Gödel's Incompleteness Theorems

Guram Bezhanishvili*

1 Introduction

In 1931, when he was only 25 years of age, the great Austrian logician Kurt Gödel (1906–1978) published an epoch-making paper [15] (for an English translation see [8, pp. 5–38]), in which he proved that an effectively definable consistent mathematical theory which is strong enough to prove Peano's postulates of elementary arithmetic **cannot** prove its own consistency.[1] In fact, Gödel first established that there always exist sentences φ in the language of Peano Arithmetic which are true, but are *undecidable*; that is, neither φ nor $\neg\varphi$ is provable from Peano's postulates. This is known as *Gödel's First Incompleteness Theorem*. This theorem is quite remarkable in its own right because it shows that Peano's well-known postulates, which by and large are considered as an axiomatic basis for elementary arithmetic, cannot prove all true statements about natural numbers. But Gödel went even further. He showed that his first incompleteness theorem implies that an effectively definable sufficiently strong consistent mathematical theory cannot prove its own consistency. This theorem became known as *Gödel's Second Incompleteness Theorem*. Since then the two theorems are referred to as *Gödel's Incompleteness Theorems*. They became landmark theorems and had a huge impact on the subsequent development of logic.

In order to give more context, we step further back in time. The idea of formalizing logic goes back to the ancient Greek philosophers. One of the first to pursue it was the great German philosopher and mathematician Gottfried Wilhelm Leibniz (1646–1716). His dream was to develop a universal symbolic language, which would reduce all debate to simple calculation. The next major figure in this pursuit was the English mathematician George Boole (1815–1864), who has provided the first successful steps in this direction. This line of research was developed to a great extent by the famous German mathematician and philosopher Gottlob Frege (1848–1925), and reached its peak in the works of Bertrand Russell (1872–1970) and Alfred North Whitehead (1861–1947). Their magnum opus *Principia Mathematica* [26] has provided relatively simple, yet rigorous formal basis for logic, and became very influential in the development of the twentieth century logic.[2]

*Mathematical Sciences; Dept. 3MB, Box 30001; New Mexico State University; Las Cruces, NM 88003; gbezhani@nmsu.edu.

[1] We recall that a theory is *consistent* if it does not prove contradiction.

[2] More details on the work of Boole, Frege, and Russell and Whitehead can be found on our webpage http://www.cs.nmsu.edu/historical-projects/; see the historical projects [23, 7]. The work of Boole has resulted in an important concept of *Boolean algebra*, which is discussed in great length in a series of historical projects [3, 2, 1], also available on our webpage.

The next major figure in the development of logical formalism was the great German mathematician David Hilbert (1862–1943), who had considerably simplified the Russell-Whitehead formalism. First-order logic as we know it today emerged from the work of Hilbert and his school. Hilbert believed that all of mathematics can be developed based on a carefully chosen finite set of axioms, which can be proven to be consistent by *finitistic methods*, and that it is decidable whether a given mathematical statement is a theorem of the system. In the 1920s Hilbert formulated this as a research program, which became known as the *Hilbert program*. The concept of "finitistic method" is not clearly defined, but one possible reading of it is that we should be able to formalize the method that provides such a consistency proof within the system itself. This is exactly where Gödel's second incompleteness theorem starts to play a crucial role. Indeed, if we have a theory T capable of axiomatizing all of mathematics, then certainly T should be able to deduce all of elementary arithmetic. But then Gödel's theorem states that if T is consistent, then it cannot prove its own consistency, thus shattering Hilbert's belief that such a theory T could exist.

