Logic in Philosophy of Mathematics

Hannes Leitgeb

1 What is Philosophy of Mathematics?

Philosophers have been fascinated by mathematics right from the beginning of philosophy, and it is easy to see why: The subject matter of mathematics—numbers, geometrical figures, calculation procedures, functions, sets, and so on—seems to be abstract, that is, not in space or time and not anything to which we could get access by any causal means. Still mathematicians seem to be able to justify theorems about numbers, geometrical figures, calculation procedures, functions, and sets in the strictest possible sense, by giving mathematical proofs for these theorems. How is this possible? Can we actually justify mathematics in this way? What exactly is a proof? What do we even mean when we say things like 'The less-than relation for natural numbers (non-negative integers) is transitive' or 'there exists a function on the real numbers which is continuous but nowhere differentiable'? Under what conditions are such statements true or false? Are all statements that are formulated in some mathematical language true or false, and is every true mathematical statement necessarily true? Are there mathematical truths which mathematicians could not prove even if they had unlimited time and economical resources? What do mathematical entities have in common with everyday objects such as chairs or the moon, and how do they differ from them? Or do they exist at all? Do we need to commit ourselves to the existence of mathematical entities when we do mathematics? Which role does mathematics play in our modern scientific theories, and does the empirical support of scientific theories translate into empirical support of the mathematical parts of these theories?

Questions like these are among the many fascinating questions that get asked by philosophers of mathematics, and as it turned out, much of the progress on these questions in the last century is due to the development of modern mathematical and philosophical logic.

2 Logic in Philosophy of Mathematics

The pioneer of both modern logic and modern philosophy of mathematics was the German mathematician and philosopher Gottlob Frege (1848–1925).[1] On the one hand, Frege devised the very first formal language in which various mathematical theorems could be formulated in absolutely precise and non-ambiguous terms, and the very first formal system in which much of the reasoning of mathematicians could be carried out in a way that made it possible in principle to check mechanically whether a sequence of statements was a proof or not. On the other hand, Frege tried to show—by proof again—that much of classical mathematics could actually be reduced to logic alone. This view of mathematics is called *Logicism*: First, Frege would define mathematical concepts such as *natural number*, 0, *less-than* (for natural numbers), and so on, on the basis of purely logical concepts such as $\neg$ (not), $\vee$ (or), $\rightarrow$ (if-then), $\exists$ (there exists), $\forall$ (for all), $=$ (identity), and the like. Secondly, once all the mathematical concepts in a mathematical theorem had been replaced by the logical concepts that defined them, he would derive the theorem by purely logical rules of inference from purely logical axioms.

While Frege's work led to incredible progress in logic and the philosophy of mathematics, the ultimate formal system that he worked with happened to be inconsistent—a contradictory statement such as $A \wedge \neg A$ could be derived in it, by first using some of Frege's axioms in order to prove the existence of a set X of all sets that do not include themselves as members, and then deriving a contradiction from the observation that X is a member of itself if and only if it is not a member of itself. This was pointed out to Frege by the British philosopher Bertrand Russell (1872–1970), who himself—together with Alfred North Whitehead —became famous through their monumental *Principia Mathematica* in which they tried to reduce mathematics to logic again, but this time without any contradictory conclusions. Modern set theory, which is still part of mathematical logic, and which had a tremendous impact on all areas of modern mathematics by becoming at the same time the universal language and the foundational system of axioms for mathematics, did not exactly follow the logicist lines of Frege or Russell and Whitehead, but at least it became clear that more or less all mathematical concepts could be reduced to combinations of logical concepts and the concept of set

[1] There is a very nice entry by Zalta (2008) on Frege in the *Stanford Encyclopedia of Philosophy* which is freely accessibly at http://plato.stanford.edu in the worldwide web. Accordingly, check out Horsten's (2007) entry on Philosophy of Mathematics there.

membership ($\in$), and more or less all known mathematical theorems could be derived from logical axioms in combination with the axioms of set theory, that is, the axioms governing $\in$.[2] However, most philosophers of mathematics today consider the concept of membership as properly mathematical, rather than purely logical, and some of the axioms of set theory are no longer counted as logical axioms either: for instance, the set-theoretic axiom of infinity, which postulates the existence of an infinite set, is now taken to be a properly mathematical axiom rather than an axiom of pure logic, since it is part of our modern conception of logic that logic ought to be neutral or silent with respect to all questions of existence.[3]

In similar ways, all modern schools in the philosophy of mathematics owe a lot to logical ideas, logical concepts, and logical results.[4]

For instance, classical mathematics' once great rival *Intuitionism* (just as the closely related school of *Constructivism*) rejects the logical law of the excluded middle ($A \vee \neg A$) and demands generally for the proof of an existence claim $\exists x P(x)$ a procedure by which an instance $P(a)$ can be actually constructed or determined. The development of formal systems of intuitionistic logic, intuitionistic arithmetic, and constructive set theory, and their comparison with formal systems of classical logic and mathematics, all of which being topics of mathematical logic, certainly added significantly to the understanding of this view of mathematics, even though the founder of Intuitionism, the Dutch mathematician Luitzen E.J. Brouwer (1881–1966), deliberately understood mathematics in informal and non-logic-oriented terms.

More recently, *Structuralism*, which takes the existence and properties of

[2]Recently, *category theory* has been challenging the role of set theory as being the universal framework for mathematics. It is hotly debated whether there are parts of modern mathematics that go beyond set theory but not beyond category theory.

[3]Since the 1980s, *Neo-Logicists/Neo-Fregeans* have revived some of Frege's original ideas on Logicism by trying to reduce mathematics to principles of second-order logic and so-called abstraction principles. Second-order logic—in contrast with first-order logic, which is the logic of quantifier expressions $\exists x$ and $\forall x$ that speak about individual objects x—includes logical axioms and rules for quantifier expressions $\exists P$ and $\forall P$ which quantify over sets P (or properties P or concepts P, depending on one's favourite interpretation). It is still a matter of controversy whether second-order logic is proper logic or whether it is really set theory "in disguise", as W.V. Quine maintained. Abstraction principles are principles which state the identity conditions of objects of a certain kind, such as numbers. See Hale and Wright (2001) for the most extensive Neo-Fregean reconstruction of mathematics to date.

[4]There are entries in the *Stanford Encyclopedia of Philosophy* on all of these schools. The standard collection of articles on all of them is Benacerraf and Putnam (1983).

mathematical individuals to be derivative from the existence and properties of structures—so that e.g. the nature of the natural number 2 is exhausted by the fact that it comes third in the successor relation $0-1-2-3-4-\ldots$—has traded on logical methods. For instance: While first-order axiomatisations of arithmetic are provably incapable of excluding unintended interpretations of the symbols of arithmetic[5], there are second-order axiomatisations of arithmetic—most famously, the system of the second-order (Dedekind-)Peano axioms—which are provably categorical in the sense that every two models of such a system of axioms are isomorphic to each other, that is, have the same structure. This categoricity theorem, which was proven by the German mathematician Richard Dedekind already in the 19th century, and which may be called a model-theoretic theorem (model theory being a part of logic again), explains therefore how and why we might be able to get epistemic access to mathematical structures such as the one of the natural numbers: by their categorical second-order axiomatisations (see Shapiro 1997 for more on this).

