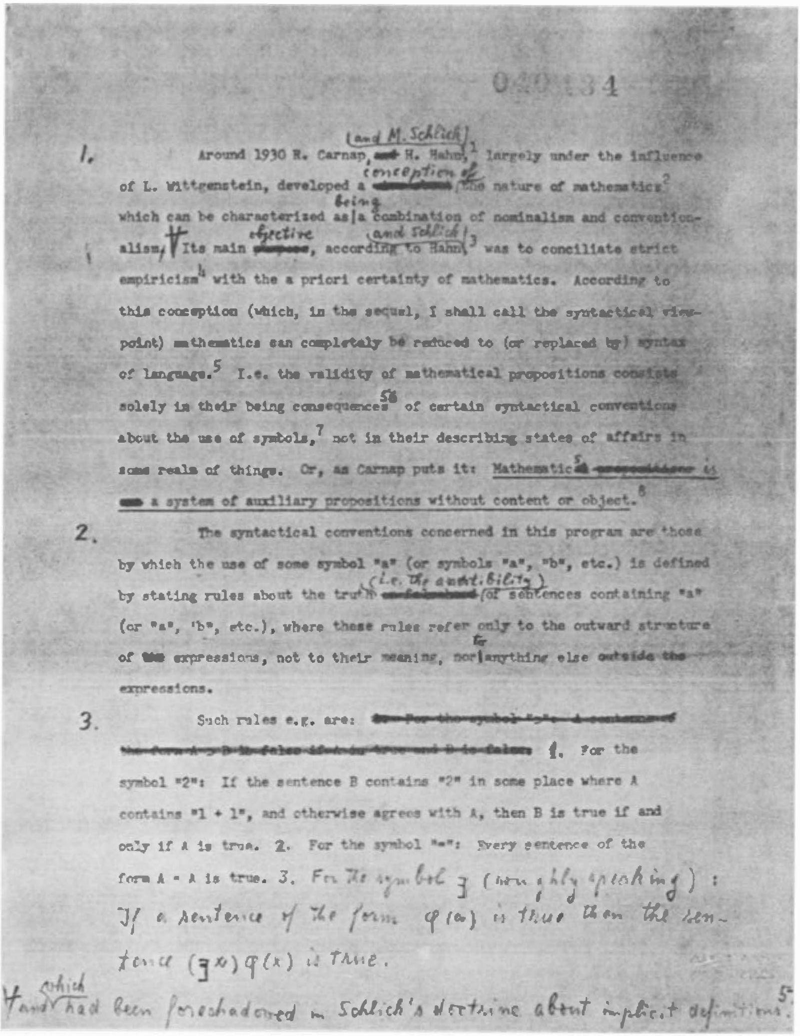
27 [[It is possible to maintain the objectivity (i.e. extramentality [?]) of concepts and still reject Platonism namely if something like the theory explained in the second half of footnote is assumed (except that now the inner sense (i.e. the faculty of inner self-perception) takes the place of the outer senses).]]

35 To be more exact the true situation as opposed to the view criticized is the following. 1. The meanings of the mathematical terms are not reducible to the linguistic rules about their use except for a very restricted domain of mathematics. 2. Even where such a reduction is possible the linguistic rules cannot be considered to be something man-made and propositions about them to be lacking objective content because these rules are based on the idea of a finite manifold (in the form of finite sequences of symbols) and this idea (with all its properties) is entirely independent of any convention and free choice (hence is something objective). In fact its theory is equivalent to arithmetic.

37 The chief difference between finitism (in the Hilbertian sense) and intuitionism [[is 1. Finitism restricts the application of the propositional connectives (~, ⊃, etc.) to decidable propositions (except of course in so far as their meaning could be defined in terms of other ideas admitted). 2. Finitism introduces]] is that the latter 1. does not admit any "abstract" concepts (such as "not", "implies", "there is", "function [?] of integers", or axioms referring to them) as primitive terms (or axioms) of mathematics proper. [[2. admits as objects of mathematics proper only objects of finite (and intuitively penetrable) complexity such as integers or finite sequences of symbols and concepts definable for them without the use of transfinite quantification (i.e. by induction alone) but remains within the domain of the "... [?]". Cnf. footn. [?].]]

45 If e.g. some new mathematical axiom had a great number of easily arrived at consequences verifiable separately without making use of this axiom there would be exactly as much reason to believe it to be true as for a well established law of physics. Nevertheless it might for other reasons be very unlikely that this axiom could ever be proved (or proved consistent) by constructive methods.

48 If however nominalists renouncing a proof wanted to be content with empirical evidence for this fact they would acknowledge thereby that their whole theory precisely is based on one of the objectively subsisting mathematical facts whose non existence it wants to prove.



040134

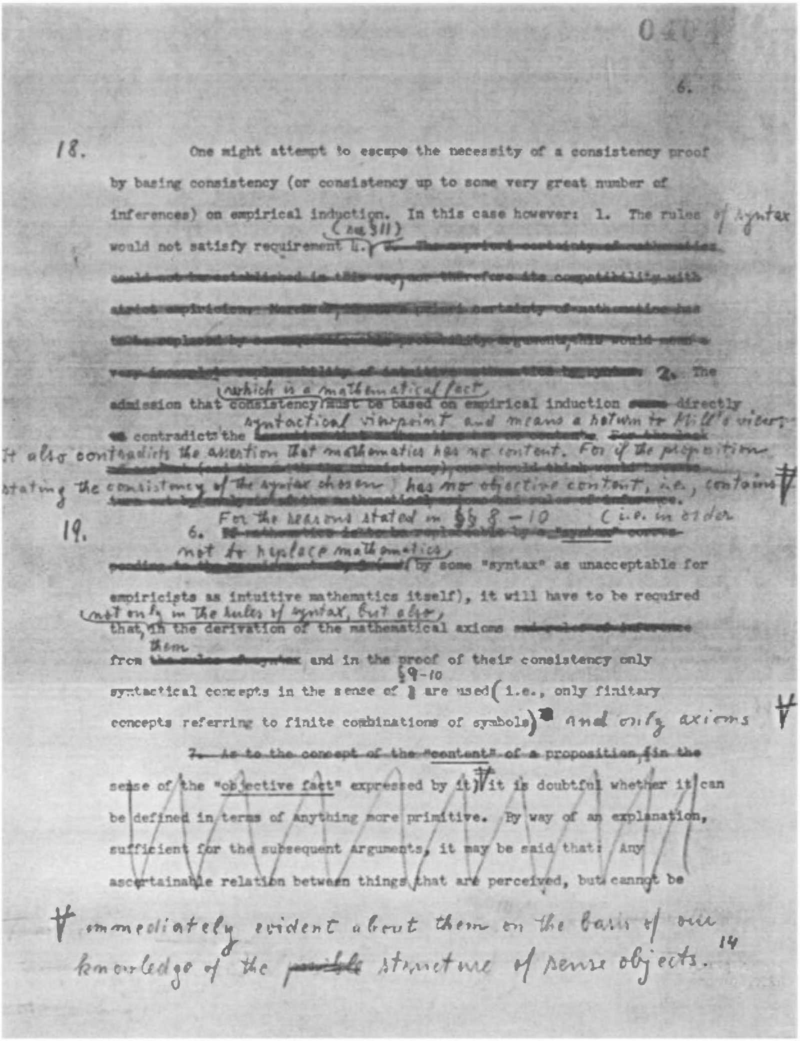
1\. Around 1930 R. Carnap, H. Hahn, (and M. Schlick)[1] largely under the influence of L. Wittgenstein, developed a conception of the nature of mathematics[2] which can be characterized as being a combination of nominalism and conventionalism, † Its main objective, according to Hahn (and Schlick)[3] was to conciliate strict empiricism[4] with the a priori certainty of mathematics. According to this conception (which, in the sequel, I shall call the syntactical viewpoint) mathematics can completely be reduced to (or replaced by) syntax of language.[5] I.e. the validity of mathematical propositions consists solely in their being consequences[56] of certain syntactical conventions about the use of symbols,[7] not in their describing states of affairs in some realm of things. Or, as Carnap puts it: Mathematics[5] is a system of auxiliary propositions without content or object.[8]

2\. The syntactical conventions concerned in this program are those by which the use of some symbol "a" (or symbols "a", "b", etc.) is defined by stating rules about the truth (i.e. the assertibility) of sentences containing "a" (or "a", "b", etc.), where these rules refer only to the outward structure of expressions, not to their meaning, nor to anything else.

3\. Such rules e.g. are: 1. For the symbol "2": If the sentence B contains "2" in some place where A contains "1 + 1", and otherwise agrees with A, then B is true if and only if A is true. 2. For the symbol "=": Every sentence of the form A = A is true. 3. For the symbol ∃ (roughly speaking): If a sentence of the form φ(a) is true then the sentence (∃x) φ(x) is true.

† and which had been foreshadowed in Schlick's doctrine about implicit definitions. 5.

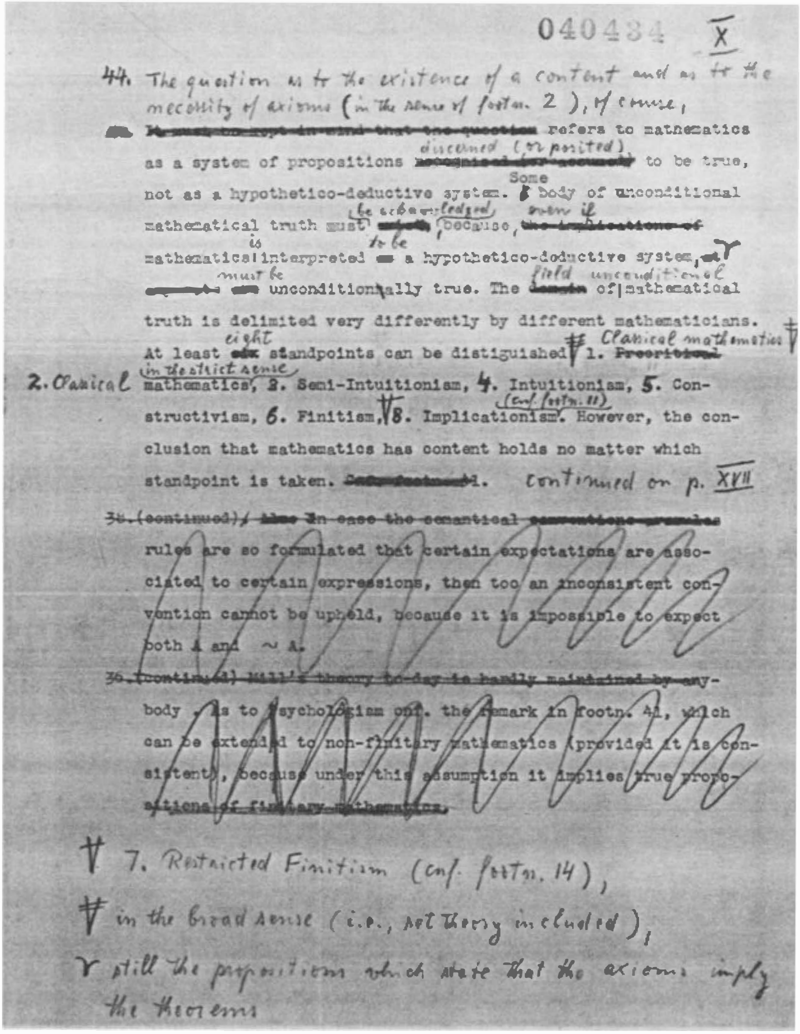
5\. Essay on Carnap, version II. First folio of the main text.


0404

6.