2 Gödel's life

Kurt Friedrich Gödel was born on April 28, 1906, in Brno, which currently is part of Czech Republic, but back then was part of Austro-Hungarian Empire. He was born into an ethnic German family. Kurt was an extremely curious child, to the extent that he became known as "Herr Warum" (Mr. Why). From 1912 to 1916 Gödel attended a Lutheran school, and from 1916 to 1924 he was enrolled in a German gymnasium, excelling with honors in all his subjects, particularly in mathematics, languages, and religion. In 1924 Kurt moved to Vienna, where he entered the University of Vienna. At the university he became a participant of the famous *Vienna Circle*, led by Moritz Schlick (1882–1936), and including Hans Hahn (1879–1934) (of the Hahn-Banach theorem), and Rudolf Carnap (1891–1970). An important event of those years of Gödel's life was attending Hilbert's lecture in Bologna on completeness and consistency of mathematical systems.

In 1928 Hilbert and his student Wilhelm Ackermann (1896–1962) published an influential book [21]. One of the open problems posed in the book was whether a formula is provable in a first-order theory T iff it is true in all models of T. This problem became known as the *completeness problem*. Gödel chose this as a topic of his dissertation, which he completed in 1929 under the supervision of Hahn. In the dissertation Gödel gave an affirmative solution of the problem. The obtained theorem became known as *Gödel's Completeness Theorem*.[3] He was awarded the doctorate in 1930. The same year Gödel's paper appeared in press [14], which was based on his dissertation.

In 1931 Gödel published his epoch-making paper [15]. It contained his two incompleteness theorems, which became the most celebrated theorems in logic. The incompleteness theorems have dramatically changed our perception of logic, and made the author one of the greatest logicians of all time. They not only have several fundamental consequences, but also gave birth to several new branches of logic. The incompleteness theorems remain a constant source of inspiration for new generations of logicians, and after almost 80 years since their

[3]For more information on the completeness theorem we refer to the historical project [5], which is available on our webpage `http://www.cs.nmsu.edu/historical-projects/`.

publication, still remain a topic of discussion and debate.

In 1932 Gödel earned his *Habilitation* at the University of Vienna, and in 1933 he became a *Privatdocent* there. The same year the Nazis came to power in Germany and one of the most turbulent eras for Europe has started.

It is in 1933 that Gödel started to visit the United States on a consistent basis. In 1933 he delivered an address to the annual meeting of the American Mathematical Society. The same year he visited Alonzo Church (1903–1995) in Princeton. In 1934 he gave a series of lectures at Princeton. Church's two brilliant students Stephen Kleene (1909–1994) and Barkley Rosser (1907–1989) took notes, which were subsequently published in [16]. This is when Church was contemplating how to formalize the concept of an *effectively calculable* function, which has resulted in the famous *Church's thesis*. More on this, as well as on Gödel's initial rejection of Church's proposal, and his later acceptance of it thanks to the impressive work of Alan Turing (1912–1954) can be found in the historical project "Church's Thesis"; see [4, pp. 253–265]. Gödel would visit Princeton again in 1935 and 1938, and would eventually settle there in 1940.

In 1936 Schlick was assassinated by a pro-Nazi student. This triggered a severe nervous breakdown in Gödel, who developed paranoid symptoms, including a fear of being poisoned. This would hunt him throughout his life, and would eventually become the reason of his death.

In 1938 the Nazi Germany annexed Austria de facto into Greater Germany (the *Anschluss*). The same year Gödel married his partner of 10 years Adele Nimbursky,[4] and the couple decided to escape the Nazi Germany and settle in the United States. In 1939 World War II started. The same year Kurt and Adele left Vienna for Princeton. Instead of crossing the Atlantic, which was rather dangerous at the time, the Gödels decided to take the trans-Siberian railway to the Pacific, sailed from Japan to San Francisco, and crossed the United States by train to Princeton, where Gödel accepted a position at the Institute for Advanced Study. An interesting story of their travels is told in [10].