Similarly, *Nominalism* about mathematics, according to which there are no abstract objects at all even though mathematical theories may still be used to shorten longish purely logical derivations of statements about the physical world from other statements about the physical world, experienced a much richer, more precise, and much more sophisticated revival by means of logical theorems than it had ever had before in the traditional metaphysical debates between Realists and Nominalists: Hartry Field's *Science Without Numbers* (Field 1980), which formulates physical theories without reference to numbers on the basis of representation theorems that were once proven in what one might now call the model theory of geometry, is the most famous example.

But without doubt the most drastic impact that a logical result ever had on a school in the philosophy of mathematics is the impact that Kurt Gödel's (1931) famous Incompleteness Theorems[6] had on *Formalism*, which

[5]There is a whole branch of mathematical logic which deals with such non-standard models of arithmetic or with non-standard models of mathematical theories more generally.

[6]Boolos, Burgess, and Jeffrey (2002) is a classical introduction to the theorems. Smith (2009) is a recent and detailed reconstruction of the Incompleteness Theorems which is accessible also to non-mathematicians. Hájek and Pudlák (1991) is recent mathematical treatment of formal systems of first-order arithmetic in general. Raatikainen (2005) and Torkel Franzén's *Gödel's Theorem. An Incomplete Guide to Its Use and Abuse* (Franzen 2005) give excellent surveys of the philosophical significance of the Incompleteness Theo-

is the topic of the next section.

3 Formalism and Formal Systems

David Hilbert (1862–1943), one of the most famous mathematicians of his time and the most important proponent of Formalism ever, only accepted a particular fragment of the arithmetic of natural numbers—call it *elementary arithmetic*—as being completely beyond doubt, in light of our immediate intuitive grasp of natural numbers as sequences of strokes that one could manipulate by erasing or adding strokes according to elementary rules.[7] The rest of mathematics was acceptable to Hilbert only as some kind of formal game with symbols, where the meaning that got ascribed to these symbols in the course of the game was not relevant at all. Instead, the rules of such mathematical games were to be formulated by purely syntactical means and this had to be done as precisely as Frege once had formulated the rules of his system of logic, or otherwise the rules of the game would remain unclear. Once again it should be possible in principle to check for each sequence of formulas—for each sequence of moves in the game—whether the sequence was a proof according to the rules of the game or not. So, first of all, all the mathematical theories that one cared about needed to be reconstructed in terms of formal systems.

Any such a formal system $\mathcal{S}$—which today is called also 'recursively axiomatised theory'—is given by:

1. A *formal language* $\mathcal{L}_{\mathcal{S}}$ which is specified by listing the syntactic rules by which the formulas of the language can be generated in a systematic manner. For instance, since $2 + 2 = 4$ is a formula of the language of elementary arithmetic, also $\neg\, 2 + 2 = 4$ (it is not the case that $2 + 2 = 4$) is a formula of the same language. Of course, we regard $2+2 = 4$ as true and $\neg\, 2+2 = 4$ as false, since we automatically supply these symbols with a particular intended interpretation, but that does not actually matter at this point—it only matters that both are well-formed expressions. Similarly, if the constant symbol 4 is replaced by the variable x, if 2 gets replaced by y, and if before the resulting formula

rems. See also Jeremy Avigad's article on "Proof Theory" in this volume.

[7] It is an open question which axiomatised fragment of arithmetic exactly Hilbert regarded as mathematically and philosophically unproblematic: whether first-order Peano Arithmetic or the system now called Primitive Recursive Arithmetic or something else.

$y + y = x$ we put the quantifier expressions $\forall x \exists y$, we end up with yet another formula of the language of elementary arithmetic: $\forall x \exists y\, y+y = x$ (for every natural number x there exists a natural number y, such that $y + y = x$). According to our intended interpretation of arithmetical signs, this formula says that every natural number x is an even number, which is, obviously, false, but again the formula nevertheless belongs to the language of elementary arithmetic. And again, as far as the formal system is concerned, this intended interpretation is irrelevant.[8]

These are the symbols that one can use in order to build up a formula of one particular language—as, e.g., the language of elementary arithmetic: constant symbols for particular natural numbers (0, 1, 2,...), variables for numbers (x, y,...), arithmetical function signs ($+$, $\times$), arithmetical relation symbols ($=$, $<$), and logical symbols. If one wanted to speak about natural numbers in terms of a more expressive language, then one could always add further symbols. For instance, while we will only deal with first-order languages of arithmetic below, in which $\forall$ and $\exists$ quantify solely over natural numbers, we could have turned to second-order languages of arithmetic in which $\forall$ and $\exists$ may also be used to quantify over *sets* of natural numbers, and so on. In any case, one usually assumes that the language of a formal system contains only countably many formulas, and that a computer is in principle capable of listing all these formulas once the syntactic rules of the formal system have been turned into a corresponding computer program.[9]

2. The set of *axioms* of $\mathcal{S}$, i.e., a set of formulas of $\mathcal{L}_\mathcal{S}$. If there are to be just finitely many axioms, then these axioms may simply be listed explicitly. Otherwise, if there are to be infinitely many axioms, then one may state a procedure by which one would be able to determine in finitely many steps whether a given formula of $\mathcal{L}_\mathcal{S}$ is to be counted as an axiom or not. In either case, the set of axioms of a formal system

[8] But whenever one has some interpretation of a formula in the language a formal system in mind, and the formula is such that it is true or false relative to that interpretation, one often says 'statement' or 'sentence' instead of 'formula'.

[9] Of course, in the heyday of Hilbert's Formalism, computers in the modern sense were not available as yet. In fact, logical investigations that were closely related to formalist views of mathematics and to the Incompleteness Theorems constituted a significant source of both inspiration and theoretical support for the emergence of modern computers and the development of computer science.

is assumed to be a *decidable set*, that is, a set of formulas in a formal language for which a computer program could be written, such that the program would output 'yes' if it received as an input a formula that is a member of the set of axioms, but which would otherwise output 'no'.[10]

In standard formal systems one finds two types of axioms: *logical axioms* and *eigenaxioms*. The former express general logical laws, whereas the latter determine the particular (in our case, mathematical) content of the formal system or theory $\mathcal{S}$. For instance, the formal system of elementary arithmetic contains both logical axioms and arithmetical axioms: $\forall x\, x = x$ is among the former, whereas $\forall x\, x + 0 = x$ is among the latter.