18. One might attempt to escape the necessity of a consistency proof
by basing consistency (or consistency up to some very great number of
inferences) on empirical induction. In this case however: 1. The rules of syntax
(see §11)
would not satisfy requirement 4. [illegible]
[illegible]
[illegible]
[illegible]
[illegible] 2. The
(which is a mathematical fact)
admission that consistency must be based on empirical induction ~~some~~ directly
syntactical viewpoint and means a return to Mill's views;
~~we~~ contradicts the [illegible]
It also contradicts the assertion that mathematics has no content. For if the proposition
[illegible]
stating the consistency of the syntax chosen) has no objective content, i.e., contains
[illegible].
For the reasons stated in §§ 8–10 (i.e. in order
19. 6. [illegible]
not to replace mathematics
[illegible] by some "syntax" as unacceptable for
empiricists as intuitive mathematics itself), it will have to be required
(not only in the rules of syntax, but also)
that, in the derivation of the mathematical axioms [illegible]
them
from [illegible] and in the proof of their consistency only
§9–10
syntactical concepts in the sense of § are used (i.e., only finitary
concepts referring to finite combinations of symbols) and only axioms
~~7. As to the concept of the "content" of a proposition (in the~~
~~sense of the "objective fact" expressed by it) it is doubtful whether it can~~
~~be defined in terms of anything more primitive. By way of an explanation,~~
~~sufficient for the subsequent arguments, it may be said that: Any~~
~~ascertainable relation between things that are perceived, but cannot be~~
immediately evident about them on the basis of our
knowledge of the ~~possible~~ structure of sense objects.[14]

6. Essay on Carnap, version II. Folio 6, where several kinds of corrections and changes can be seen.



040434 X

44. The question as to the existence of a content and as to the necessity of axioms (in the sense of footn. 2), of course, ~~It must be kept in mind that the question~~ refers to mathematics as a system of propositions discerned (or posited) ~~recognized (or assumed)~~ to be true, not as a hypothetico-deductive system. Some ~~A~~ body of unconditional mathematical truth must be acknowledged ~~exist~~, because, even if ~~the implications of~~ mathematics is interpreted to be a hypothetico-deductive system, still the propositions which state that the axioms imply the theorems must be unconditionally true. The field ~~domain~~ of unconditional mathematical truth is delimited very differently by different mathematicians. At least eight ~~six~~ standpoints can be distinguished: 1. Classical mathematics in the broad sense (i.e., set theory included), 2. Classical mathematics in the strict sense, 3. Semi-Intuitionism, 4. Intuitionism, 5. Constructivism, 6. Finitism, 7. Restricted Finitism (cf. footn. 14), 8. Implicationism (cf. footn. 11). However, the conclusion that mathematics has content holds no matter which standpoint is taken. ~~Cf. footnote 1.~~ continued on p. XVII

~~34. (continued) Also In case the semantical conventions or rules rules are so formulated that certain expectations are associated to certain expressions, then too an inconsistent convention cannot be upheld, because it is impossible to expect both A and ~ A.~~

~~36. (continued) Mill's theory to-day is hardly maintained by anybody. As to psychologism cf. the remark in footn. 41, which can be extended to non-finitary mathematics (provided it is consistent), because under this assumption it implies true propositions of finitary mathematics.~~

7. Essay on Carnap, version II. Folio X of the footnotes. Many of the folios pertaining to this group are single space typed.

3

# Is mathematics syntax of language?, II

1.

Around 1930 R. Carnap, H. Hahn and M. Schlick,[1] largely under the influence of L. Wittgenstein, developed a conception of the nature of mathematics[2] which can be characterized as being a combination of nominalism and conventionalism and which had been foreshadowed in Schlick's doctrine about implicit definitions.[3] Its main objective, according to Hahn and Schlick,[4] was to conciliate strict empiricism[5] with the a priori certainty of mathematics. According to this conception (which, in the sequel, I shall call the syntactical viewpoint) mathematics can completely be reduced to (or replaced by) syntax of language.[6] I.e. the validity of mathematical propositions consists solely in their being consequences[7] of certain syntactical conventions about the use of symbols,[8] not in their describing states of affairs in some realm of things. Or, as Carnap puts it: *Mathematics is a system of auxiliary propositions without content or object.*[9]

2.

The syntactical conventions concerned in this program are those by which the use of some symbol "*a*" (or symbols "*a*", "*b*", etc.) is defined by stating rules about the truth (i.e. the assertibility) of sentences containing "*a*" (or "*a*", "*b*", etc.), where these rules refer only to the outward structure of expressions, not to their meaning, nor to anything else outside the expressions.

3.

Such rules e.g. are: 1. For the symbol "2": If the sentence B contains "2" in some place where $A$ contains "1 + 1", and otherwise agrees with $A$, then $B$ is true if and only if $A$ is true. 2. For the symbol "=": Every sentence of the form $A = A$ is true. 3. For the symbol $\exists$ (roughly seaking): If a sentence of the form $\varphi(a)$ is true then the sentence $(\exists x)\varphi(x)$ is true.

4.

In his *Log. Synt. of Lang.* (p. 101 [?]–129) Carnap has carried this program out. Another method of carrying it through can be derived from a paper by F. P. Ramsey.[10] Finally much of the work of the Hilbert School about the formalization and consistency of mathematics can be interpreted to be a partial elaboration of this view,[11] although the authors of these papers, for the most part, favor different philosophical opinions.

5.

All these developments no doubt are interesting from a technical point of view, moreover they have contributed much to the clarification of some fundamental concepts. However, they prove that: 1. *Mathematics can be replaced by syntax of language.* 2. *Mathematical propositions have no content,* only if the terms "syntax", "content", etc. are taken in a very generalized (or a too restricted) sense, whereby these results become unfit to serve the afore-mentioned purpose of the syntactical program or the support of the philosophical views in question (such as nominalism or empiricism). On the other hand, if these terms are taken in their original sense (which also is the one required by the objective of the syntactical program and the philosophical questions involved), then assertion No. 1 (except for a limited section of mathematics) is disprovable. As to assertion No. 2 the examination of the syntactical viewpoint; perhaps more than anything else, leads to the conclusion that there *do* exist mathematical objects and facts which are exactly as objective (i.e. independent of our conventions or constructions) as physical or psychological objects and facts, although, of course, they are objects and facts of an entirely different nature. And this is true even for those sections of mathematics which *can* be reduced to syntax in the *original* sense of the term. Therefore the view that physics and other empirical sciences describe some "realm of things", while mathematics does not (cnf. the difference between "*Realwissenschaft*" and "*Formalwissenschaft*" in the passage quoted in footn. 9), seems hardly tenable.

6.

Starting with the first item, assertion No. 1 above, I shall now give a list of the meanings of the terms occurring in it, as they seem to me to be required by the objective of the syntactical program and the philosophical questions involved.

7.

1. Since the syntactical program aims at dispensing with mathematical intuition without impairing the usefulness of mathematics for the empirical sciences, it will have to be required of a satisfactory elaboration that mathematics is covered to the full extent to which it can be used in the empirical sciences (in particular for deriving verifiable consequences from laws of nature). This however is the case for all classical mathematics. Therefore "mathematics" will have to mean classical mathematics. But nothing is changed in the results stated in the sequel if it means intuitionistic mathematics. In both cases mathematics (in assertion No. 1) must be considered to be a system of propositions discernible to be true. For the question (see §1 and footn. 5 and 6) is exactly whether the intuitive content[12] of mathematics can be disregarded and nevertheless the theorems of mathematics be asserted and applied.[13]

8.

2. "*Language*" will have to mean some symbolism which can actually be exhibited and used in the empirical world. In particular it will have to be required that its sentences consist of a finite number of symbols. For sentences of infinite length (since they do not exist and cannot be produced in the empirical world) evidently are purely mathematical objects. Thus, such objects, instead of being avoided, would be assumed right from the beginning.

9.

3. For the same reason it will have to be required of the "*rules of syntax*" that they be "finitary", i.e., that they do not contain such phrases as e.g.: "If there exists an infinite set of expressions with a certain property", nor even: "If all expressions of a certain infinite set have a certain property" [[without naming some procedure for constructing it (in case it exists.)]]. For these phrases have no meaning if attention is confined to what is, or can be, given in sense perceptions, but rather some intuition for mathematical objects, such as

the totality of all possible expressions, or the intuitive concept of a correct proof, is necessary for them to be understood.[14]

10.

The occurrence of the phrases mentioned also contradicts the ordinary concept of a syntactical rule, which requires that the rule should refer only to the structures of given finite expressions, not to infinite sets of expressions, which would make the application of the rule depend on the solution of difficult, and possibly unsolvable, mathematical problems.

11.

4. Moreover a rule about the truth of propositions can be called *"syntactical"* only if it is clear from its formulation, or if it can be proved, that it does not imply the truth or falsehood of any "factual" proposition (i.e. one whose truth, owing to the semantical rules of the language, depends on experiences of some kind). This requirement in particular follows from the fact that mathematics is to be developed as an a priori science and that it is the lack of content upon which its a priori admissibility in spite of strict empiricism is to be based. The requirement under discussion implies that the rules of syntax must be demonstrably consistent, since from an inconsistency *every* proposition follows,[15] all factual propositions included.

12.

5. According to what was said in the end of footn. 6 and in the beginning of §7 the phrase: "Mathematics is *replaceable* by syntax of language" will have to mean: 1. that the formal axioms and rules of inference of mathematics can be deduced from the rules of syntax,[16] and 2. that the applications of mathematics to the empirical world, which formerly were based on the intuitive truth of the mathematical axioms, can be justified by syntactical considerations. The second item, however, again requires a consistency proof for the syntactical rules, as can be seen from the following example:

13.

If mathematics is interpreted to be a system of objectively true propositions, then, e.g., on the ground of a proof for Goldbach's

Conjecture (which says that every even number is the sum of two primes) it can be predicted that a computing machine (which is empirically known to work reliably) will find two primes whose sum is some large number N. In order to make the same prediction (for every N), if mathematics is interpreted syntactically, the consistency [[(and only the consistency)]] of the rules of syntax must be known. For a failure of the machine to furnish such a decomposition of N would entail the existence of a formal disproof of Goldbach's Conjecture (provided that the formalism of finitary arithmetic follows from the rules of syntax). On the other hand, from the mere fact that Goldbach's Conjecture follows from some arbitrarily assumed rules for handling certain symbols (even if they imply the formalism of finitary arithmetic), nothing whatsoever can be concluded about the result the machine will yield.

14.

The situation is quite similar for the prediction (on the basis of empirically known physical laws) that a bridge constructed in a certain manner will not break under a certain load; the single parts of the bridge playing the same role as the elements of the computing machine.

15.

Similarly also in pure mathematics one needs a consistency proof in order to draw the usual conclusions from mathematical theorems, e.g., in order to conclude from a proof of Goldbach's Conjecture that a finitary procedure for decomposing an even number N into two primes will yield a positive result for every N. But this exactly is the question one is interested in, if he wants to solve Goldbach's problem. Moreover it [[can be formulated as a syntactical question, and hence is meaningful from the syntactical point of view]] is a question regarding ascertainable facts. Therefore syntax would not be a substitute for intuitive mathematics in the sense of footn. 6, if this conclusion could not be drawn on its basis. (Note that the question discussed in this paragraph, in order to be meaningful, and not trivial, from the syntactical viewpoint, must refer to the actual performance of a computation and, therefore, like the two preceding questions concerns the applications of mathematics to reality.)

16.

Of course there is no question that by applying mechanically certain rules for handling mathematical symbols one arrives at the same formulas as by using mathematical intuition[17] (e.g., also at those formulas which express the two predictions and the mathematical conclusion dealt with in §§ 13–16). However mathematical intuition in addition creates the conviction that, if these formulas express observable facts and were obtained by applying mathematics to verified physical laws (or if they express ascertainable mathematical facts), then these facts will be brought out by observation (or computation). Therefore, syntax, if it is to be an acceptable substitute for intuition, must also yield sufficient reason for this conviction. And for this purpose a consistency proof is necessary.

17.

That the consistency of the syntactical rules must be known in order to apply mathematics can, more generally, be seen as follows: According to the syntactical viewpoint the whole purpose of introducing, in addition to the propositions expressing directly observable facts, others, which by definition have various mutual derivability relations to the former, is to be able to apply a principle which, roughly speaking, reads as follows: If some comparatively simple proposition *A* of the second kind implies the facts of a certain category which were observed, then any proposition about observables of this category which follows from *A* has a high probability of being verified. But this conclusion evidently presupposes that the negations of consequences of *A* are not likewise consequences of *A*.

18.