1940 marks another famous contribution of Gödel to mathematics, namely to set theory, when he showed that the Axiom of Choice (AC) and the Generalized Continuum Hypothesis (GCH) are consistent with ZF (the Zermelo-Fraenkel set theory) [17]. This he did by constructing a model of ZF in which both (AC) and (GCH) hold. The model became known as Gödel's *constructible universe*. This provided a partial solution of the first problem of the set of 23 problems that Hilbert presented to the International Congress of Mathematicians in Paris in 1900, which have shaped the twentieth century mathematics. The second half of the solution was given by Paul Cohen (1934–2007) in 1964, who showed that there exist models of ZF in which the negation of both (AC) and (GCH) hold. In doing so, Cohen introduced a technique of *forcing*, which became instrumental in proving independence results. For his work Cohen was awarded the Fields Medal in 1966, which is the most prestigious award in mathematics. It is awarded every four years to a mathematician under 40 years of age who made a significant contribution to mathematics. It is only surprising that Gödel was never awarded the Fields Medal. In fact, Cohen's 1966 Fields Medal continues to be the only Fields Medal to have been awarded for a work in mathematical logic.

Gödel became a permanent member of the Institute for Advanced Study in 1946. Around

[4]Gödel's parents had opposed their relationship because Adele was a divorced dancer, six years older than Gödel.

this time his interests turned to philosophy and physics. In 1951 Gödel demonstrated the existence of paradoxical solutions to Albert Einstein's (1879–1955) field equations in general relativity. His solutions became known as the *Gödel metric*; these "rotating universes" would allow time travel and caused Einstein to have doubts about his own theory. The same year he was awarded the first Albert Einstein Award.

Gödel became a full professor at the Institute for Advanced Study in 1953. He was a big admirer of Leibniz. In the early 1970s, Gödel circulated an elaboration of Leibniz's version of Anselm of Canterbury's (c. 1033–1109) ontological proof of God's existence. This became known as *Gödel's ontological proof*. In 1974 Gödel was awarded the National Medal of Science. He became an emeritus professor at the Institute in 1976.

Throughout his life Gödel suffered periods of mental instability, which have gotten worse later in his life. He had an obsessive fear of being poisoned. In order for him to eat, his wife Adele had to taste the food for him. Late in 1977 Adele was hospitalized for six months and could not taste Gödel's food anymore. In her absence he starved himself to death. When he died he weighed only 65 pounds (approximately 30 kg).

There are many books written and many stories told about Gödel. His friendship with Einstein was legendary. The two would walk together to and from the Institute for Advanced Study. Their conversations remained mystery to the other members of the Institute. Such was the mental prowess of Gödel that economist Oskar Morgenstern (1902–1977), who was a mutual friend of Gödel and Einstein, recounted that toward the end of his life Einstein confided that "he came to the Institute merely...to have the privilege of walking home with Gödel."

Gödel's legacy is enormous. The Kurt Gödel Society was founded in 1987. It is an international organization, which promotes research in logic, philosophy, and the history of mathematics. But Gödel is most remembered for his celebrated incompleteness theorems, to which we now turn. More information about Gödel's life and contributions can be found in the following books [24, 8, 29, 22, 30, 27, 31, 9, 32, 28, 20, 33, 12, 18].

3 Self-referential statements and paradoxes

In order to understand better the main idea behind Gödel's proof of his incompleteness theorems, we need to give a brief account of self-referential statements, which were the main source of paradoxes that haunted the foundations of mathematics.

We start our brief account of self-referential statements with the well-known *Liar Paradox*, which is attributed to Epimenides—a semi-mythical Greek philosopher of sixth century B.C. Consider the following sentence: "This sentence is false." Is the sentence true or false?

Exercise 1 First assume that the sentence is true. What can you conclude? Now assume that the sentence is false. What conclusion can you reach? Is there a paradox hidden in the sentence? Why? How would you resolve it? Provide your explanations.

There are many variations of the liar paradox. One is the well-known *Barber Paradox* that Russell was very fond of. Suppose there is a town with a male barber who shaves those and only those men in town who do not shave themselves. Then the question is whether the barber shaves himself.