3. The set of *logical rules* (rules of inference) of $\mathcal{S}$ which is specified again by listing the rules explicitly or by stating a procedure by which one is able to determine within finitely many steps what counts as a rule of the system. Each rule is of the form

$$\frac{\begin{array}{c} A_1 \\ A_2 \\ \vdots \\ A_n \end{array}}{B}$$

where $A_1, A_2, \ldots, A_n$ are called the *premises* of the rule and B is called its *conclusion* (the line indicates the transition from the premises to the conclusion). In typical formal systems, the rules are required to be logically valid, i.e., strictly truth-preserving. For example, Modus Ponens,

$$\frac{\begin{array}{c} A \\ A \to B \end{array}}{B}$$

[10] According to the traditional view on axioms which we inherited from the ancient Greek, only such sentences were to be chosen as axioms which one could see to be true by mere inspection, that is, which were not in need of proof. However, according to the modern terminology, 'axiom' does not have any such epistemic connotation anymore at all.

is typically chosen to be among the rules of a formal system.[11]

4. The set of *theorems* of $\mathcal{S}$ which is determined by the specifications of the language, the axioms, and the rules of $\mathcal{S}$: a formula A of $\mathcal{L}_{\mathcal{S}}$ is defined to be a theorem of $\mathcal{S}$ if and only if A is derivable from the axioms of $\mathcal{S}$ by means of applications of the logical rules of $\mathcal{S}$. In other words: there is a formal derivation (formal proof) that starts from the axioms of $\mathcal{S}$ and which then "leads" to A by means of finitely many applications of rules of $\mathcal{S}$. Accordingly, we say that A is *disproven* in $\mathcal{S}$ if and only if $\neg A$ is derivable in $\mathcal{S}$. These rather vague definitions may actually be turned into exact and purely mathematical definitions. In particular, there are no "practicality" conditions built into the concept of *formal derivability*: for instance, even if the shortest derivation of a formula in a formal system takes more steps than there are particles in our physical universe, the formula still counts as derivable in the formal system.

On the one hand, such formal systems may be regarded as potential answers to the question: What is your theory? Famously, Euclid was the first to suggest that scientific disciplines—in his case: Euclidean geometry—should be built up deductively by axioms and rules (but of course his axiomatisation of geometry was not quite a formal system in our modern sense yet). Later, Newton gave a "Euclid-style" axiomatisation of mechanics. In the 19th and 20th century, various mathematical disciplines were reconstructed successfully in terms of more or less precisely defined formal systems: geometry (David Hilbert), arithmetic (Richard Dedekind, Giuseppe Peano), set theory (Ernst Zermelo, Abraham Fraenkel, Albert Thoralf Skolem), and so forth. Further down below, we will deal mainly with formal systems of first-order arithmetic.

On the other hand, a formal system may be viewed as determining a computer with a particular kind of computer program (or, more formally speaking, a so-called Turing machine): For the theorems of a formal system can be enumerated by a computer program as follows. Let a procedure generate systematically all finite sequences of formulas in the language of that formal system, which is easy to implement. By the definition of a formal system, for each formula in any such sequence it can be determined in finitely

[11] In systems of so-called natural deduction, logical rules are permitted to have a much more complicated form than the one presented here. But we concentrate only on rules in the sense of Hilbert.

many steps whether the formula is an axiom of the system or whether it is the result of applying the rules of the system to some of the formulas that occur earlier in that sequence; if either of these two properties apply to all members of a sequence, which again can be determined in finitely many steps, then let the computer output the last formula in that sequence, for in that case the sequence is a derivation of the last formula of the sequence by the axioms and rules of the formal system; otherwise move on to the next sequence. In this way it is easy to see that for every formal system there exists a computer with a particular computer program, such that all and only the theorems of the formal system are enumerated by the computer with that program. It is also possible to reverse this procedure: for every computer with a computer program that enumerates only formulas in a particular formal language, one can show that a formal system exists with precisely that language whose theorems are exactly the formulas enumerated by that program. (This was proven by William Craig in the 1950s.) So there is in fact a tight correspondence between formal systems and computers with programs of a particular kind.

It was clear to Hilbert that in order for the formal systems that one intended to use in mathematical areas outside of arithmetic not to interfere with the "real mathematics" of natural numbers, it was necessary for them to be consistent. Indeed, it should not even be possible to derive a formula in any such system that would contradict a formula for which there was an elementary arithmetical proof. For otherwise the sum of that formal system with elementary arithmetic would be inconsistent again, which would mean that the formal system would be in conflict with the part of mathematics that for Hilbert was sacrosanct. For the same reason, the formal system could not be joined with arithmetic then for the purpose of deriving arithmetical statements in non-arithmetical ways, since one can show that in any inconsistent formal system that includes the standard axioms and rules of logic, the derivation of arithmetical formulas and indeed of all formulas whatsoever gets all *too* simple— for every formula whatsoever can be derived. Ideally, therefore, one would *prove* the consistency of the formal systems that one intended to employ in mathematics. And in order for these proofs to be beyond doubt again, it would be necessary to prove the consistency of these formal systems just by the means of elementary arithmetic. Thus, the consistency statements in question had to be translated into statements about natural numbers first, after which one could have hopes to prove them in the same ways as we prove theorems in arithmetic. This research project

became known as Hilbert's program (see Zach's (2003) entry in the *Stanford Encyclopedia of Philosophy*). It is this program in the foundations and the philosophy of mathematics that got seriously undermined by the two most famous theorems of mathematical logic: the Incompleteness Theorems proven by the Austrian mathematician Kurt Gödel (1906–1978) in 1931.[12]

In the next four sections it will be explained very briefly what these theorems say. This is not meant as a substitute for a thorough mathematical treatment of the Incompleteness Theorems. Ideally, having read the present article, you will become so interested in the topic that you will turn to a detailed formal exposition of the theorems.

4 The First Incompleteness Theorem

Theorem 1 (First Incompleteness Theorem) *Any consistent formal system $\mathcal{S}$ which includes a sufficient amount of arithmetic is incomplete with regard to statements of the language of elementary arithmetic, i.e., there are such statements that can neither be proven nor disproven in $\mathcal{S}$.*

(Strictly speaking this is not exactly Gödel's own version of the theorem but rather Barkley Rosser's later slight modification of it, but never mind.)

What do we mean by 'sufficient amount of arithmetic' and 'incomplete with regard to statements in the language of elementary arithmetic' here? Let us deal with them one after the other.