One might attempt to escape the necessity of a consistency proof by basing consistency (or consistency up to some very great number of inferences) on empirical induction. In this case however: 1. The rules of syntax would not satisfy requirement 4. (see §11) [[2. The a priori certainty of mathematics could not be established in this way, nor therefore its compatibility with strict empiricism. Moreover, if the a priori certainty of mathematics has to be replaced by some questionable probability argument, this would mean a very incomplete replaceability of intuitive mathematics by syntax.]] 2. The admission that consistency which is a mathematical fact must be based on empirical induction directly contradicts the syntactical viewpoint and means a return to Mill's view. It also contradicts the assertion that mathematics has no content. For if a proposition stating the consistency of the syntax chosen has no objective content, i.e., contains nothing beyond what we produce ourselves by our own syntactical stipulations (cnf. footn. 7), its truth must be known to us due to the knowledge we have of our own acts, provided only that we give a complete account of them to ourselves. And these conclusions cannot be escaped by interpreting a proposition about consistency to be a physical proposition about written or spoken language. For this physical proposition, too, is analytic, although it contains non-logical concepts (exactly as, e.g., the proposition: "Either it will rain or it will not rain").

19.

6. For the reasons stated in §§ 8–10 (i.e. in order not to replace mathematics by some "syntax" as unacceptable for empiricists as intuitive mathematics itself), it will have to be required that, not only in the rules of syntax, but also in the derivation of the mathematical axioms from them and in the proof of their consistency only syntactical concepts in the sense of § 9–10 are used (i.e., only finitary concepts referring to finite combinations of symbols) [[[18]]] and only axioms immediately evident about them on the basis of our knowledge of the structure of sense objects.[19]

[[7. As to the concept of the *"content"* of a proposition, in the sense of the *"objective fact"* expressed by it, it is doubtful whether it can be defined in terms of anything more primitive. By way of an explanation, sufficient for the subsequent arguments, it may be said that: Any ascertainable relation between things that are perceived, but cannot be created by us, which moreover is such that its subsisting in no way depends on our arbitrary acts or decisions, constitutes an objective fact (although, vice versa, every objective fact need not necessarily satisfy all these conditions).]]

20.

The requirements 2, 3, 6 are particularly important in order to attain the objective of the syntactical program. For, if mathematical intuition is to be dispensed with by means of syntax, it certainly will have to be required that the use of the "abstract" concepts[20] of mathematics and of the axioms referring to them, which cannot be understood or known without mathematical intuition, be based on considerations about concretely presentable finite combinations of symbols. If, instead, in the formulation of the syntactical rules some of the very same abstract concepts are being used, or in the consistency proof some of the axioms usually assumed about them, then the whole program completely changes its meaning and is turned into its downright opposite: Instead of clarifying the meanings of the abstract mathematical forms by explaining them in terms of syntactical rules, abstract mathematical concepts are necessary in order to formulate the syntactical rules; and instead of justifying the mathematical axioms about these concepts by reducing them to syntactical rules, these axioms (or at least some of them) are necessary in order to justify the syntactical rules as consistent.

21.

Therefore, only if the syntactical program could be carried through under the requirements 1–6, the assertion that mathematics is syntax of language would have any relevance for the problem of nominalism and conventionalism versus realism, or empiricism versus rationalism. Whether it would entail that mathematics has no content is a different question (cnf. §34).

22.

Now, do the elaborations of the syntactical viewpoint which have actually been given satisfy the requirements 1–6? By no means. Ramsey's ideas necessitate admitting propositions of infinite (and even non-denumerable) length. Carnap uses non-finitary syntactical rules and arguments. Formalism under the requirements 2–6 has yielded syntactical foundation only for a small part of mathematics. Only if requirement 6 or 5 is substantially weakened (cnf. §24 and footn. 39), there is some hope that it may yield a foundation for all mathematics.

23.

Now the question arises whether this failure applies only to the particular attempts that have been made or has deeper reasons. The answer to this question depends on the precise extension of "finitary combinatorial reasoning". Now an explicit definition of this concept has not been given yet. But, in view of the work by G. Gentzen,[21] there can be no doubt that all such reasoning can be expressed in the formalism of classical number theory. Since, however, owing to a general theorem, a consistency proof for a system containing primitive recursive number theory can never be expressed in this system, it follows that not even classical number theory, still less any more comprehensive systems, can be proved consistent by finitary reasoning.[22]

24.

The fact that, in addition to the concepts directly referring to concretely given combinations of symbols [[(which are all contained in finitism)]], certain "abstract" concepts (together with their axioms) must be admitted, in order to make a consistency proof for classical mathematics possible, was recognized also by leading proponents of formalism.[23]

25.

So, in carrying through the syntactical program, *the requirements 1–6 cannot be satisfied simultaneously,* i.e., *the intuition for, or the assumptions about, the abstract and transfinite concepts of mathematics cannot be replaced by considerations about finite combinations of symbols and their finitary properties and relations,* where "replacing" here means that the same consequences as to ascertainable facts (including general propositions about them[24]) can be derived in either case, and "ascertainable fact" means any numerical equation for computable functions or trivially equivalent propositions.[25]

26.

This holds no matter whether "mathematics" is understood to mean classical or intuitionistic or constructivistic mathematics, or even intuitionistic number theory.[26] Only if mathematics itself is confined to some level[27] of finitary combinatorial reasoning (and moreover for certain sections of mathematics which contain abstract concepts and their axioms only with [[substantially weakened axioms]] artificial restrictions or which do not contain primitive recursive number theory) the syntactical program can be carried through under the requirements 2–6.

27.

But in these cases an even larger section of finitary mathematics, than the one contained in the system, is necessary for the consistency proof. *Therefore, also for finitary mathematics, the syntactical interpretation gives no rational foundation for the belief in its correctness.* Moreover, if syntax is understood in the strict sense explained in the end of footn. 19 (which is necessary in order to build up mathematics without introducing any purely mathematical objects by the help of mathematical intuition (cnf. footn. 5)) then, very likely, not even finitary mathematics can be founded on syntax. For it probably is demonstrable that the consistency of very great integers, even for a moderate number of proof steps, cannot be proved, in a feasible number of steps, without using very great numbers in the consistency proof.

28.

In general the concepts and axioms occurring in the section of mathematics considered need not all occur among (or be derivable from) those sufficient for a consistency proof. There exist certain possibilities of replacing some of them by others.[28] But the very fact that, if some of the essential concepts or axioms are omitted, others of equal power become necessary reveals a principle which could perhaps be called the *"Non-eliminability of the mathematical content of an axiomatic system by the syntactical interpretation."* This principle can be made precise in several ways, e.g. thus: The axioms used in the consistency proof must have at least the same (in fact even a slightly greater) demonstrative power within finitary combinatorics (or finitary number theory) than those proved consistent; or thus: It must be possible [[in a certain sense]], by means of the concepts and axioms used in the consistency proof, to construct a "model" for those proved consistent,[29] i.e., to define concepts demonstrably, satisfying the given axioms and not satisfying any proposition disprovable from them.[30] So in this slightly weakened sense the axioms used do imply those proved consistent. Finally also the above-mentioned theorem which says that the consistency of a system *S* of axioms containing arithmetic cannot be proved in any subsystem of *S* belongs into this order of ideas.

[[In what sense the results stated in the preceding paragraphs justify the inference that mathematical objects and facts do exist will be discussed later (cnf. §43). This conclusion however can be reached much more easily and directly by other arguments, which I shall explain now.]]

29.

On the ground of these results it can be said that *the scheme of the syntactical program to replace intuition by rules for the use of symbols fails because this replacing destroys any reason for expecting consistency, which is fundamental for both pure and applied mathematics, and because for the consistency proof a mathematical intuition of the same power is necessary as for discerning the truth of the mathematical axioms.*

30.

This formulation of the non-feasibility of the syntactical program (which also applies to finitary mathematics) is particularly well suited for elucidating the question as to whether mathematics is void of content [[in the sense that no mathematical objects or facts exist]]. For, if *prima facie* content of mathematics were only a wrong appearance, it would have to be possible to build up mathematics satisfactorily without making use of this "pseudo" content.

31.

More precisely the situation can be described as follows: That mathematics *does* have content (in any acceptable sense of the term) appears from the fact that, in whatever way it (or any part of it) is built up, one always needs certain undefined terms and certain axioms about them.[31] *For these axioms there exists no other rational foundation except that either they can directly be perceived to be true (owing to the meaning of the terms or by an intuition of the objects falling under them); or* *else that they are assumed (like physical hypotheses) on the ground of inductive arguments, e.g., their success in the applications.*[32] The former case would [[unquestionably]] seem to apply at least to some mathematical axioms, e.g., the *modus ponens* and complete induction.[33] In the latter case the mathematical character of the axioms, in spite of their inductive foundation, appears in the circumstance that [[they are necessary in order to derive verifiable consequences from the laws of physics, exactly as the laws of physics themselves, and that moreover]] they have consequences not obtainable without them in that part of mathematics to which the former case applies, i.e., whose primitive terms have an immediately understandable meaning [[This circumstance also shows that these axioms of the second kind remain mathematical even though an inductive character may be attributed to them.]] (e.g. the axioms of infinity mentioned in footn. 45 have number-theoretical consequences).

32.

*A third possibility, namely to posit the mathematical axioms by convention, does not exist.* For, before any such convention can be made, mathematical axioms of the same power are necessary already in order to prove the consistency of the envisaged convention. A consistency proof, however, is indispensable because it belongs to the essence of a convention, as opposed to a proposition with content, that it cannot be disproved and that, in particular, it does not imply any propositions which can be falsified by observation (cnf. §11). Without a consistency proof the "convention" itself, since open to disproof, would really be an assumption.[34] Brought to its shortest form this proof runs as follows: If mathematical intuition is accepted as a source of knowledge, the existence of a content of mathematics evidently is admitted. If it is rejected, mathematics becomes open to disproof and for this reason has content.

33.

All attempts of analyzing away the given or assumed facts expressed in the axioms of mathematics have failed and must necessarily fail: The axioms demonstrably cannot be replaced by definitions and the rule of substituting the *definiens* for the *definiendum.*[35]

If mathematics is reduced to logic, then axioms about the primitive terms of logic[36] must be assumed, some of which are so far from trivial, that they are rejected as false by many mathematicians. If the mathematical axioms are replaced by syntactical rules, one needs axioms of the same power about the primitive terms of syntax or about abstract concepts to be used in the syntactical considerations.

34.

All this also applies to finitary mathematics, which, therefore, likewise has content. Hence,

[[ ... [?] symbols fall under certain finitary rule belong to the class of those mathematical facts which, in a certain sense, are the simplest. This class contains, e.g., also the numerical equations, like 2 + 3 = 5. That such an equation is true, in syntactical interpretation, means the combinatorial fact that it follows (in a finite number of steps) from the general rule defining the arithmetical operation occurring (e.g. $(x + y) + 1 = x + (y + 1)$ in the case of addition). If, following Frege, we give a meaning to the arithmetical symbols, the equation expresses a similar combinatorial fact in terms of the primitive concepts of logic (i.e. the concepts necessary are then more fully analyzed).

These facts of finite combinatorics, both as to their logical structure and their immediacy and certainty, correspond exactly to the sense perceptions in physics. Like the sense perceptions they are not objectively the simplest (these are the axioms concerning the primitive terms of logic), but they are the most obvious to us.

These arguments show also finite combinatorics by no means is void of content. Therefore]]

if all of mathematics could be reduced to finitary syntax (in accordance with the requirements 2–6) this would mean, not that it has no content, but only that its content would not be larger than that of finitary combinatorics. (Note that in this case a model for mathematics could be defined in terms of finitary combinatorics and that, given finitary combinatorics, the rest of mathematics could really be introduced by conventions.) In reality it turns out to be infinitely larger. It should be noted, however, that this is only the result of a mathematical investigation. A priori it is conceivable that the syntax necessary for building up and applying mathematics would require only an insignificant part of finitary combinatorics, as would be the case, e.g., if all mathematics could be derived from explicit definitions and the law of identity. In that case there would be some justification to calling mathematics "void of content".

35.

It can be shown (see §37) that *the reasoning which leads to the conclusion that no mathematical facts exist is nothing but a* petitio principii, *i.e. "fact" from the beginning is identified with "empirical fact", i.e. "fact in the world of sense perception"*.[37] In this sense the voidness of content of mathematics can be admitted, but it ceases to have anything to do with the philosophical questions mentioned in the beginning of this paper, since also Platonists should agree that mathematics has no content of this kind. For its content, according to Platonism, does not consist in facts perceptible with the senses, but in relations between concepts or other ideal objects.