Exercise 2 Reason that this scenario is paradoxical. Try to resolve the problem. Provide your explanations.

Another version of the paradox is a story from the Spanish Inquisition. A particularly cruel Spanish inquisitor told a Jew to make one statement about himself. If the statement was false, he would be executed; if the statement was true, he would be hanged. The Jew responded by declaring "Today I am going to be executed!"

Exercise 3 Is the declaration of the Jew paradoxical? What should the inquisitor do? Explain your reasoning.

Similar paradoxes also occur in mathematics, namely in set theory. We next address the celebrated *Russell Paradox*, which showed that set theory as developed by the founder of set theory Georg Cantor (1845–1918) and Frege is inconsistent. Let $S = \{X : X \notin X\}$; that is, S is the *collection* of all those sets that do not belong to itself.

Exercise 4 Is it true that $S \in S$? Is it true that $S \notin S$? Explain. What can you conclude about S? What can you say about set theory developed by Cantor and Frege? Try to give your own resolution of the problem. Provide explanations of your answers.

A similar paradox arises when dealing with ordinals in set theory. This was already noticed by Cantor, who discovered ordinals, but was first stated clearly by the Italian mathematician Cesare Burali-Forti (1861–1931). The paradox is known as the *Burali-Forti Paradox*. In order to discuss it, we recall that an *ordinal* is a set α satisfying the following two conditions:

1. α is *transitive*; that is, $\gamma \in \beta$ and $\beta \in \alpha$ imply $\gamma \in \alpha$.
2. α is *well-ordered*; that is, each nonempty subset β of α has a $\in$-least element (in other words, there exists $\gamma \in \beta$ such that for each $\delta \in \beta$ we have $\gamma \in \delta$ and $\delta \notin \gamma$).

Ordinals have many useful properties. For example, for each ordinal α, the following are easy to prove: (i) if $\beta \in \alpha$, then β is also an ordinal; (ii) $\beta \subseteq \alpha$ iff $\beta \in \alpha$ or $\beta = \alpha$; and (iii) $\alpha = \bigcup\{\beta : \beta \in \alpha\}$. (These are good exercises to do!) The most important property of ordinals that we will use is that no ordinal can be an element of itself.

Exercise 5 Let α be an ordinal. Show that $\alpha \notin \alpha$.

Now let $\mathbf{O}$ be the *collection* of all ordinals. Burali-Forti's paradox arises when we ask whether or not $\mathbf{O}$ is an element of itself.

Exercise 6 Is $\mathbf{O}$ transitive? Is $\mathbf{O}$ well-ordered? Justify your answers.

Exercise 7 Based on Exercise 6, can you conclude that $\mathbf{O} \in \mathbf{O}$? Does this lead to a contradiction? Why? Can you resolve the Burali-Forti paradox? Explain your answers.

As we have seen, self-referential statements are a source of paradoxes in set theory as developed by Cantor and Frege. Since all of mathematics can be built on the basis of set theory, it is desirable to free set theory of all paradoxes. This is exactly what Hilbert set to do. He wanted to develop mathematics on a strict axiomatic basis and show that it is free of contradictions. But, as Gödel showed, Hibert's program contained serious flaws. In the remaining of this section we give an intuitive idea behind Gödel's incompleteness theorems.

Let $\mathcal{L}$ be a formal (read: first-order) language and let T be a theory in $\mathcal{L}$. We assume that T is consistent, which means that T does not prove contradiction. Let φ assert its own unprovability in T. By this we mean that φ states that it is not provable in T. For now suppose that $\mathcal{L}$ is capable of expressing φ. Then it is legitimate to ask whether or not T proves φ.

Exercise 8 Assume that φ is provable in T. What can you conclude? Explain your answer.

Exercise 9 Now assume that the negation of φ is provable in T. What conclusion can you reach? Explain your answer.

Based on your answers to Exercises 8 and 9, you should be able to conclude that neither φ nor $\neg\varphi$ is provable in T. Therefore, φ is undecidable in T. However, based on how we defined φ, it is true.