• *"Sufficient" amount of arithmetic*: A formal system $\mathcal{S}$ contains a "sufficient" amount of elementary arithmetic, in the sense of the First Incompleteness Theorem, if and only if:

1. The language $\mathcal{L}_\mathcal{S}$ of the formal system $\mathcal{S}$ either includes the language of elementary arithmetic itself or the function symbols and relation symbols of elementary arithmetic are definable in the language $\mathcal{L}_\mathcal{S}$. For example, while the standard language of set theory does not include the + sign for natural numbers, it is possible to define + and, for that matter, also the natural numbers themselves, just by means of

[12] While most philosophers of mathematics regard Hilbert's program as dead, in light of the Incompleteness Theorems, this does not mean that Formalism about mathematics is dead. Even more importantly, proof theory, that is, the part of mathematical logic that emerged from the failure of Hilbert's original program, is still an important area of logic.

set-theoretic vocabulary. Indeed, for every formula in the language of elementary arithmetic there is a "translation" of that formula into the standard language of set theory.

2. The axioms and rules of a formal system of arithmetic of a particular kind—which we will simply term 'elementary arithmetic' again—are either derivable in $\mathcal{S}$ themselves or they are derivable in $\mathcal{S}$ if the latter is extended by the definitions to which we referred in 1. We won't state the exact axioms and rules for elementary arithmetic here: The logical axioms and rules are just the ones of classical first-order logic. The arithmetical ones can be found in, for instance, appendix A.2 of Franzen (2005). Let me just stress that elementary arithmetic in this sense is *very* elementary: it contains much less arithmetic than what is presupposed by *any* mathematics course at a university.

• *Incomplete with regard to statements in the language of elementary arithmetic*: Let $\mathcal{S}$ be a formal system, such that $\mathcal{L}_{\mathcal{S}}$ includes the language of elementary arithmetic either directly or by the help of definitions (as explained before): $\mathcal{S}$ is then defined to be incomplete with regard to statements in the language of elementary arithmetic if and only if there is a formula A of the language of elementary arithmetic such that neither A nor $\neg A$ is a theorem of $\mathcal{S}$.

At this point, the amazing power of the First Incompleteness Theorem should become clear enough: Take any formal system $\mathcal{S}$ you like—any theory which one determines much in the way in which scientists determine their theories systematically by stating axioms. As long as $\mathcal{S}$ does not contain any formula A and its negation $\neg A$ simultaneously, and as long $\mathcal{S}$ includes a sufficient amount of elementary arithmetic, $\mathcal{S}$ is incomplete with regard to statements of elementary arithmetic, that is, there are such statements that can neither be proven nor disproven in $\mathcal{S}$. For example, take standard axiomatic set theory, which is an enormously strong formal system, in fact so strong that virtually all of the mathematical theorems that got proven by mathematicians so far can be derived in it: If it is consistent (which we believe to be the case), then since it contains a sufficient amount of arithmetic (and *much* more), it must be incomplete; there is a statement A in the language of elementary arithmetic, such that neither A nor $\neg A$ is derivable in it. Since neither A nor $\neg A$ is derivable in the theory, the result of adding, e.g., A to the theory is a strict consistent extension again. Still this extension will be

incomplete, by yet another application of the First Incompleteness Theorem: There is then a different sentence A' such that the extended theory proves neither A' nor $\neg A'$. And so forth.

Sometimes one can find the First Incompleteness Theorem stated in a version that employs a *truth predicate* ('is true'): For every consistent formal system $\mathcal{S}$ which includes a sufficient amount of arithmetic there is a *true* statement A in the language of elementary arithmetic, such that A is not derivable in $\mathcal{S}$. This is because among the statements A and $\neg A$ that any of the formal systems in question fails to derive according to the First Incompleteness Theorem, there must at least be one true sentence, since by the metalinguistic version of the law of the excluded middle, for every sentence A in the language of elementary arithmetic, either A is true or $\neg A$ is true.

There is nothing wrong whatsoever about this *semantic* reformulation of the purely *syntactic* First Incompleteness Theorem, as long as one has the right kind of understanding of 'true'. A precise theory of truth that supplies this understanding was introduced and developed by Alfred Tarski in the 1930s (see Tarski 1936). Indeed, even a weak, so-called "deflationary" theory of truth, together with classical logic, is sufficient to conclude the semantic variant of the First Incompleteness Theorem from Gödel's actual theorem. But note that the question of whether a formula A of the language of a formal system is true or false only makes sense if a certain *interpretation* of that language is presupposed—without assigning some sort of meaning to A it is simply not determined whether A is true or not. This interpretation or meaning-assignment is usually regarded as unproblematic at least in the case of the formulas of the language of elementary arithmetic.

How did Gödel manage to prove the First Incompleteness Theorem? Without going into any details, and without giving the proof itself, we would at least like to convey the two essential ideas that Gödel put together in order to give the proof: the *arithmetisation of syntax* on the one hand, and *self-referentiality* on the other.

5 The Arithmetisation of Syntax

If a formal system $\mathcal{S}$ contains the language of arithmetic and does not have any false arithmetical statements among its theorems, then obviously (parts of) the system can be interpreted as referring to natural numbers and as stating some of the mathematical properties of these numbers. However, if

natural numbers are at the same time regarded as *codes* for syntactic expressions such as terms, formulas, and derivations, then the arithmetic sentences of $\mathcal{S}$ can also be interpreted as referring indirectly to these syntactic expressions—by referring to the codes of these expressions—and stating indirectly some of the syntactic properties of these expressions—by stating them in terms of mathematical properties of their natural number codes. This is the ingenious thought behind Gödel's arithmetisation of syntax.

Here is one way in which this arithmetisation can be achieved[13]:

1. The coding mapping (Gödel coding or Gödelisation):

   Let us exemplify the idea in terms of a simple informal example. Consider the English alphabet:

   $$a, b, c, \ldots$$

   (We will suppress quotation marks where we can, but strictly speaking the English alphabet includes 'a', 'b', 'c', ..., rather than a, b, c, ...).

   Now we can construct a coding function that maps finite sequences of English letters to natural numbers. For example:

   - Assign 1 to a, 2 to b, 3 to c, 4 to d, and so forth.
   - Let p_1 be the first prime number (2), p_2 the second prime number (3), p_3 the third prime number (5), and so on.
   - Now assume that we are given a sequence of English letters; as one says in this context, one is given a "word" or string, where it does not matter whether such a "word" or string is meaningful or not. But just for the fun of it, let us choose a sequence of letters that does have a meaning: *bad*.

     We can then encode this word by means of a number as follows:

     (a) Take p_1 to the power of the number that was assigned to the first letter in the word. In our case: $p_1^2 = 2^2$.

     (b) Take p_2 to the power of the number that was assigned to the second letter in the word. In our case: $p_2^1 = 3^1$.

[13]There are more efficient coding procedures than the one described here, but again never mind.

(c) Take p_3 to the power of the number that was assigned to the third letter in the word. In our case: $p_3^4 = 5^4$.
(Since our word consists of three letters, we are done.)