36.

All pure mathematics that has been developed, as well as its applications in the empirical sciences, it is true, can be replaced (in the sense of footn. 6 or §12[38]): I. By non-finitary syntax. II. By finitary syntactical rules plus an empirical knowledge of their consistency. III Some sections of mathematics can even be replaced by syntax under the requirements 2–6.[39] In all these schemes the syntactical rules replacing mathematics are really "void of content" in this sense that they do not imply the truth or falsehood of any sentence stating an empirical fact (i.e. they are compatible with all distributions of truth-values for atomic – i.e. indecomposable – empirical propositions), although in case II this is known only empirically. But to conclude from this state of affairs that mathematics (or the section of mathematics concerned) has no content is possible only owing to the *petitio principii* mentioned. This can be brought to complete evidence as follows:

37.

The possibility of building up mathematics (or parts of mathematics) in the ways described under I, II, III, IV (cnf. footn. 39) and its lack of content in the sense explained, evidently is due solely to these two circumstances: 1. That pure mathematics implies nothing about the truth values of those propositions which contain no logical or mathematical symbols (i.e., within the field of mathematics and natural science, the atomic empirical propositions), 2. that mathematics (or rather the part of mathematics concerned) follows from a finite number of axioms and formal rules, known at the time when the language is constructed.[40]

38.

These two conditions, however, might very well be satisfied also for some portion of empirical science with respect to the rest of it. We might, e.g., possess an additional sense that would show to us a second reality, so widely separated from space-time reality, as not to allow us any conclusions about the latter, and moreover so regular, that it could be described by a finite number of laws. We could then, by an arbitrary decision,[41] recognize only the first reality as such, and declare propositions referring to the other one to be without content and true only in consequence of syntactical conventions. These could be chosen as to make exactly those propositions true which could be seen to be true with the supposed additional sense.[42]

39.

I even think that this comes pretty close to the true state of affairs (cnf. footn. 33), except that this additional sense (i.e. reason) is not counted as a sense, [[because it differs too much from the others, in particular by the fact that the truth of universal propositions is directly (i.e. without induction) perceivable by it.]] because its objects are quite different from those of all other senses. For, while with the latter we perceive "the particular", with reason we perceive "the general".[43]

40.

But can it at least be said that, owing to the syntactical interpretations that are feasible, the intuitive content of mathematics can be disregarded without loss except that mentioned in footn. 38? It cannot. For [[under I (or III, IV)]] either the intuitive content of mathematics, to a very large extent, is necessary for the syntactical considerations replacing mathematics [[Under II, it is true, this applies only to a very small extent, and not at all if attention is confined to marks on paper and syntax is interpreted to be "physics of symbols". However, in the latter case 2 + 2 = 4 becomes an empirical fact and that this means purposely closing one's eyes in the face of obvious facts,]] or else consistency, which is necessary for the application of mathematical theorems (cnf. §§ 13–16), must be based on empirical induction. I believe that, at least for finitary and some parts of intuitionistic mathematics, practically everybody will agree that the consistency proof based on mathematical intuition is incomparably better.

41.

The preceding paragraph also answers the first of the two questions raised in footn. 13, namely as follows: In order to arrive at the same conclusion about ascertainable facts as one who uses mathematical intuition, the syntactical interpretation, without an extensive use of mathematical intuition, is not sufficient, but rather at least one mathematical fact must be known empirically, namely the consistency of the syntactical system used. However, to accept some mathematical proposition as true and admit that it can only be known with the help of empirical induction is incompatible with the syntactical viewpoint (cnf. §18).

42.

From later publications of Carnap it appears that today he would hardly uphold the formulation quoted in footn. 9. From what he says in *Rev. Int. Phil.* 11 (1950), p. 35, e.g., it follows that at present he does not object to associating, in scientific semantics, mathematical objects to formulas as their meaning or denotation. However he maintains (in the article just quoted) that the philosophical question about the objective existence of mathematical objects does not refer to this "internal" existence, but means whether these objects, formally introduced by axioms, "really" exist. An answer to this question is asserted to have no "cognitive content", i.e. the question is considered to be meaningless, while formerly it was answered negatively; or else Carnap has changed his opinion about internal existence in mathematics.

43.

At any rate Carnap's present standpoint as to the existence of mathematical objects does not lend itself to founding any difference between "*Real-*" and "*Formal*"-*wissenchaft*, in the sense of the passage quoted in footn. 9. For exactly the same considerations, according to Carnap, apply to the question of the existence of objects of physics and even of everyday life. Moreover, also as to the *meaning* of "internal" existence there is no difference between the two cases. For the relevant existential assertions in either case can only be based either on direct perception or on assumptions which are made because they imply correct answers to questions regarding ascertainable facts (cnf. §44), and which can be disproved, but not on conventions (cnf. §32). On the other hand, if the meaning of the term "convention" is so enlarged that conventions do not need any consistency proof, then physical hypotheses also can be classed as conventions. [[Also the relationship of the objects of science to our experience is basically the same in the *Real-* and *Formalwissenschaft*. For the experience of understanding and discerning the truth of mathematical facts corresponds exactly to the experience of sense perception (cnf. footn. 33).]] Nor is it true of the mathematical "conventions" (not even of those of finitary mathematics, cnf. footn. 19) that in case of an inconsistency with observation we never would (or ought to) change them, but rather the laws of nature (e.g. by assuming systematic errors). The appearance to the contrary is due solely to the convincing power of the very same mathematical intuition, which conventionalism is supposed to make unnecessary.

44.

It is exactly from the syntactical viewpoint, which does not acknowledge the direct evidence of mathematical axioms, nor wants to make use of an "understanding" of the primitive terms of mathematical or physical theories, nor distinguishes between "causation" and other constant connections, that it is least of all possible to make a difference on principle between mathematics and natural science.

For as far as verifiable consequences of theories are concerned the mathematical axioms are exactly as necessary for obtaining them as the laws of nature (cnf. footn. 41). If, e.g., the impredicative axioms of analysis are necessary for the solution of some problem of mathematical physics, these axioms will imply predictions about observable facts not obtainable without them. Moreover it is perfectly conceivable that an inconsistency with observation may be due not to some wrong physical assumptions but to an inconsistency of these axioms. This shows that, from the syntactical point of view, the mathematical axioms can be looked upon as part of the physical theory which is only well defined after they have been given. Also for the considerations given in §17 it makes no difference whether *A* is a mere description of the experiences with the help of logical terms or contains physical concepts.

45.

It seems to me that this whole situation viewed from Carnap's own standpoint can only mean that the mathematical axioms are certain irreducible, largely existential, hypotheses which are exactly as necessary for the scientific description of reality as, e.g., the hypothesis of the existence of elementary particles or of some field satisfying certain equations. [[Also in both cases these hypotheses are disprovable, in physics by deriving some wrong propositions about sense perceptions, in mathematics by deriving some wrong numerical equation (which, in two valued logic, is equivalent to deriving an inconsistency).]]

46.

Therefore, also from the empirical standpoint, *there is not the slightest reason to answer the question of the objective existence of mathematical and empirical objects and facts differently, while it is the essence of the syntactical viewpoint to distinguish the two exactly in this respect.* (Cnf. the quotation in footn. 9).

47.

There is real difference between the two that the acceptance of mathematical axioms (unlike that of laws of nature) so far has been based exclusively on their intuitive evidence, and not on the success of their consequences.[44] However, in view of the fact that a large proportion of mathematical problems (even of the number-theoretical problems of Goldbach type mentioned in footn. 24) may not be solvable in this fashion and that, moreover, axioms with an incomparably greater demonstrative power might be discovered empirically, it is not impossible that this attitude will be given up sometime in the future.[45] Today already, to recognize the intuitionistic critique to be justified, and continue to use classical mathematics in the applications, is something of this nature.

48.

I do not want to conclude this paper without mentioning the paradoxical fact that, although any kind of nominalism or conventionalism, in mathematics turns out to be fundamentally wrong, nevertheless the syntactical conception perhaps has contributed more to the clarification of the situation than any other of the philosophical views proposed: on the one hand by the negative results to which lead the attempts to carry it through, on the other hand by the emphasis it puts on a difference of fundamental importance, namely the difference between empirical and conceptual truth, upon which it reflects a bright light by identifying it with the difference between empirical and conventional truth. [[But it must of course be admitted that there is a fundamental opposition in the character of these two kinds of truth.]]

49.

I believe the true meaning of the opposition between things and concepts or between factual and conceptual truth is not yet completely understood in contemporary philosophy, but so much at least is clear that in both cases one is faced with "solid facts", which are entirely outside the reach of our arbitrary decisions.

# Gödel's footnotes

1 Cnf: R. Carnap, *Erk.* 5 (1935), 30; *Einheitswissenschaft,* Heft 3 (1934). H. Hahn, *ibid.* Heft 2 (1933), French translations of these papers in *Act. Sci. Ind.* 291, 226. Moreover: H. Hahn, *Erk.* 1 (1930), p. 96; *Erk.* 2 (1931), p. 135; *Krise u. Neuaufbau in den exakten Wissenschaften,* Leipzig 1933. M. Schlick, *Gesammelte Aufsätze,* 1938, p. 145, p. 222.

2 The terms "mathematical", "mathematics" throughout this paper are used as synonymous with "logico-mathematical", "logic and mathematics." Moreover the term "axiom" is always used in the sense of "formal axiom or rule of inference" and is applied to any assertion that can be used in a complete proof without being [[expressed]] proved, in both contensive and formalized mathematics. The term "contensive" was suggested as a translation of "*inhaltlich*" by H. B. Curry in [?] [see also Curry, *Foundations of mathematical logic,* 1963, p. 14].

3 Cnf: M. Schlick, *Allgemeine Erkenntnislehre,* 1. Aufl., 1918, p. 30.

4 Cnf.: H. Hahn, *Act. Sci. Ind.* 226, p. 13, 19 and M. Schlick, *l. c.*, p. 147

5 The tenet of empiricism in question evidently is this that all knowledge is based on sense perceptions and their digest, and that we do not possess an intuition into some independent realm of abstract (mathematical) objects. The objective of the syntactical program, therefore, could also be stated thus: to build up mathematics as an a priori science without using mathematical intuition or referring to any mathematical objects or facts.

6 The precise formulation on the basis of the papers quoted would be: *Mathematics is syntax of language.* Since however mathematics is usually presented as a science of certain objects, about which certain propositions are asserted to be demonstrably true, the

question is whether it can be *replaced* by syntax; i.e. whether what is asserted in mathematics can be *interpreted* in terms of syntax, and whether on the basis of this interpretation the same conclusions as to ascertainable facts can be drawn, if mathematical theorems are applied.

7 Combining this assertion with another one of the syntactical school, namely that a logical consequence only repeats part of the content of the premisses, one arrives at the conclusion that a statement to the effect that some mathematical proposition is true does nothing else but repeat part of the conventions about the use of the symbols. This shows most clearly the lack of objective content of mathematical propositions on the basis of this view. However some proponents of the syntactical conception apparently would prefer to regard the derivability of a mathematical theorem as an empirical fact and to place the voidness of content of mathematical propositions in the circumstance that nothing except the arbitrarily chosen rules of syntax is used in deriving them. Cnf., however, what is said in §§ 30–33 and 43.

8 E.g., according to Hahn, *Act. Sci. Ind.* 226, p. 26, the laws of contradiction and of excluded middle express certain conventions about the use of the sign of negation.