Exercise 10 Reason why and explain your reasoning.

Thus, φ is an example of a true sentence in the language of T, which is undecidable in T. This is Gödel's first incompleteness theorem in a nutshell! Of course, the key assumption that we made is that φ can be expressed in the language of T. This is a rather nontrivial assumption. Gödel's fundamental contribution was in showing that if T is an effectively definable mathematical theory which is capable of proving Peano's postulates of elementary arithmetic, then we can express sentences like φ in the language of T. How to do this will be discussed later in the project. Right now we turn our attention to an informal discussion of how to obtain Gödel's second incompleteness theorem from the first incompleteness theorem.

Let T be as above. If sentences like φ are expressible in the language $\mathcal{L}$ of T, then we can also express in $\mathcal{L}$ a sentence asserting that T is a consistent theory.

Exercise 11 How would you express that T is consistent in $\mathcal{L}$? Explain your answer.

Let ψ be such a sentence, and consider $\psi \to \varphi$.

Exercise 12 Give your own reasoning that $\psi \to \varphi$ expresses "If T is consistent, then I'm not provable in T."

Exercise 13 Next show that $\psi \to \varphi$ is provable in T. Hint: Recall that φ expresses its own unprovability in T and use Gödel's first incompleteness theorem.

Exercise 14 Finally, show that if T is consistent, then T cannot prove its own consistency. Hint: Use Exercise 13.

Thus, we arrive at Gödel's second incompleteness theorem! Again, the informal reasoning given above would become formal only if we are able to formalize sentences like ψ and φ within T. As Gödel has shown, this can be done if T is an effectively definable mathematical theory which is capable of proving Peano's postulates of elementary arithmetic. The remaining of the project is dedicated to this task.

4 Peano Arithmetic

In order to describe Gödel's technique in detail, we need to give a formal account of Peano Arithmetic. The first development of arithmetic can already be found in Euclid's "Elements" (Books VII–IX). More formal development of arithmetic was undertaken much later, in the second half of the nineteenth century, first by Hermann Grassmann (1809–1877) [19], and later by Frege [13], Richard Dedekind (1833–1916) [11], and Giuseppe Peano (1858–1932) [25]. The works of Frege, Dedekind, and Peano were independent of each other. Peano's 1889 paper became an instant classic. His postulates defining an axiomatic theory of arithmetic became known as *Peano's postulates*, and the first-order theory of arithmetic based on Peano's postulates is known as *Peano Arithmetic*. It is usually denoted by **PA**.[5]

We recall that the language $\mathcal{L}$ of **PA** is the standard first-order language with equality containing the constant 0 (zero), the unary function symbol s (the successor function), and two binary function symbols $+$ (addition) and $\cdot$ (multiplication). The axioms of **PA** are:

1. The standard axioms of first-order logic with equality.[6]
2. $\forall x \neg(s(x) = 0)$.
3. $\forall x \forall y(s(x) = s(y) \rightarrow x = y)$.
4. $\forall x(x + 0 = x)$.
5. $\forall x \forall y(x + s(y) = s(x + y))$.
6. $\forall x(x0 = 0)$.
7. $\forall x \forall y(x \cdot s(y) = xy + x)$.
8. $[\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(s(x)))] \rightarrow \forall x \varphi(x)$, where $\varphi(x)$ is a formula of $\mathcal{L}$ with one free variable x.

Note that axiom 2 states that 0 is not a successor, axiom 3 states that the successor function is 1-1, axioms 4 and 5 define addition, and axioms 6 and 7 define multiplication. Finally, axiom 8 is the first-order axiom-schema of mathematical induction.

This appears to be a rather simple first-order theory. However, as we will see shortly, it is rather powerful, mostly thanks to the axiom-schema of mathematical induction.