- Finally, multiply the results of these calculations. In our case: $2^2 \cdot 3^1 \cdot 5^4 = 4 \cdot 3 \cdot 625 = 7500$.
- So 7500 is the Gödel code of our given word *bad*.

This encoding has two remarkable properties: First of all, the coding function is *computable*, that is, a computer could be programmed to determine the code of any given input string in finitely many computation steps. Secondly, it's a *proper* encoding, in the sense that it is possible to reconstruct from every code number the unique original word that it encodes—in other words, *decoding* works properly:

- Let a Gödel code n be given. In the example: 7500.
  We want to reconstruct which word is encoded by it.
- Determine the prime factorization of n. In our case: $2^2 \cdot 3^1 \cdot 5^4$.
  (According to the Fundamental Theorem of Arithmetic, this factorisation is determined uniquely, up to the ordering of factors.)
- Now we proceed as follows:
  (a) Consider the power of the first prime factor and take its corresponding English letter. In our case: *b*.
  (b) Consider the power of the second prime factor and take its corresponding English letter. In our case: *a*.
  (c) Consider the power of the third prime factor and take its corresponding English letter. In our case: *d*.
  (Since here our n has only three prime factors, we are done.)
- Finally, write down the resulting letters consecutively.
  In our case: *bad*
- So *bad* is the word that is encoded by 7500.

In the case of the Incompleteness Theorems, Gödel introduced such a coding function for the expressions of the language $\mathcal{L}_\mathcal{S}$ of any system $\mathcal{S}$ that is referred to by the theorems. Since derivations from the axioms of $\mathcal{S}$ by means of the rules of $\mathcal{S}$ are also nothing else but (perhaps long) strings of symbols, they can be encoded by numbers in this way,

too. The existence of a "mechanical" coding function such as the one sketched above is itself a mathematical fact. So there is nothing "non-mathematical" about Gödel's theorems or proofs. Gödel simply uses the mathematical fact that such a coding exists in his mathematical proof of the Incompleteness Theorems.

2. Gödel was able to prove that (i) all the standard syntactic properties, relations, and operations that are to do with terms, formulas, and derivations, correspond to more or less "nice" arithmetical properties, relations, and operations of their corresponding Gödel codes, and that (ii) all of the latter "nice" arithmetical properties, relations, and operations can be represented in elementary arithmetic (the formal system that we dealt with in the last section). But what does it mean to say that an arithmetical property, relation or operations is *representable* in a formal system? This can be made more precise as follows (we will only do this for properties and relations): a property or relation R of natural numbers, where R has k arguments, is representable in a formal system $\mathcal{S}$ the language of which includes the language of elementary arithmetic if and only if there is a formula $A[x_1, \ldots, x_k]$ of $\mathcal{L}_\mathcal{S}$ with k variables, such that for all natural numbers $n_1, \ldots, n_k$:

   (a) if $R(n_1, \ldots, n_k)$, then $\mathcal{S}$ proves $A(n_1, \ldots, n_k)$,
   (b) if not $R(n_1, \ldots, n_r)$, then $\mathcal{S}$ proves $\neg A(n_1, \ldots, n_k)$.

   We also say that such a case that the formula A represents R in $\mathcal{S}$.

   This is still not perfectly precise. Within $R(n_1, \ldots, n_k)$ on the left-hand side, for instance, n_1 is meant to stand for a natural number, while in $A(n_1, \ldots, n_k)$ on the right-hand side n_1 is meant to stand for a *symbol* for a natural number. This could be made explicit by replacing n_1 on the right-hand side by, say, $\overline{n_1}$, but as long as one is clear about what is meant here, let's not be too pedantic.

   What Gödel proved was that by means of coding, all the usual syntactic properties, relations, and operations for formal expressions are representable in elementary arithmetic and therefore also in every formal system that includes elementary arithmetic as a subsystem. In particular, the syntactic relation

   - $Pr_\mathcal{S}(x, y)$: the sequence with Gödel code x is a derivation in the formal system $\mathcal{S}$ of the sentence with Gödel code y

is representable in all systems that extend elementary arithmetic. Once the relation $Pr_{\mathcal{S}}$ is represented in this way in terms of an arithmetical formula, it is easy to define *provability in* $\mathcal{S}$, that is,

- $Prov_{\mathcal{S}}(y)$: the sentence with Gödel code y is provable in $\mathcal{S}$

on the basis of it. Simply take $Prov_{\mathcal{S}}(y)$ to be: $\exists x Pr_{\mathcal{S}}(x, y)$.

Given this representation of provability in $\mathcal{S}$ in terms of an arithmetical formula, one can ask which properties of the concept *provability in* $\mathcal{S}$ can be proven in $\mathcal{S}$. It turned out that in standard formal systems of arithmetic, such as the well-known system of so-called Peano arithmetic which is a proper extension of the system of elementary arithmetic in the last section, various important such properties can be proven to hold of *provability in* $\mathcal{S}$. In fact, this will be exactly what the phrase 'certain amount of arithmetic' in our formulation of the Second Incompleteness Theorem is meant to say, namely that a few essential properties of formal provability can be proven to hold (see below). But before we turn to the Second Incompleteness Theorem, we still need to supply the second of Gödel's ideas that were crucial to his proof of the First Incompleteness Theorem.

6 Self-Referentiality

In Gödel's original proof of the First Incompleteness Theorem, the existence of a "self-referential" sentence plays an important role, where a self-referential sentence is simply one that speaks about itself (via coding) or about a sentence that is provably equivalent to itself.

As Gödel showed, there is a sentence G of the language of elementary arithmetic, such that if $\mathcal{S}$ satisfies the assumptions mentioned in the First Incompleteness Theorem, then $\mathcal{S}$ proves the equivalence formula

$$G \leftrightarrow \neg Prov_{\mathcal{S}}(\ulcorner G \urcorner)$$

where, generally, if B is a formula in $\mathcal{L}_{\mathcal{S}}$, $\ulcorner B \urcorner$ is an arithmetical term that denotes the Gödel code of B (and where a computer could be programmed to determine that term given B as input).

In this sense, up to provable equivalence, G says about itself via reference to its own code:

I am not provable in $\mathcal{S}$.

Gödel derives the existence of G from a general fixed-point lemma (also called diagonalization lemma), which says that for every formula $A[x]$ in $\mathcal{L}_\mathcal{S}$ with one free variable x[14], there exists a sentence in $\mathcal{L}_\mathcal{S}$ again that says about itself (or about a sentence that is provably equivalent to itself) that it has the property that is expressed by $A[x]$. The proof of this lemma is actually constructive, that is, Gödel constructs the fixed-point formula in question by a concrete procedure that could be run on a computer given the input $A[x]$. If $A[x]$ is chosen to be $\neg Prov_\mathcal{S}(x)$, then this yields the existence of G above.