9 Cnf. *Erk.* 5 (1935), p. 36; *Act. Sci. Ind.* 291, p. 37. The whole passage reads as follows: "Wenn zu der Realwissenschaft die Formalwissenschaft hinzugefügt wird, so wird damit kein neues Gegenstandsgebiet eingeführt, wie manche Philosophen glauben, die den 'realen' Gegenständen der Realwissenschaft die 'formalen' oder 'geistigen' oder 'idealen' Gegenstände der Formalwissenschaft gegenürberstellen. Die Formalwissenschaft hat überhaupt keine Gegenstande; sie ist ein System gegenstandsfreier, gehaltleerer Hilfssätze." ["In adjoining the formal sciences to the factual sciences *no new area of subject matter* is introduced, despite the contrary opinion of some philosophers who believe that the "real" objects of the factual sciences must be contrasted with the "formal", "*geistig*" or "ideal" objects of the formal

sciences. *The formal sciences do not have any objects at all;* they are systems of auxiliary statements without objects and without content". In Feigl/Brodbeck (eds.), *Readings in the philosophy of science,* New York, 1953, p. 128.] I would like to say right here that Carnap today would hardly uphold the formulations I have quoted (cnf. §42). Moreover some of them were given only by Hahn and Schlick, and probably would never have been subscribed to by Carnap. [[That nevertheless I am discussing them in detail has two reasons, namely: 1. While the program itself and its elaboration, as far as it is feasible, have been presented in detail in several publications, the negative results as to its feasibility in its most straightforward and philosophically most interesting sense have nowhere been discussed. 2. The syntactical program in the form I am presenting it here is a priori perfectly sound and poses an interesting problem. Only a thorough mathematical investigation can decide on its feasibility.]] However, I am not concerned in this paper with a detailed evaluation of what Carnap has said about the subject, but rather my purpose is to discuss the relationship between syntax and mathematics from an angle which, I believe, has been neglected in the publications about the subject. For, while the syntactical program itself and its elaboration, as far as it is possible, have been presented in detail, the negative results as to its feasibility in its most straightforward and philosophically most interesting sense have never been discussed sufficiently.

10 Cnf. *Proc. Lond. Math. Soc.,* 2 s. vol 25 (1926), p. 338. Reprinted in *The foundations of mathematics and other logical essays,* by F. P. Ramsey, 1931.

11 Because the axioms and rules of inference of formal systems can be interpreted to be syntactical rules which state that: 1. All formulas of a certain structure are true. 2. Formulas obtained from true formulas by certain formal operations are also true. Moreover consistency will turn out to be the key problem also for the syntactical viewpoint.

12 [[The existence, as a psychological fact, of an intuition covering the axioms of classical mathematics can hardly be doubted, if the in-

tuitionistic critique is recognized to be justified. In fact this intuition (which may be called the "natural mathematical intuition") has a high degree of precision, as can be seen from the fact that mathematicians never disagree as to the question of the correctness of a proof, even if they are not familiar with the precise axiomatization of classical mathematics.]] The existence, as psychological fact, of an intuition covering the axioms of classical mathematics can hardly be doubted even by adherents of the Brouwerean school, except that the latter will explain this psychological fact by the circumstance that we are all subject to the same kind of errors, if we are not sufficiently careful in our thinking.

13 In consequence of the intuitionistic critique it has become customary to regard, not the theorems of classical mathematics, but only their derivability from certain axioms, to be mathematical truth. But thereby mathematics, to a very large extent, loses its applicability (cnf. §§ 13–16), unless the consistency of the axioms is known, which however cannot be known on the basis of this "implicationistic" standpoint. For the consistency of a system of axioms is not equivalent to a proposition of the form: *B* follows from the axioms *A*. Therefore if only implicationistic mathematics could be replaced by syntax, mathematics as to its applications could not. [[Moreover the delimitation of mathematical truth given by Implicationism is very artificial. For it acknowledges only particular mathematical truths (i.e., propositions of the same type as $2 + 2 = 4$ and their consequences of the form: [?]), while we perceive with the same certainty and distinctness that $a + b = b + a$ holds for *all* integers *a, b*. For a more detailed exposition of implicationism cnf. K. Menger, *Blaetter fuer Deutsche Philosophie*, Bd. 4, 1930, p. 324.]]
The concept of "discernible truth" [[(*inhaltliche Richtigkeit*)]], as applied to the propositions of classical mathematics themselves, may be rejected as meaningless. However, what is necessary here is only to imagine some mathematician who believes in the truth of classical mathematics (i.e., in the "natural mathematical intuition" mentioned in footn. 12) and to inquire, whether the consequences he can arrive at on the basis of this belief as to propo-

sitions considered to be meaningful can also be obtained on the basis of some syntactical interpretation without using mathematical intuition. Cnf. the answer to this question in §41. Moreover, if the syntactical viewpoint with regard to mathematics, in contradistinction to the factual sciences, is to make any sense, *some* concept of truth, other than truth by syntactical convention (let us call it "objective truth") must be admitted and the question then is whether, due to some syntactical interpretation, the consequences of the mathematical axioms occurring in the applications, can be obtained without knowing or assuming the objective truth (in the sense admitted) of the mathematical axioms, or at least a considerable number of them. Cnf. the answer to this question given in footn. 34.

14 Carnap, to the objection that Platonism is implied by transfinite rules, replies (*Log. Synt.*, p. 114) that one may know how to handle the transfinite concepts (in inferences, definitions, etc.) without making any metaphysical assumptions about the objective existence of the abstract entities concerned. This, of course, is true in the same sense as one also may know how to handle the concepts of physical objects without ascribing to these objects any existence in a metaphysical sense. But nevertheless, before one can rationally use them in science, he must assign to them reality (or objectivity) at least in some immanent (Kantian) sense, in order to distinguish them from wrong (i.e. disprovable) physical hypotheses. The same, therefore, applies to the transfinite mathematical entities, whose existence also can be disproved, namely by an inconsistency derived from them. Hence, if mathematics is based on transfinite syntax, it is implied that there are two different realities of equal standing, which is exactly what the syntactical viewpoint denies (cnf. footn. 9). To finitary rules evidently this argument does not apply, because the objects to which they refer may be considered to be parts of the physical world (either existing already or producible).

15 This holds both for classical and intuitionistic logic and, therefore, also for the syntactical rules replacing them (cnf. §12). Note that

under rather general assumptions about the logic chosen consistency of the syntactical rules vice versa implies their compatibility with all possible sense experiences that can have occurred at any time.

16 If mathematics is to be replaceable by syntax only in the restricted sense that each individual observable fact which is a mathematical consequence of laws of nature or of other observable facts and empirical induction should also follow on the basis of the syntactical interpretation, then it is not necessary that the mathematical axioms themselves be derivable from the rules of syntax, since the same individual consequences as to observable may follow also with the help of different general concepts or axioms. As to this approach cnf. footn. 39. [[In particular it then is sufficient if the inference illustrated by Goldbach's Conjecture on §13 can be drawn finitarily for each particular N. However also in this weakened sense mathematics cannot actually, but only theoretically, be replaced by syntax.]]

17 Except that *all* mathematical intuition very likely cannot be expressed in *one* formal system (cnf. footn. 45 and 40) and demonstrably in none that itself is intuitively evident.

18 [[It is important to note that, if mathematics is to be completely replaceable (or interpretable, or justifiable) by syntax, then, not only each proposition demonstrable in mathematics, but also the mathematical procedures of proof, must be derivable from the rules of syntax.]]

19 I believe that what must be understood by "syntax", if syntax is not to presuppose some Platonistic realm of all possible combinations of symbols or other abstract entities (cnf. footn. 20), is exactly equivalent to Hilbert's "Finitism", i.e., it consists of those concepts and reasonings, referring to finite combinations of symbols, which are contained within the limits of "that which is directly given in sensual intuition" (*"das unmittelbar anschaulich Gegebene"*). (cnf. *Math. Ann.* 95 (1925), p. 171–173). The section

of mathematics thus defined is equivalent with recursive number theory (cnf. Hilbert-Bernays, *Grundlagen der Mathematik*, Bd. 1, p. 20–34 and p. 307–346), except that it may rightly be argued (cnf. P. Bernays, *L'enseignement math.*, 34 (1935), p. 61) that exorbitantly great integers must not occur in finitary proofs, because they are theoretical constructions which are as far apart from the "immediately given" (and even from anything given in space-time reality) as the infinite, and, therefore, cannot be known to be meaningful or consistent, unless we trust some abstract mathematical intuition. If restrictions of this kind are introduced (namely to the effect that the integers and the number of elements of combinations referred to in a theorem or its proof must not be above some limit), then the negative results as to finitary consistency proofs mentioned in the sequel remain valid (cnf. footn. 22) and stronger ones probably can be obtained (cnf. §27).

20 Such abstract concepts, e.g., are: "proof", "function", "there is", "infinite set", where the first three of these terms are to be understood in their original "contensive" [see footn. 2] meaning, i.e.: "Proof" does not mean a sequence of expressions satisfying certain formal conditions, but a sequence of thoughts (or rather forms of thought) creating conviction in a sound mind; "function" does not mean an expression of the formalism, but an understandable and precise rule associating mathematical objects with mathematical objects (in the simplest case integers with integers); "there is" means objective existence irrespective of actual producibility.

21 On the occasion of G. Gentzen's consistency proof for number theory (cnf. *Forsch. Log. Grundleg. exakt. Wiss.*, N. F. Heft 4, 1938) it was ascertained up to which ordinal number definitions and proofs by transfinite induction can be expressed in the formalism of classical number theory. (Cnf. D. Hilbert and P. Bernays, *Grundlagen der Mathematik*, Bd. 2 (1939), p. 360–374). Thereby it became evident that those which cannot be so expressed are not finitary, while, on the other hand, all finitary proofs can be represented as inductions with respect to certain ordinal numbers.

22 It may be argued that for all applications consistency up to some huge, but finite, number N of simple proof-steps is sufficient and that, moreover, from a strictly empirical standpoint the question of consistency is meaningful only up to some number N of proof-steps which is not absolutely beyond what can somehow be realized in our world. However, also in this weakened sense consistency cannot actually be proved by finitary reasoning, because the length of a finitary consistency proof in this sense would have to be of at least the same order of magnitude as N. Whether for each N there actually exist such limited consistency proofs of this length is an interesting question (cnf. footn. 39). (Note that, in order that the length of a proof, i.e. the number of simple proof-steps, be a true measure of the degree of complication, and for other reasons, it must be stipulated that each definition and each assertion occurring in a proof contain fewer than *K* symbols, where *K* is chosen once and for all.)

23 Cnf. P. Bernays, *Entretiens de Zurich,* 1938, ed. by F. Gonseth, 1941, pp. 144, 147; *L'enseignement math.,* 34 (1935, pp. 68, 69, 94; *Rev. Internat. Phil.,* No 27–28, 1954, Fasc. 1–2, p. 2; also: G. Gentzen, Trav. oln [?] IX. *Congr. Int. Phil.,* VI, p. 203 (published in: *Act. Sci. Ind.* 535, 1937). In the more recent papers of the formalistic school the term "finitary" has been replaced by "constructivistic", in order to indicate that it is necessary to use certain parts of intuitionistic mathematics which are not contained within the limits of that which is directly given in sensual intuition.

24 Thereby I mean universal propositions the special instances of which are ascertainable facts, such as, e.g., Goldbach's Conjecture.

25 E.g., the statement that the formalized system of classical analysis is consistent is a general proposition about ascertainable facts (concerning individual proof figures) which can be proved from transfinite axioms, but not finitarily.

26 Of course this result disproves assertion 1. in §5 only if some essentially non-finitary part of mathematics is consistent (and,

therefore, its consequences as to ascertainable facts are correct). But it is hardly possible to doubt the consistency of intuitionistic number theory and believe in that of finitary number theory. Moreover the [[empirical evidence for the consistency of classical mathematics is about as strong as that for some well established physical theory]] restriction imposed by this remark on the validity of the disproof of assertion No. 1 only means that it cannot be asserted with greater confidence than the theorems of classical or intuitionistic mathematics.

27 Finitary mathematics can be split up into a hierarchy of levels in such manner that the lower levels can be proved consistent in the higher ones.

28 E.g. the non-constructive concept of "there is" (referring to integers) can be replaced by "intuitionistic number-theoretical proof", or by "computable function of finite type", or by [["constructive accessible ordinal"]] "ordinal number $< \varepsilon_0$". Also the general concept of "set" can be replaced by that of "ordinal number" in conjunction with that of "recursive function of ordinal numbers". The concept of "integer" (with the axiom of complete induction) can be replaced by the concept of "set" (and its axioms). Finally it is not impossible that the non-constructive concepts of mathematics can be replaced by constructive ones, provided "constructivity" is taken in a sufficiently wide sense.