Exercise 15 Prove that the following are theorems of **PA**:

- $\forall x \forall y \forall z[(x + y) + z = x + (y + z)]$ (associativity of $+$).
- $\forall x \forall y(x + y = y + x)$ (commutativity of $+$). Hint: First show that $\forall x(x = 0 + x)$ and $\forall x \forall y[s(x) + y = s(x + y)]$.
- $\forall x \forall y \forall z[(xy)z = x(yz)]$ (associativity of $\cdot$).

[5] More on Peano's paper can be found in the historical project [6], which is available on our webpage http://www.cs.nmsu.edu/historical-projects/.

[6] We are assuming the basic knowledge of first-order logic; consult, e.g., the historical project [5], which is available on our webpage.

- $\forall x \forall y \forall z[x(y+z) = (xy) + (xz)]$ (left distributivity).
- $\forall x \forall y(xy = yx)$ (commutativity of $\cdot$). Hint: First show that $\forall x(x = 0x)$ and $\forall x \forall y[s(x) \cdot y = xy + y]$.
- $\forall x \forall y \forall z[(x+y)z = (xz) + (yz)]$ (right distributivity).
- $\forall x \forall y \forall z(x + z = y + z \rightarrow x = y)$ (cancellation law for $+$).
- $\forall x \forall y \forall z[(\neg(z = 0) \wedge xz = yz) \rightarrow x = y]$ (cancellation law for $\cdot$).

In **PA** we can introduce names for natural numbers as follows:

$1 = s(0)$

$2 = s(s(0)) = s(1)$

$3 = s(s(s(0))) = s(s(1)) = s(2)$

$\vdots$

Exercise 16 Prove that the following are theorems of **PA**:

- $\forall x[s(x) = x + 1]$.
- $\forall x(x \cdot 1 = x)$.
- $\forall x[\neg(x = 0) \rightarrow \exists y(x = s(y))]$.

In **PA** we can define $<$ and $\leq$ as follows:

$x < y$ iff $\exists z[\neg(z = 0) \wedge (x + z = y)]$.

$x \leq y$ iff $(x < y) \vee (x = y)$.

Exercise 17 Prove that the following are theorems of **PA**:

- $<$ is a strict linear order (that is, $<$ is irreflexive, transitive, and satisfies the trichotomy law).
- $\leq$ is a linear order (that is, $\leq$ is reflexive, transitive, and $\forall x \forall y[(x \leq y) \vee (y \leq x)]$).
- $\forall x(0 \leq x)$.
- $\forall x(0 < x + 1)$.
- $\forall x \forall y(x < y \leftrightarrow x + 1 \leq y)$.
- $\forall x \forall y(x \leq y \leftrightarrow x < y + 1)$.
- $\forall x(x < x + 1)$.
- $0 < 1$, $1 < 2$, $2 < 3$, ...

We can also show that the axiom-schema of mathematical induction implies the strong induction axiom, which implies the least number principle, which in turn implies the method of infinite descent. Thus, all three principles of strong induction, least number principle, and method of infinite descent are provable in **PA**.

Exercise 18 Prove that the following are theorems of **PA**:

- $\forall x[\forall y(y < x \rightarrow \varphi(y)) \rightarrow \varphi(x)] \rightarrow \forall x\varphi(x)$ (strong induction).
- $\exists x\varphi(x) \rightarrow \exists y[\varphi(y) \wedge \forall z(z < y \rightarrow \neg\varphi(z))]$ (least number principle).
- $\forall x[\varphi(x) \rightarrow \exists y(y < x \wedge \varphi(y))] \rightarrow \forall x\neg\varphi(x)$ (method of infinite descent).

In **PA** we can define divisibility as follows:

$x|y$ iff $\exists z(y = xz)$.