The sentence G is reminiscent of the famous "Liar" sentence which says

I am not true.

and which seems to lead to a contradiction by means of the following bit of informal reasoning: Assume the Liar sentence is true; then what it says must be the case; but what it says is that it is not true. So we have: If the Liar sentence is true, then it is not true. Now assume the Liar sentence is not true: but that is exactly what it says; so it must be true after all. So we also have: If the Liar sentence is not true, then it is true. Summing up, this yields: the Liar sentence is true if and only if it is not true. But that is a logical contradiction in classical logic. There is now a whole area in philosophical logic called 'formal theories of truth', which deals with the Liar paradox and with ways of avoiding its contradictory consequences. In spite of the similarity between the Liar sentence and Gödel's sentence G above, it is important to realize that the two are very different still: First of all, while the Liar sentence includes a truth predicate, which is a *semantic* predicate, Gödel's sentence G includes a provability predicate for some formal system, where provability in that system is not a semantic but a *syntactic* notion, or, via coding: an *arithmetical* notion. Indeed, G is but a sentence in the language $\mathcal{L}_\mathcal{S}$ that says something about natural numbers and their arithmetical properties or relationships or operations, even when we know that *qua* coding G says something about its own unprovability. Secondly, while the Liar seems to commit us to the proof of a contradiction, Gödel's sentence G does not: for instance, take $\mathcal{S}$ to be elementary arithmetic itself; then if $\mathcal{S}$ is consistent, which no one doubts, $\mathcal{S}$ neither derives G nor $\neg G$. Since, in particular, G is not derivable in $\mathcal{S}$, G is actually a true statement.

[14]Free means: not in the range of a quantifier expression $\exists x$ or $\forall x$.

But no contradiction follows from that: it just means that $\mathcal{S}$ misses out on one particular truth, that is, G.

We turn now to the second of Gödel's Incompleteness Theorems.

7 The Second Incompleteness Theorem

Theorem 2 (Second Incompleteness Theorem) *For any consistent formal system $\mathcal{S}$ that includes a sufficient amount of arithmetic, the consistency of $\mathcal{S}$ is not provable in $\mathcal{S}$ itself.*

We are already familiar with some of the "ingredients" of this theorem. Mainly we will have to deal with the notion *certain sufficient amount of arithmetic* again—which amounts to something *different* here from what it meant in the case of the First Incompleteness Theorem—and the notion of *unprovability of the consistency of $\mathcal{S}$ in $\mathcal{S}$ itself*:

• *Sufficient amount of arithmetic*: $\mathcal{S}$ includes a sufficient amount of arithmetic in the sense of the Second Incompleteness Theorem if (i) it includes a sufficient amount of elementary arithmetic in the sense of the First Incompleteness Theorem, that is, it includes a formal system of arithmetic of a particular kind as a subsystem, and additionally (ii) $\mathcal{S}$ and $Prov_{\mathcal{S}}$ taken together satisfy the following conditions which are called the *provability conditions* (or Hilbert-Bernays-Löb conditions)[15]:

1. $\mathcal{S}$ proves: $Prov_{\mathcal{S}}(\ulcorner A \rightarrow B \urcorner) \rightarrow (Prov_{\mathcal{S}}(\ulcorner A \urcorner) \rightarrow Prov_{\mathcal{S}}(\ulcorner B \urcorner))$.

2. $\mathcal{S}$ proves: $Prov_{\mathcal{S}}(\ulcorner A \urcorner) \rightarrow Prov_{\mathcal{S}}(\ulcorner Prov_{\mathcal{S}}(\ulcorner A \urcorner) \urcorner)$.

3. If $\mathcal{S}$ proves A, then $\mathcal{S}$ proves $Prov_{\mathcal{S}}(\ulcorner A \urcorner)$.

where $\ulcorner A \urcorner$ and $\ulcorner B \urcorner$ are symbols again for the Gödel codes of whatever sentences of $\mathcal{L}_{\mathcal{S}}$ that replace the letters A and B, respectively.

[15] This is a non-trivial requirement: one can actually define $Prov_{\mathcal{S}}(y)$ in a way such that according to the standard interpretation of the arithmetical signs the set of formulas whose codes have the very property that is expressed by $Prov_{\mathcal{S}}(y)$ is precisely the set of formulas derivable in $\mathcal{S}$, but still elementary arithmetic is not able to derive the provability conditions for $Prov_{\mathcal{S}}(y)$ defined in such way. However, given Gödel's own definition of $Prov_{\mathcal{S}}(y)$, the provability conditions are indeed satisfied for $Prov_{\mathcal{S}}(y)$ and all suffiently strong formal systems of arithmetic (including the system of first-order Peano arithmetic).

In modern philosophical logic, statements of a similar form as these figure prominently in *modal logic*, however there '$Prov_{\mathcal{S}}$' usually gets replaced by some sentential operator $\Box$—which might stand for necessity or knowability or for something else—and instead of the symbols for the codes of formulas it is the formulas themselves that get stated (see Yde Venema's article on "Modal Logic and (Co-)Algebra" in this volume). For instance, $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$, $\Box A \rightarrow \Box\Box A$, and if $\vdash A$ then $\vdash \Box A$, are normally all counted as valid principles of the logic of metaphysical necessity. This means: the set of necessary propositions is closed under Modus Ponens, if a proposition is necessary then it is necessary that this is so, and everything that is provable from the logical principles for necessity is itself necessary. There is a whole subarea of modal logic that deals with the study of provability in formal systems such as first-order Peano arithmetic just by means of the syntactical and logical resources of modal logic and of its famous possible-worlds semantics (which was developed by Saul Kripke in the 1950s and 1960s). In this way, Hilbert's metamathematics—mathematics about mathematics—can be given a modal formalisation.[16]

Finally, we can now also say more exactly what the 'unprovability of the consistency of $\mathcal{S}$ in $\mathcal{S}$ itself' means:

• *The consistency of $\mathcal{S}$ is not provable in $\mathcal{S}$ itself*: Given our previous definition of $Prov_{\mathcal{S}}$, and given that $\mathcal{S}$ contains elementary arithmetic as a subsystem, it is not hard to see that the consistency of $\mathcal{S}$ can be expressed in terms of an arithmetical formula, for example, by means of:

$$\neg Prov_{\mathcal{S}}(\ulcorner 0 = 1 \urcorner).$$

Let us abbreviate this last formula by $Con_{\mathcal{S}}$. So 'the consistency of $\mathcal{S}$ is not provable in $\mathcal{S}$ itself' simply means: The formula $Con_{\mathcal{S}}$ is not provable in $\mathcal{S}$.

The proof of the Second Incompleteness Theorem consists mainly of a formalisation of the proof of Gödel's original version of the First Incompleteness Theorem within the formal language of $\mathcal{S}$ itself, which becomes possible on the basis of the provability conditions.