29 That this follows from a consistency proof was formulated as a conjecture by Carnap himself in the discussion at the *Deutsche Naturforschertagung* in Koenigsberg 1930. (Cnf. *Erk.* 2 (1930), p. 143.)

30 Without the second requirement a model always exists, even for inconsistent systems (on account of the fact that the sign of negation need not be interpreted by negation). But such models are of little interest, because they cannot replace the axioms under consideration as to their applications in other fields or in empirical science.

31 The question as to the existence of a content and as to the necessity of axioms (in the sense of footn. 2), of course, refers to mathematics as a system of propositions discerned (or posited) to be true, not as a hypothetico-deductive system. Some body of unconditional mathematical truth must be acknowledged, because, even if mathematics is interpreted to be a hypothetico-deductive system, still the proposition which states that the axioms imply the theorems must be unconditionally true. The field of unconditional mathematical truth is delimited very differently by different mathematicians. At least eight standpoints can be distinguished. They may be characterized by the following catchwords: 1. Classical mathematics in the broad sense (i.e., set theory included), 2. Classical mathematics in the strict sense, 3. Semi-Intuitionism, 4. Intuitionism, 5. Constructivism, 6. Finitism, 7. Restricted Finitism (cnf. footn. 19), 8. Implicationism (cnf. footn. 13). However, the conclusion that mathematics has content holds no matter which standpoint is taken. [[It can even be said that it appears most clearly in propositions of *this* type that mathematical theorems express objective facts. For while the definitions of 5, 7, 12 and the rules of computation for + and = (namely: $x + (y + 1) = (x + y) + 1$, the transitivity of =, etc.) seemingly can be interpreted to be conventions, the statement that $5 + 7 = 12$ follows from these conventions evidently expresses an objective (combinatorial) fact.]] It can even be said that, if the most restricted standpoint (implicationism) is taken, it can be seen most easily that mathematical theorems express objective facts. For the singular combinatorial facts concerned are unequivocally ascertainable relations between the primitive terms of combinatorics, such as: "pair", "equality", "iteration", and they can least of all be true by convention, because an inconsistency in the rules for the use of these terms can only be excluded by an intuition which has the meanings of the very same terms as its object (cnf. §32).

32 As to the meaning of "success" in the case of mathematical axioms cnf. §44 and footn. 44.

33 It seems arbitrary to me to consider propositions like "This is red" immediate data, but not so to consider the directly perceivable

mathematical facts mentioned. For, disregarding the greater precision and greater complication of the latter, the difference only lies in the fact that in the first case a relation between an undefined concept and an individual object is perceived, while in the second case it is a relation between two or more undefined concepts. Note that the axioms mentioned, although they do not follow from definitions, nevertheless can be classed as analytic propositions. For "truth owing to the meaning of the undefined terms" has exactly the same significance for undefined terms that "truth owing to the definitions" has for defined terms. It may be argued that these and similar axioms (cnf. footn. 45) should rather be called synthetic a priori, because a mere understanding of the concepts does not suffice for knowing their truth, but rather an intuition of the objects falling under them is necessary. However,

[[an intuition of objects which are *not* concepts (and which are infinite in number) can lead to a positive knowledge of general propositions only if infinitely many objects can somehow be apprehended in one glance (e.g. in one composite object) or if, for conceptual reasons, it is known that all the objects concerned are equal. Both these conditions perhaps are satisfied in geometry (the first one due to continuous motion), but not for the axioms of set-theoretic mathematics.]]

it can be answered that some kind of survey of all possible objects falling under it is implied in the complete understanding of a concept. At any rate the situation, for the axioms mentioned, is quite different from that prevailing, e.g., for the proposition: "Intuitive space is three dimensional". Here reference is made to a particular non-conceptual object (space) which is not defined in general terms, but directly given, so that this proposition in no sense follows from an understanding of the concepts occurring.

34 This answers the question raised in the end of footn. 13. "Objective truth" of axioms may be understood in two senses, namely to be: 1. the subsistence of the facts described in them, or 2. the verifiability of their consequences in the field of the directly observable, which, in the case of mathematics, e.g., may mean

verifiability in conjunction with certain simple laws of nature, such as the invariability of written characters or the reliability of short term memory. [[3. Assumption of objective truth may simply mean the decision to base one's expectations and actions on the axioms concerned.]] But, no matter which meaning of "objective truth" is chosen, provided only that there is to be any reason for expecting that the applications of mathematics will lead to correct assertions as to observable facts, a knowledge (with certainty or probability) of the objective truth of the mathematical axioms (or at least a large portion of them) cannot be dispensed with on the ground of any syntactical interpretation. For, unless the axioms used in the consistency proof are known to be true in either sense 1. or 2., the consistency proof is of no avail for drawing the conclusions sketched in §§ 13–17.

[[Moreover, if one takes the empirical standpoint, this assumption [the assumption of the "objective truth of the mathematical axioms"; or their "validity", according to another version] cannot consist in the syntactical convention that the axioms are to be true by definition. For this presupposes that the consistency of the axioms is known, since otherwise the axioms one day could be disproved, while it belongs to the essence of any convention, as opposed to a proposition with content, that it cannot be disproved (cnf. moreover requirement 4 on §11). But a knowledge of the consistency of the axioms can only be obtained with the help of mathematical intuition. If, on the other hand, the concept of convention is so enlarged as to comprise this case, the laws of nature also become conventions, so that mathematics as little, or as much, rests on syntactical conventions as natural sciences, and the distinction between *"Real-"*and *"Formal"-wissenschaft* in the sense of the quotation in footn. 9 loses every justification.]]

[[Carnap, it is true, has never asserted explicitly that reduction of mathematics to syntax means reduction to convention. However]] The passage quoted in footn. 9 seems to imply that mathematical truth, in Carnap's opinion, consists in nothing but conventions (cnf. also footn. 7). For, if there were in it some objective element (i.e. one not subject to our arbitrary choice), then this evidently would be the object of *Formalwissenschaft*. Hence the philosophical opinion explicitly rejected in the passage quoted would then be correct (unless, at least for some parts of mathematical truth, one advocates psychologism, or Mill's opinion, which are both rejected by proponents of the syntactical viewpoint).

It is to be noted that the arguments adduced in the text do much more than disprove the syntactical conception as formulated in the preceding paragraph. They show in addition that, in a certain sense, the content of mathematics is infinite. For there exist a practically unlimited series of axioms (cnf. footn. 45) each of which expresses some new and independent mathematical fact, since it cannot be reduced to a convention on the basis of the preceding axioms.

In order to disprove the assertion that mathematics has no objective content whatsoever (which would not be tenable even if mathematics were reducible to explicit definitions, cnf. §34), it is sufficient to point out that it is not in the least subject to our options, and therefore an objective fact, that such and such theorems are implied by such and such conventions. These facts are of a combinatorial nature (cnf. footn. 31), and to interpret them to be physical facts about marks on paper evidently means a return to Mill's view as to this particular kind of mathematical facts.

It may be added that, no matter whether the Platonistic or the syntactical view is taken, the truth of mathematical propositions is not a bit more conventional than that of factual propositions. For, also in mathematics, what is conventional is solely which symbols are associated with meanings, but, once these conventions have been made, the truth or falsehood of propositions is objectively determined. This follows from the fact that it is exactly by the rules of syntax that, according to the syntactical viewpoint, the meaning of the mathematical symbols is defined.

[[The foregoing proofs, it is true, only show that *finite* beings cannot interpret mathematics as a system of conventions (and human beings not even the mathematics of moderately large finite systems). However, if we were able to perform a transfinite number of mental acts, mathematics could be so.]]

There remains the possibility to see the voidness of content of *some* mathematical propositions, e.g. those referred to in the two preceding paragraphs, not in their conventional character, but in the alleged fact that, if the psychological difficulties of ascertaining their truth are disregarded, they all turn out to assert the same trivial thing, namely $p \supset p$ or something similar. However, this view clearly is psychologism. For if it were true that all those propositions assert $p \supset p$ in different ways, the substance of their content would be that by such and such combinations of concepts (ideas) $p \supset p$ is asserted. But this exactly would be a psychological fact. Cnf. also footn. 36.

The only way to eliminate all mathematical facts, without falling into psychologism or Mill's view, evidently would be to define the truth of every mathematical theorem by a separate convention. But in that case, of course, mathematics would lose all its interest and all its applicability. For also in the applications it is always a question of ascertaining some mathematical fact (cnf. §17). Mill's theory today is hardly maintained by anybody.

35 Neither classical nor intuitionistic mathematics, nor any system in which they can be proved consistent, nor even finitary mathematics, can be based on definitions alone, where "definition" means an always applicable rule for eliminating the symbol defined and "T" (truth by definition) and "F" (falsity by definition) are the only undefined terms. For this would yield a decision procedure for all propositions occurring.

36 The indispensableness of primitive terms and of axioms for them, also in logic, refutes the view that in logical inference the conclusion only reiterates tautologically part of the content of the premisses. For the very fact that the conclusion is contained (and asserted already) in the premisses is due to the meaning of the primitive terms of logic and, therefore, is an objective fact concerning (and specific for) these meanings.

Similarly the answer to the assertion mentioned in footn. 8 is that in order to understand these "conventions" and apprehend their possibility (i.e., their consistency) one either needs the concept of negation under some other name, or some other logical or combinatorial concepts with their axioms.

37 What induces to identifying "fact" and "synthetic fact" (i.e. to disregarding facts consisting in relations between concepts) is the circumstance that, due to definitions, relations between concepts apparently can be reduced to relations between logical concepts (cnf., however, footn. 43) and that the logical concepts in their regular use (i.e., as connectives and operators) somehow don't seem to belong to the subject-matter of the proposition, but rather to be the means by which something is said about the subject-matter. But, if a proposition is true due to nothing but properties of the means of expression, it cannot say anything about the subject-matter; but neither about the means of expression, or they would not be means of expression, but subject-matter. However, even if this were admitted, it would only prove that mathematical propositions in certain formulations express no facts, because these are hidden in the means of expression. But this would imply neither the non-existence nor the inexpressibility of mathematical facts. For mathematical concepts can also be made subjects of propositions.

38 I.e., in particular, apart from the loss of aesthetic values and the paralysing of mathematical invention consequential (at least under the schemes II and IV [for scheme IV, see footnote 39]) on disregarding the content of mathematical propositions.

39 Moreover, IV: All classical mathematics, at least theoretically, can be replaced by finitary syntax in a certain weakened sense. This is accomplished by dropping the requirement (contained in §12) that the logical rules of inference, too, should be derivable from the rules of syntax, and requiring only that this should be possible for each mathematical theorem. Then there do exist syntactical systems which can finitarily be proved consistent and which (provided classical mathematics is consistent) are equivalent with all classical mathematics (although this equivalence cannot be proved finitarily). They are obtained by stipulating that a sentence *A* is true if, for some *n*, *A* can be proved in *n* steps, but the negation of *A* cannot be proved in *n* or fewer than *n* steps. (This definition is due to L. Kalmar.)

If the question raised in footnote 22 were answered affirmatively, the truth, in this sense, of the theorems of mathematics could actually be proved by finitary reasoning in a feasible number of steps. This scheme would then be the closest existing approximation to a satisfactory syntactical foundation of mathematics. But still syntax [[would remain a very incomplete substitute for intuitive mathematics]] could not replace mathematics in the sense explained in footn. 5, because the inference illustrated by Goldbach's problem in §13 could, in general, only be drawn for integers below some limit *M*, although *M* could be enlarged successively. [[In fact, because of the degree of complication necessarily involved, it could, in general, not even be drawn for very great single numbers N. And this is true of every finitary syntax as to inferences about individual numbers drawn from transfinite axioms.]]
However, the scheme IV, combined with the treatment of empirical induction indicated in §17, opens up a possibility to set up a finitary syntactical system which can replace mathematics in the restricted sense explained in footn. 16. But this depends on an affirmative answer to the question raised in footn. 22 and, moreover, it is very likely demonstrable that in any finitary syntax which can replace transfinite mathematics in this restricted sense the syntactical considerations necessarily go far beyond restricted finitism (cnf. footn. 19 and §27). Therefore these syntactical schemes are open to the same objections as transfinite syntax (namely that they have to refer to purely mathematical objects which cannot at all be realized in physical or psychical reality).
Incidentally, it must not be said that questions regarding the actual feasibility of proofs are of no philosophical interest. For the main function of mathematics (as of every conceptual thinking) is to bring the vast manifold of particularities of the world under control. Therefore a substitute for mathematics that would fail at this point could not replace mathematics in one of its most essential aspects. If, e.g., in some syntactical system, in order to apply some mathematical theorem to the number of electrons in our world, it were necessary to perform a number of operations about equal to the number of electrons, this system certainly could not replace mathematics satisfactorily even for empiricists.