Exercise 19 Prove that the following are theorems of **PA**:

- $\forall x(x|x)$.
- $\forall x(x|0)$.
- $\forall x(1|x)$.
- $\forall x\forall y\forall z(x|y \wedge y|z \rightarrow x|z)$.
- $\forall x\forall y[(\neg(x = 0) \wedge x|y) \rightarrow x \leq y]$.
- $\forall x\forall y[(x|y \wedge y|x) \rightarrow x = y]$.
- $\forall x\forall y\forall z(x|y \rightarrow x|yz)$.
- $\forall x\forall y\forall z[(x|y \wedge x|z) \rightarrow x|(y + z)]$.

In **PA** we can also express and prove the division algorithm and the fundamental theorem of arithmetic.

Exercise 20 First express and then prove the division algorithm in **PA**.

Exercise 21 First express in **PA** that p is a prime number. Then show that it is provable in **PA** that if p is a prime number and p divides xy, then p divides x or p divides y. Hint: Use the division algorithm.

Exercise 22 How would you prove in **PA** Euclid's theorem that there are infinitely many prime numbers? Hint: Can you show that for each x there exists a prime number p such that $x < p$? What can you conclude from there?

Exercise 23 How would you express and prove the fundamental theorem of arithmetic in **PA**? Hint: Can you code in **PA** the sequence of prime numbers? Can you code in **PA** the sequence of pairs of prime numbers? Can you code in **PA** the sequence of n-tuples of prime numbers?

We conclude this section by addressing first-order models of **PA**. Let $\mathbb{N}$ denote the set of natural numbers and let $s(n) = n + 1$ for all $n \in \mathbb{N}$.

Exercise 24 Show that $\mathfrak{N} = (\mathbb{N}, 0, s, +, \cdot)$ is a model of **PA**.

Exercise 25 Are there models of **PA** which are not isomorphic to $\mathfrak{N}$? Justify your answer. Hint: Use the Löwenheim-Skolem theorem.[7]

Exercise 26 Justify that there exist even countable models of **PA**, which are not isomorphic to $\mathfrak{N}$. This is known as *Skolem's paradox*. Hint: Use the compactness and Löwenheim-Skolem theorems.[8]

In fact, there exist uncountably many non-isomorphic countable models of **PA**, known as *nonstandard* models of **PA**. Try to reason why. This is a nontrivial, but good exercise.

5 The first incompleteness theorem

Now that we have learned so much about **PA**, we are ready for Gödel's first incompleteness theorem. As we have seen, there are many non-isomorphic models of **PA**. In fact, there are many non-isomorphic countable models of **PA**. Therefore, it should be of less surprise that there might exist sentences φ which are true in $\mathfrak{N}$, but not derivable in **PA**.

Exercise 27 Reason why.

What we are after is a sentence φ (in the language of **PA**) which is true in $\mathfrak{N}$, but is undecidable in **PA**; that is, neither φ nor $\neg\varphi$ is provable in **PA**. As we saw in Section 3, if φ expresses its own unprovability in **PA**, then φ is undecidable in **PA**, yet it is true. The question, of course, is whether we can express φ within **PA**. This is the very question this section is dedicated to.

5.1 Gödel numbers

The main idea of Gödel was to translate the language $\mathcal{L}$ of **PA** into a numeric code in such a way that $\mathcal{L}$ can talk about itself. This is achieved by means of *Gödel numbers*. Since $\mathcal{L}$ is countable, there is a bijection $\alpha : \mathcal{L} \to \mathbb{N}$. With each expression $e = s_1 \dots s_n$ of $\mathcal{L}$, where each s_i is a symbol of $\mathcal{L}$, we associate a unique natural number as follows. Define a function β from the set of expressions of $\mathcal{L}$ to $\mathbb{N}$ by

$$\beta(e) = 2^{\alpha(s_1)} 3^{\alpha(s_2)} \dots p_n^{\alpha(s_n)},$$

where $e = s_1 \dots s_n$ and $2, 3, \dots, p_n$ are the first n prime numbers.