Just as in the case of the First Incompleteness Theorem, containing a sufficient amount of arithmetic *up to translation by means of definitions* is

[16]For a standard reference on modal provability logic, see Boolos (1993).

actually enough to get the Second Incompleteness Theorem going. For instance, the Second Incompleteness Theorem implies that if standard set theory is consistent, then since it contains elementary arithmetic, and since it satisfies the provability conditions given Gödel's own definition of $Prov_{\mathcal{S}}(y)$, it is not able to derive its own consistency *even though we all believe it is true that standard set theory is consistent* and *even though the consistency of standard set theory can be expressed in purely mathematical terms by Gödel's coding methods*. If not even set theory is able to derive its own consistency (assuming it is consistent), then of course no weak system of arithmetic will be able to derive the consistency of set theory either. Hence, the formal system of standard set theory is an important part of (formalised) mathematics that is beyond the reach of Hilbert's program which aimed to prove the consistency of such formal systems by elementary arithmetical means. In this sense, Hilbert's program has failed, and it was proven to be a failure by a logical theorem.

8 What Does This Show About Provability?

One of the very recent trends in the philosophy of mathematics is an emphasis on *mathematical practice*, that is, on what single mathematicians—and the mathematical community as a whole—actually produce and do, and by what methods and principles they operate. This trend is usually regarded to be opposed to, or at least critical of, the application of logical methods in the philosophy of mathematics, since these are regarded as too abstract and too far removed from the concepts that one actually needs in order to study the activities of "real" mathematicians. As we have found, the Incompleteness Theorems deal with derivability in particular formal systems. But, at least at first glance, when "real" mathematicians prove theorems, they do not seem to put forward any formal systems in which they carry out these proofs: They do not specify any formal language by syntactic rules, they do not fix any particular set of statements as "the" set of mathematical axioms, they usually do not refer to derivation rules at all, and when they check whether a sequence of statements is a correct proof, the core of what they are doing does not seem to involve syntactic procedures. In a nutshell: what mathematicians mean when they speak of proof and provability, and what they do when they actually decide whether something is a proof, seems to differ substantially from what we called derivation and derivability in a formal system. "Real"

provability does not seem to be relativized to any formal system but is rather *absolute* and *informal* (as was emphasized by Gödel himself, and later again by John Myhill in his "Some Remarks on the Notion of Proof"—see Myhill 1960).[17] So how exactly does absolute/informal proof and provability relate to proof and provability in formal systems?

Surprisingly, *some* information about this can be derived from the Incompleteness Theorems again. On their basis, it turns out to be possible to prove significant statements even about computability, the human mind, and the in-principle proving capabilities of human mathematicians. Indeed, this is a point at which artificial intelligence and cognitive science meet logic and the philosophy of mathematics. And logical methods are found to throw some light even on provability in the sense of mathematical practice (or at least on something close to that).[18]

Let

- T be the set of *true arithmetical* statements,
- K be the set of *humanly knowable arithmetical* statements,
- S_e be the set of all arithmetical statements *enumerated by the computer (Turing machine)* e according to the program of that computer,
- K' be the set of *humanly knowable* statements.

The exact definition of 'arithmetical statement' is not so important, but 'sentence in the language of elementary arithmetic' in the sense of the first section is a possible option. With respect to the enumeration of arithmetical statements by a computer, think of computer programs which in discrete steps determine an arithmetical statement and which then print it on a screen. It is not presupposed that any such program terminates after finitely many steps—e might well be a computer and a program which print arithmetical statements indefinitely.

Here are two arguments for theses that are philosophically important, where each of the arguments relies on one of the Incompleteness Theorems:

[17] Leitgeb (2009) discusses the differences between formal provability on the one hand, and informal or absolute provability by mathematicians on the other, in logical and philosophical detail.

[18] In the following, we follow Stewart Shapiro's treatment of the topic in his "Incompleteness, Mechanism, and Optimism" (Shapiro 1998) to a large extent.

1. Argument:

   - Assume the so-called *Mechanistic Thesis*, that is: There is a Turing machine e, such that $K = S_e$.
   - But that means K is enumerable by a Turing machine.
   - However, by the First Incompleteness Theorem, the set T is not enumerable by a Turing machine, since if it were so then T would coincide with the set of theorems of the formal system that corresponded to that Turing machine, which is excluded by the semantic (truth-theoretic) variant of the First Incompleteness Theorem.
   - So it follows that $K \neq T$.
   - Summing up, this argument shows that the following thesis is entailed by the First Incompleteness Theorem:

     $$\text{Thesis 1: Mechanistic Thesis} \rightarrow K \neq T$$

     In words: If the set of humanly knowable arithmetical statements can be enumerated by a Turing machine, then there are true arithmetical statements which are not humanly knowable.

2. Argument:

   - Assume the *Mechanistic Thesis* again, that is: There is a Turing machine e, such that $K = S_e$.
   - Assume that the Hilbert-Bernays-Löb provability conditions hold for S_e and a particular provability predicate.
   - Assume for contradiction that the statement '$K = S_e$' is a member of K':
   - Since, trivially, '$K \subseteq T$' is a member of K' (because we know from epistemology that knowledge implies truth), it follows that '$S_e \subseteq T$' is a member of K', too (by our equality assumption from above together with standard assumptions on the closure of K' under logical derivations).
   - Hence, it also follows that 'S_e is consistent' is a member of K' (because we know that every set of true statements must be consistent, where the consistency statement is formulated by means of the provability predicate referred to before).

- But 'S_e is consistent' is an arithmetical statement (as being given by the arithmetization of syntax). Thus, 'S_e is consistent' is even a member of K.
- But this contradicts the Second Incompleteness Theorem: since K certainly contains "sufficient" arithmetic, S_e does so, too; S_e would furthermore be enumerable by a Turing machine and would therefore coincide with the set of theorems of a formal system; by assumption, the provability conditions are satisfied; finally, by what we have just seen, S_e would include a statement that in a straightforward way expresses that S_e is consistent; but that is exactly the type of situation that is excluded by the Second Incompleteness Theorem. Contradiction.
- So by reductio we have shown: '$K = S_e$' is not a member of K'.
- Summing up, this argument proves that the following thesis is entailed by the Second Incompleteness Theorem:

  Thesis 2: Mechanistic Thesis (and provability conditions) $\rightarrow$ '$K = S_e$' is not a member of K'.

  In words: If the set of humanly knowable arithmetical statements can be enumerated by a Turing machine (and the provability conditions hold for that set and for an arithmetical formula that determines that set), then it is not humanly knowable by which Turing machine the set of humanly knowable arithmetical statements can be enumerated.

It should be clear that the two arguments from above are not fully formalised themselves. In particular, all informally stated claims that involve quotation marks and/or the membership sign should be made precise. But this can be done: see Carlson (2000) for a much more precise and sophisticated treatment.