40 Note that item 2 is true only for mathematics at some given stage of its historical development and that, therefore, a definite syntactical interpretation also can be given only for mathematics in this sense (cnf. footn. 45 and 17).

41 The arbitrariness involved in Carnap's definition of content consists in the fact that it is based on the relation of logical consequence. I.e. logically equivalent sentences by definition have the same content. The content thus defined, therefore, should rather be called "empirical content". [[One could similarly define a "mathematical content" by using some other relation (e.g., "consequence in the calculus of quantifiers"); and a "logical content" by using the relation "equivalent by substitution of *definiens* and *definiendum* for each other".]]
By using other relations one could similarly define various kinds of "mathematical content" (corresponding to the various sections of mathematics), e.g., in Carnap's language II, a "set-theoretical content" by means of the relation of "consequence owing to the calculus of propositions and quantifiers". On the other hand, by using the relation of "consequence due to laws of nature" some kind of "historical content" could be defined relative to which laws of nature would be "void of content".
That it is arbitrary to call mathematics void of content because, without laws of nature, it has no verifiable consequences also appears from the fact that the same is true for laws of nature without mathematics or logic. Cnf. also §44.
[[If, on the other hand, content is defined, not by some relation of consequence, but by the *significata* of expressions, then, since in constructing a language one can choose freely what, if anything, one wants to associate to expressions as their *significatum*, one can make mathematical expressions void of content in this sense too. But the same can be done also for many factual propositions, in both cases without impairing the applicability of the propositions in question, since consequences from them can be derived formally. However the expectation that the application will not lead to wrong propositions about observables rests solely on a discernment of the content, or part of the content, of these propositions, or of other propositions about them.

This can be seen as follows: The conventions in question comprise an infinity of applications to special cases. Therefore they cannot be expressed by enumeration of all these cases, but must be formulated in general rules. The general concepts occurring in these rules, such as "substitution" or "juxtaposition", it is true, can be analysed further, but one necessarily ends up with certain undefined concepts.

Note that also by interpreting mathematics to be "physics of symbols" (cnf. II, §36) the necessity of a knowledge of primitive terms and facts concerning them (such as the consistency of the system) cannot be escaped. Only these terms and facts are transferred into the field of physics.

Moreover it is impossible that the primitive terms in any way can be created or constructed by ourselves. For the essence of any general concept consists in its relatedness to an infinity of special instances, namely in the fact that some of them fall under it and others do not, while everything which we (who are finite beings) can produce has a finite degree of complication. Therefore, whenever we construct some concept, we construct it out of other concepts. Hence the primitive concepts and the facts concerning them are objective at least in this sense that they are only perceived, not made, by us. Moreover they are objective also in the sense that they are not something merely psychological, because confronted with the same (or equal) objects we always make the same assertions, so that to differences in our assertions differences in the objects correspond. But this question does not concern us here since I am not engaged in disproving psychologism, but the syntactical viewpoint.]]

42 It might be objected that the analogy between mathematical intuition and the supposed additional sense breaks down insofar as the general laws holding for the supposed second reality could be disproved by further observations. However, the same would happen for mathematics if an inconsistency arose. For a disproval of observed laws also is nothing else but an inconsistency between different methods of ascertaining the same thing, since empirical induction and the application of laws of nature also are

such methods. The "inexhaustibility" of mathematics makes the similarity between reason and the senses (cnf. §39) still closer, because it shows that there exists a practically unlimited number of independent perceptions also of this "sense".

In reply to another possible objection it should be noted that, exactly as mathematics, the second reality, although implying nothing about the facts in the first reality, nevertheless might help us in knowing the latter, e.g., if it contained schematic pictures of all situations possible in the first reality.

43 What the organs of sense convey us is, strictly speaking, only the impressions of particular objects. For propositions such as "This is red" one already needs reason (although on a primitive level) in order to understand the general concept "red" and the copula "is". Mathematical knowledge (at least in the interpretation given to it by logicism, as opposed to Kantianism) is purely conceptual [[i.e., no particular objects occur in it]]. But such conceptual (analytic) knowledge may (and actually does) exist also outside mathematics, i.e., with regard to non-mathematical primitive terms.

It is clear that the "general", exactly as the "particular", is not created, but only perceived by us. For every general concept comprises a potential infinity of special instances of application, while everything we can produce is finite. Concepts, therefore, can only be obtained by the help of other concepts (even in case they are introduced by rules for the use of symbols). Therefore at least the primitive concepts cannot be constructed by us. But the defined concepts are only combinations of the primitive ones, and the fact that these combinations are meaningful and the meanings they have, again, are not within the reach of our constructions, but rather implied by the meanings of the primitive terms.

It should be noted that the direct perceptibility of certain mathematical objects is by no means the only reason for asserting their existence. Even empiricists, who do not accept this kind of perception (except perhaps for the sense qualities) as a source of knowledge, must nevertheless assert the existence of mathematical objects, if they consistently apply their own criteria of objectivity (cnf. §§ 43–46 and footn. 14).

44 "Success", within mathematics, of some new mathematical axiom would mean that many of its consequences could be verified on the basis of the former axioms, the proofs, however, being more difficult, and moreover, it would solve important problems not solvable before. Note that also the consistency of some new axiom, provided it yields a substantially stronger system, is indemonstrable on the basis of the preceding axioms and their consistency. But the undecidability, from the former axioms, of certain questions decidable by it might be so demonstrable.

45 What can be said today is this: In order to solve all problems of Goldbach type (cnf. footn. 24) of a certain degree of complication *k*, a system of axioms is necessary whose degree of complication is of at least the same order of magnitude as *k* (where the degree of complication is measured by the number of symbols necessary to formulate the problem – or the system of axioms–, either in primitive terms, or in defined terms if the number of symbols occurring in the definitions is also counted). Since it seems reasonable to assume that there exist only a finite number of primitive terms (i.e. terms whose meaning is understandable without definition) and that, moreover, the degree of complication of axioms (i.e. of propositions discernible to be true without proof) has an upper bound, it seems to follow that there exist propositions of Goldbach type which, for the human mind, are undecidable in an absolute sense. Their degree of complication, it is true, might be so great that they could not even be formulated within a reasonable length of time. However, it is to be noted that all of *present day* mathematics can be derived from a handful of very simple axioms about a very few primitive terms. Hence the problems undecidable *from these axioms* can easily be formulated. Therefore, even if solutions are desired only for all those problems which can be formulated in a few pages, mathematics will have to change completely its character; namely the few axioms being used today will have to be supplemented by a vast number of new ones. This is not entirely impossible. In fact the change required would not be greater than was that from Greek to modern mathematics.

Moreover there do exist unexplored series of axioms which are analytic in this sense that they only explicate the content of the concepts occurring in them (cnf. footn. 33), e.g., the axioms of infinity in set-theory, which assert the existence of sets of greater and greater cardinality. However, if modern standards of mathematical rigour are applied, the concepts concerned, e.g., the general concept of set, cannot be considered sufficiently precise or constructive to serve as the basis for unimpeachable mathematical proofs. That there should exist sufficiently powerful series of axioms not based on some rather questionable concept is not very likely. Therefore, even if there do exist enough analytic axioms for the solution of all problems characterized above, still a more or less inductive character probably will have to be assigned to them.

IS MATHEMATICS SYNTAX OF LANGUAGE?

by Kurt Gödel

It is well known that Carnap has carried through, in great detail, the conception that mathematics is syntax of language. However, not enough attention has been paid to the fact that the philosophical assertions which form the original content and the chief interest of this conception have by no means been proved thereby. Quite on the contrary, this, as well as any other possible, execution of the syntactical scheme rather tend to bring the falsehood of these assertions to light. I am speaking of the following assertions:

I. Mathematical intuition, for all scientifically relevant purposes, in particular for drawing the conclusions as to observable facts occurring in applied mathematics, can be replaced by conventions about the use of symbols.

II. In contradistinction to the other sciences, which describe certain objects and facts, there do not exist any mathematical objects or facts Mathematical propositions, because they are nothing but consequences of conventions about the use of symbols, are compatible with all possible experiences, and, therefore, are devoid of any content.

III. The conception of mathematics as a system of conventions makes the a priori validity of mathematics compatible with strict empiricism. For we know a priori, and without having to appeal to any a priori intuition, that conventions about the use of symbols cannot be disproved by experience.

It seems to me that these three assertions are refutable, as far as any philosophical assertion can be refutable in the present state of philosophy. That this is possible only in a limited sense follows from the fact that the philosophical, as well as the other very general , terms occurring in philosophical assertions are not well defined and

8. Essay on Carnap, version VI. First folio.

# 4

# Is mathematics syntax of language?, VI

It is well known that Carnap has carried through, in great detail, the conception that mathematics is syntax of language. However, not enough attention has been paid to the fact that the philosophical assertions which form the original content and the chief interest of this conception have by no means been proved thereby. Quite on the contrary, this, as well as any other possible, execution of the syntactical scheme tend to bring the falsehood of these assertions to light. I am speaking of the following assertions:

I. Mathematical intuition, for all scientifically relevant purposes, in particular for drawing the conclusions as to observable facts occurring in applied mathematics, can be replaced by conventions about the use of symbols.

II. In contradistinction to the other sciences, which describe certain objects and facts, there do not exist any mathematical objects or facts. Mathematical propositions, because they are nothing but consequences of conventions about the use of symbols, are compatible with all possible experiences, and, therefore, are devoid of any content.

III. The conception of mathematics as a system of conventions makes the a priori validity of mathematics compatible with strict empiricism. For we know a priori, and without having to appeal to any a priori intuition, that conventions about the use of symbols cannot be disproved by experience.

It seems to me that these three assertions are refutable, as far as any philosophical assertion can be refutable in the present state of philosophy. That this is possible only in a limited sense follows from the fact that the philosophical, as well as the other very general, terms occurring in philosophical assertions are not well defined and admit of various interpretations. In particular this applies to the terms occurring in I, II, III above, namely: "replacing", "content", "fact", "disproving". Therefore, what can be shown is only that those meanings of the terms which make the assertions in question true are artificial and widely divergent from the original ones, or even that, by a kind of *petitio principii*, the answers to the questions are anticipated by definitions which deal differently with analogous situations.[1] The purpose of the following considerations is to show this.

I. 1. As far as assertion I is concerned, it is true that by applying certain rules about the use of symbols (to be more precise rules stating conditions under which propositions containing them are true) one arrives at the same results (also in the applications) as by applying mathematical intuition. However, in order to have any reason for the expectation that, if these rules are applied to verified laws of nature (e.g., the primitive laws of elasticity theory), one will obtain empirically correct propositions (e.g., about the carrying-power of a bridge), one must know (at least with probability) that the syntactical rules ... really are ... [?] consistent [?] i.e. compatible with all empirical facts... [?] [[ For on the basis of inconsistent rules, *all* propositions]], the empirical ones included, can be derived. But now it turns out that for proving the consistency of mathematics an intuition of the same depth (although possibly of a different kind) is needed as for recognizing the truth of the mathematical axioms.

In particular the consistency of the *abstract* mathematical concepts, such as "infinite set", "function", etc., cannot be proved consistent without again using abstract concepts, i.e., such as are not ascertainable properties or relations of finite combinations of symbols or other concrete objects. If for the proof of the consistency or admissibility of the conventions only such concepts and insights were required as are necessary in order to understand and handle symbolic conventions, then assertion I would in a sense be true. However, exactly this is demonstrably not the case. Consistency, it is true, can also be inferred empirically, but at any rate it is clear that, with an adequate meaning of the term "replace", mathematical intuition cannot be replaced by conventions, but only by conventions plus mathematical intuition or by conventions plus an empirical knowledge involving in a certain sense an equivalent mathematical content.