Both thesis 1 and thesis 2 are material implications. By classical propositional logic, they can be reformulated in terms of the following disjunctions:

The Mechanistic Thesis is false or $K \neq T$.

and

The Mechanistic Thesis is false (or the provability conditions are false) or '$K = S_e$' is not a member of K'.

The former thesis says: Either what we can know in principle about arithmetic surpasses the powers of any Turning machine, or there are arithmetical statements A and $\neg A$ for which we are for principled reasons unable to decide whether A is true or $\neg A$ is true. This is Kurt Gödel's famous dichotomy which he himself derived from his Incompleteness Theorems in his Gödel (1995).[19] The other thesis amounts to, if we ignore the part on the provability conditions (which one would need to make much more precise anyway): Either what we can know in principle about arithmetic surpasses the powers of any Turning machine, or for principled reasons we cannot know which Turing machine enumerates all and only those arithmetical truths that we can know to be true.

Is it perhaps possible to do better than these theses? That is: Is it possible to argue on the basis of the Incompleteness Theorems in favour of one of the disjuncts rather than "merely" in favour of the disjunctions from above? John Lucas (1961) and Roger Penrose (1989) thought so, when they tried to argue in such a manner just for the falsity of the Mechanistic Thesis, but careful philosophical and logical analysis of their arguments (which is still ongoing) indicates that none of their arguments is sound.

Lots of questions remain open. For instance: Does absolute/informal provability satisfy the same provability conditions as formal provability did in section 2, or are the "modal laws" of absolute/informal provability fundamentally different?[20] Is it possible to give a logical-philosophical explication of 'proof' and 'provability' in the absolute/informal sense rather than the formal sense? Is it at least possible to state a true and informative axiomatic system in which 'x is a proof' figures as a primitive term, much as 'x is known' is taken as a primitive expression in recent epistemological theories of knowledge (see Williamson 2000 and Vincent Hendricks' article on "Logic and Epistemology" in this volume)?[21] Finally: Are there true mathematical statements which are even *absolutely/informally* unprovable? While the

[19]Solomon Feferman gives an excellent presentation of this dichotomy in Feferman (2006).

[20]For instance, $Prov(\ulcorner A \urcorner) \rightarrow A$ or $\Box A \rightarrow A$ seems to be obviously logically valid for informal or absolute provability, whereas by a theorem by Löb only trivial instances of this scheme can be derived for formal provability in formal systems such as the ones that the Second Incompleteness Theorem deals with.

[21]By the way: if Williamson (2000) is right, then *knowability* does not satisfy all of the Hilbert-Bernays-Löb provability conditions, since the modal 4 or transitivity principle $\Box A \rightarrow \Box\Box A$, which is known to epistemologists as the KK principle, fails for knowability.

Incompleteness Theorems do not answer these questions just by themselves, even attempting to answer them without taking into account the Incompleteness Theorems will remain to be a futile endeavour.[22]

References

[1] Avigad, J., 2011: "Proof Theory", this volume.

[2] Benacerraf, P. and H. Putnam (eds.), 1983: *Philosophy of Mathematics: Selected Readings*, Cambridge: Cambridge University Press, 2nd edition.

[3] Boolos, G., 1993: *The Logic of Provability.* Cambridge University Press, Cambridge.

[4] Boolos, G.S., J.P. Burgess, and R.C. Jeffrey, 2002: *Computability and Logic*, Cambridge: Cambridge University Press, fourth edition.

[5] Carlson, T.J., 2000: "Knowledge, Machines, and the Consistency of Reinhardt's Strong Mechanistic Thesis", *Annals of Pure and Applied Logic* 105, 51–82.

[6] Feferman, S., 2006: "Are There Absolutely Unsolvable Problems? Gödel's Dichotomy", *Philosophia Mathematica* 14, 134–152.

[7] Field, H., 1980: *Science Without Numbers: A Defense of Nominalism*, Oxford: Blackwell.

[8] Franzén, T., 2005: *Gödel's Theorem. An Incomplete Guide to Its Use and Abuse*, Wellesley, Mass.: A K Peters.

[9] Gödel, K., 1931: "Über formal unentscheidbare Stze der Principia Mathematica und verwandter Systeme I, *Monatshefte für Mathematik und Physik* 38: 173–198.

[10] Gödel, K., 1995: "Some Basic Theorems on the Foundations of Mathematics and Their Implications", in: S. Feferman et al. (eds.), *Kurt Gödel: Collected Works III. Unpublished Essays and Lectures*, New York: Oxford University Press, 304–323.

[22] See Leitgeb (2010) for more on these questions; also google the recent conference titled 'Two Streams in the Philosophy of Mathematics: Rival Conceptions of Mathematical Proof' which dealt with many of the questions that got raised in this final section.

[11] Hájek, P. and P. Pudlák, 1991: *Metamathematics of First-Order Arithmetic. Perspectives in Mathematical Logic.* Berlin: Springer.

[12] Hale, B. and C. Wright, 2001: *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*, Oxford: Oxford University Press.

[13] Hendricks, V., 2011: "Logic and Epistemology", this volume.

[14] Horsten, L., 2007: Entry on "Philosophy of Mathematics" in the *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/philosophy-mathematics/.

[15] Leitgeb, H., 2009: "On Formal and Informal Provability", in: Ø. Linnebo and O. Bueno (eds.), *New Waves in Philosophy of Mathematics*, New York: Palgrave Macmillan, 263–299.

[16] Lucas, J.R., 1961: "Mind, Machines and Gödel", *Philosophy* 36, 112–127.

[17] Myhill, J., 1960: "Some Remarks on the Notion of Proof", *The Journal of Philosophy* 57/14, 461–471.

[18] Penrose, R., 1989: *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*, Oxford: Oxford University Press.

[19] Raatikainen, P., 2005: "On the Philosophical Relevance of Gödel's Incompleteness Theorems." *Revue Internationale de Philosophie* 59/4, 513–534.

[20] Shapiro, S., 1997: *Philosophy of Mathematics: Structure and Ontology*, Oxford: Oxford University Press.

[21] Shapiro, S., 1998: "Incompleteness, Mechanism, and Optimism", *The Bulletin of Symbolic Logic* 4/3, 273–302.

[22] Smith, P., 2009: *An Introduction to Gödels Theorems*, Cambridge: Cambridge University Press.

[23] Tarski, A., 1936: "Der Wahrheitsbegriff in den formalisierten Sprachen", *Studia Philosophica* 1, 261–404.

[24] Williamson, T., 2000: *Knowledge and Its Limits*, Oxford: Oxford University Press.

[25] Venema, Y., 2011: "Modal Logic and (Co-)Algebra", this volume.

[26] Zach, R., 2003: Entry on "Hilbert's Program" in the *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/hilbert-program/.

[27] Zalta, E., 2008: Entry on "Frege" in the *Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu/entries/frege/.