2. There remains the possibility to interpret the word "can", not in a realistic (practical), but rather in an idealized sense, e.g., by mak-

ing the fiction that we can mentally run through any finite totality, no matter how large, or even through infinite totalities. As to such schemes it is to be noted that the whole content and purpose of mathematics is to master arbitrary finite and infinite totalities and structures. It is therefore not surprising that mathematical intuition can be replaced by an immediate and complete knowledge of those totalities. All factual propositions likewise can be transformed into consequences of conventions about the use of symbols, if we assume that we can run through the totality of all objects in the word. For in that case every concept could be defined by enumeration of all objects falling under it. These "conventions" about the use of words, together with those about the use of the logical terms evidently would entail everything else.[2]

II. The fact that, in order to draw mathematical inferences in the empirical sciences, something more is necessary than conventions and their application points to the existence of a content of mathematics. Still more clearly its existence results from the following arguments:

1. If it is said that mathematical propositions have no content because nothing follows from them about experiences, it can be answered that the same is true of laws of nature. For laws of nature without mathematics or logic imply as little about experiences as mathematics without laws of nature. That mathematics, at least in some cases, does add something to the laws of nature which is not expressed in them, is best seen from examples where one has very simple laws about certain elements (e.g., those about the reactions of an electronic tube). Here mathematics adds the general laws of multiplicity, i.e. the laws as to how many tubes connected in a certain manner will react. That the latter laws are not contained in the former is seen from the fact that 1. they contain concepts not definable in terms of those occurring in the former, and that 2. new empirical inductions may be necessary in order to ascertain the latter laws, namely in case the mathematical problems in question are unsolvable. E.g. this may occur in a case like Goldbach's conjecture, which evidently implies a certain law about the reactions of a computing machine. Mathematical propositions, it is true, do not express physical properties of the structures concerned, but rather proper-

ties of the *concepts* in which we describe the structures. But in view of the example given this only shows that the properties of those concepts are something quite as objective and independent from our choice as physical properties of matter.[3]

It may be added that, unlike in the example given above, the *general* laws about the interaction of many elements may even be required for predicting the result of a *single* observation, namely in case the latter depends on an infinity, e.g., a continuum, of physical elements.

2. Mathematical axioms, in principle, are disprovable exactly like laws of nature, namely by an inconsistency derived from them. But it seems that propositions which may be wrong must have some content. If a contradiction is not acknowledged as a disproval, but only as a proof for the "inexpediency" of the conventions, the same can be done for laws of nature, which also can be interpreted to be conventions which become "inexpedient" in case a counterexample is found. It should also be noted that an inconsistent mathematical axiom, before the inconsistency is discovered, would work out in the application exactly as a wrong law of nature, since it would imply also wrong empirical propositions. From the disprovability of mathematical axioms it follows that, if mathematics is dealt with from the positivistic point of view, regarding "existence", without prejudice, "existence" should be attributed to the objects of a successful mathematics (in contradistinction to those of one contradictory) exactly as is done for the objects of a successful physics. If the possibility of a disproval of mathematical axioms is frequently disregarded, this is due solely to the convincing power of mathematical intuition. But the very starting point of the syntactical conception is the rejection of mathematical intuition. Even for consistent mathematical axioms there is a possibility of disproval, namely by the fact that they lead to wrong predictions on the basis of well verified laws of nature.[4]

3. Even if it were admitted that mathematics can be based on conventions about the use of symbols, its voidness of content still would not follow. For symbolic conventions are void of content only in so far as they *add* nothing to the theory in which they are made, but they may very well imply propositions of this theory. If, e.g., on the ground of the empirically known associativity of some physical op-

eration a convention about the dropping of brackets is introduced, then from this convention the associativity of the operation in question, i.e., an empirical proposition, follows. If a mathematical convention is introduced on the basis of its consistency, the situation is quite similar. For this fact of the consistency of the convention, again, is expressible in the main system in which it is made and the convention implies, although not this consistency itself, still certain only slightly weaker propositions, i.e., substantially the same facts as those which justified its introduction.[5]

4. Even if mathematics is built on rules of syntax, this makes it not a bit more conventional (in the sense of "arbitrary") than other sciences. For according to the positivistic point of view the rules for the *use of a symbol are the definition* of its meaning, so that different rules simply introduce different meanings, i.e. different concepts. But the choice of the concepts is free also in other sciences. Moreover syntactical rules, which introduce new symbols not as mere abbreviations for combinations of symbols present already, must be consistent [?] and compatible with all empirical possibilities and, therefore, are *very* far from arbitrary.[6]

5. There exist experiences, namely those of mathematical intuition, in which we perceive mathematical objects and facts just as immediately as physical objects, or perhaps more so. It is arbitrary to consider "this is red" an immediate datum, but not so to consider *modus ponens* or complete induction (or perhaps some simpler propositions from which the latter follows). For the difference, as far as it is relevant here, consists solely in the fact that in the first case a relationship between a concept and a particular object is perceived, while in the second case it is a relationship between concepts. Exactly as mathematical intuition is dealt with in the syntactical conception of mathematics some physical sense could also be dealt with, provided it were sufficiently separated from the other senses. I.e., one could disregard the impressions of this sense and not associate with them any objects or facts, but rather introduce the propositions concerned by "conventions about the use of symbols". A difficulty would arise, if the mutually independent impressions of this sense were so numerous, that they would make it necessary constantly to introduce new conventions. Actually, however, a somewhat similar sit-

uation does subsist for mathematics. For it turns out that, in order to solve its problems, it is necessary again and again to introduce new axioms, which can be justified only by intuition or experience. One might say that, in contradistinction to other sciences the experiences of mathematical intuition are not the object of mathematics. However, truly speaking, experiences themselves are not the object of most other sciences either. E.g., [[animals seen in hallucinations are not an object of Zoology]] color sensations are not the object of physical optics.[7]

III. As to assertion III it suffices to say that, if consistency and compatibility with [empiricism?] (which must be known in order to be able to introduce the mathematical axioms as "conventions") is based on empirical induction mathematics is not a priori true; on the other hand to prove it by mathematical intuition is not compatible with empiricism.[8]

What psychologically plays a large part in any plausibility which the thesis of the voidness of content of mathematical propositions may have are these two circumstances:

1. The logical concepts in their use in empirical propositions don't seem to belong to the subject matter of the proposition, but rather seem to be means of expression. However, if a proposition is true already due to properties of the means of expression, it cannot say anything about the subject matter. Nor can it say anything about the means of expression. For, if it did, they would not be means of expression, but subject matter.

2. No possibility is excluded by a logically true proposition, while the content of a proposition seems to consist exactly in the fact that it excludes certain possibilities. However, the first argument, even if its antecedent is admitted, does not exclude that the logical concepts may be made the subject matter of non-empirical propositions. As to the second argument it can be answered that there are different levels of possibility.[9]

# Editorial footnotes of comparison with version V

1 In the fifth version the general argument is similar, but Gödel concedes to the syntactical view "the merit of having pointed out the fundamental difference between mathematical and empirical truth", and this difference is admitted to be located "in the fact that mathematical propositions, as opposed to empirical ones, are true in virtue of the *concepts* occurring in them", with which Gödel seems to accept that mathematical propositions are, in a certain sense, analytic. What he denies is the nominalistic step of identifying concepts with terms, or symbols, in order to transform "mathematical truth into conventions and, eventually, into nothing".

2 The fifth version is a little bit clearer for we are told that the only possible rules to be used in the syntactical attempt are those which are "admissible" – non-arbitrary – rules, and admissibility entails consistency, which requires intuition. This "vicious circle" is described by writing that no matter which syntactical rules are formulated, "the power and usefulness of the mathematics resulting is proportional to the power of the mathematical intuition necessary for their proof of admissibility", whose phenomenon is called by Gödel "the non-eliminability of the content of mathematics by the syntactical interpretation". Ultimately, this means that mathematical truth has to be acknowledged to be non-conventional.

3 The corresponding argument in the fifth version is considerably bolder, as it adds the following Platonic words: "This is not surprising, since concepts are composed of primitive ones, which, as well as their properties, we can create as little as the primitive constituents of matter and their properties. However, in spite of the objective character of conceptual truth, it is quite necessary to distinguish sharply these two kinds of content and facts as 'fac-

tual' and 'conceptual'. What Carnap calls 'content' really is 'factual content'".

4 The language of the fifth version is more philosophical. We are told about "immanent existence" as being attributed by the positivists to the "correct mathematics". Moreover, Gödel adds the following strong assertion: "these mathematical objects and facts [e.g. infinite sets or properties of properties] cannot be eliminated (as, e.g., infinite points in geometry can), since there always remain primitive mathematical terms and axioms about them, either in the scientific language or the metalanguage, where 'axiom' here means a proposition assumed on the ground of its intuitive evidence or because of its success in the applications". Therefore, Gödel concludes, the positivistic point of view as for mathematical objects is inconsistent when the positivists speak about their "pseudoexistence"; the difference with physical objects "primarily lies in their different *intuitive character*", but their roles in the formalism is similar, so they could be regarded as "irreducible hypotheses of science, exactly as the assumption of a field and of the laws governing it".

5 The fifth version is again clearer, as it is, philosophically speaking, more ingenuous. Gödel starts by saying that conventions on symbols are void of content only relatively to the sense in which "they *add* nothing to a theory which implies their admissibility". The argument depends on the comparison with empirical laws as well, but here Gödel adds that any law of nature can as well be interpreted as "a convention whose admissibility derives from this law of nature". There are two possible objections. First, that in this way mathematical axioms – and thus the convention in question – are not sufficient to derive the empirically verifiable fact of consistency. Gödel's answer is that "nobody will call a law of nature (...) void of content, because it has verifiable consequences only in conjunction with other, independently known, laws". Second, that the statement of consistency itself is void of content because all instances of individual proof figures can be derived from the axioms before the convention in question has been made. Now

the answer is one of Gödel's strongest assertions about the similarity between mathematics and empirical science: "the process of formal derivation in a theory is itself a kind of observation. So the objection is about the same as if it were said that a law of nature is void of content, because the single instances of it can be ascertained without its help, namely by direct observation".

6 An extension of this argument appears in the fifth version. It might sound as belonging to a rather formalist philosophy of mathematics, as it is based on the "fact" that definitions in some way create their object, but this impression is complemented by the insistence on the similarity between mathematics and empirical science. Gödel writes that "what can be asserted on the basis of the definitions is exactly as objectively determined in mathematics as in other sciences. Viewed from this angle the content of mathematics appears in the fact that definitions implicitly assert the existence of the object defined. In particular if a symbol is introduced by stating rules as to which sentences containing it are true, then from these rules much the same conclusions can be drawn as could be from the assumption of the existence of an object satisfying those rules".

7 In the fifth version the parallelism between mathematics and empirical science is emphasized, as it is pointed out that the syntactical program is possible only because mathematical propositions follow from a small number of primitive ones, and they are separable from other propositions because no empirical propositions follow from them. Thus, there is a great similarity between mathematical intuition and empirical perception, for if the objects of physical sense were so regular and separated from those of the other senses, "we could interpret also the propositions based on impressions of *this* sense to be syntactical conventions without content and associate no facts or objects with them or their constituents". Gödel arrives even at the assertion that there is no substantial difference between mathematics and other sciences for "a general mathematical theorem, in a sense, has the mathematical experiences relating to the special cases as its object".

8 The fifth version is again clearer, now because it resorts to admissibility. Thus, if the syntactical rules we introduce to replace mathematics are non-admissible, certain empirical consequences are derivable, which can be disproved by experience. If, on the other hand, our rules are admissible (then consistent), any a priori knowledge of this fact can only be obtained by mathematical intuition.

9 In the fifth version the last sentence is somewhat clarified, as a distinction between physical and logical possibility is added. Also, there was a third item dealing with the fact that certain sentences are determined by semantical rules to be true under all circumstances (e.g., "It will rain or it will not rain tomorrow"), which are usually admitted to be void of content. Thus, they cannot be denied because of their structure, or because of the meanings of the terms occurring in them, which could be interpreted by saying that logic is a part of syntax. Gödel's answer is, however, that "what is regarded as the content of a proposition largely is a question of what one is interested in. E.g., one may very well say that the proposition mentioned above, although it says nothing about rain, does express a property of 'not' and 'or'".

# Index