[7] The Löwenheim-Skolem theorem states that if a first-order theory T has an infinite model, then it has models of any infinite cardinality. Versions of it were proved by the German mathematician Leopold Löwenheim (1878–1957) in 1915 and by the Norwegian mathematician Thoralf Skolem (1887–1963) in 1920. The most general form of the theorem was proved by the famous Polish mathematician Alfred Tarski (1901–1983) in 1928 and the Russian mathematician Anatoly Maltsev (1909–1967) in 1936. More on the Löwenheim-Skolem theorem can be found in the historical project [5], which is available on our webpage http://www.cs.nmsu.edu/historical-projects/.

[8] The compactness theorem states that if Γ is a set of first-order sentences such that each finite subset of Γ has a model, then Γ has a model. The compactness theorem was first proved by Maltsev in 1936.

Exercise 28 Show that β is 1-1. Also give an argument that both n and $\alpha(s_i)$ can be computed from $\beta(e)$.

Since we are interested in whether or not a given sentence is provable in **PA**, and as proofs are finite sequences of expressions of $\mathcal{L}$, we would also like to associate a unique natural number with each finite sequence $\mathcal{E} = (e_1, \ldots, e_n)$ of expressions of $\mathcal{L}$. Define a function γ from the set of finite sequences of expressions of $\mathcal{L}$ to $\mathbb{N}$ by

$$\gamma(\mathcal{E}) = 2^{\beta(e_1)} 3^{\beta(e_2)} \ldots p_n^{\beta(e_n)},$$

where $\mathcal{E} = (e_1, \ldots, e_n)$ and $2, 3, \ldots, p_n$ are the first n prime numbers.

Let $\gamma(e) = \gamma(\mathcal{E})$ if $\mathcal{E} = (e)$; also let $\gamma(s) = \gamma(e)$ if $e = s$. We refer to the numbers $\gamma(s)$, $\gamma(e)$, and $\gamma(\mathcal{E})$ as the *Gödel numbers* of s, e, and $\mathcal{E}$, respectively. It is common to denote them by $\ulcorner s \urcorner$, $\ulcorner e \urcorner$, and $\ulcorner \mathcal{E} \urcorner$, respectively.

Exercise 29 Show that γ is 1-1. Also give an argument that both n and $\beta(e_i)$ can be computed from $\gamma(\mathcal{E})$.

Let $P(x_1, \ldots, x_n)$ be a property of natural numbers. We call $P(x_1, \ldots, x_n)$ *expressible* (in **PA**) if there exists a formula $\varphi(x_1, \ldots, x_n)$ of $\mathcal{L}$ such that for all $m_1, \ldots, m_n \in \mathbb{N}$ we have $P(m_1, \ldots, m_n)$ holds in $\mathfrak{N}$ iff $\mathfrak{N} \models \varphi(\mathbf{m_1}, \ldots, \mathbf{m_n})$. Here $\mathbf{m_1}, \ldots, \mathbf{m_n}$ are the names in $\mathcal{L}$ of the natural numbers $m_1, \ldots, m_n \in \mathbb{N}$. We call a property P of symbols, expressions, and sequences of expressions of $\mathcal{L}$ *expressible* if the corresponding property of their Gödel numbers is expressible.

Exercise 30 Show that the property "is a variable (of the language $\mathcal{L}$)" is expressible. Hint: Show that the corresponding property on Gödel numbers is expressible.

Exercise 31 Show that the property "is a term (of $\mathcal{L}$)" is expressible.

Exercise 32 Show that the property "is a formula (of $\mathcal{L}$)" is expressible.

Exercise 33 Show that the property "is a sentence" (of $\mathcal{L}$)" is expressible.

Exercise 34 Show that the property "is an axiom (of **PA**)" is expressible.

Exercise 35 Show that the property "is a proof (in **PA**)" is expressible.

Exercise 36 Show that the property "$\mathcal{E}$ is a proof (in **PA**) of the formula φ (of $\mathcal{L}$)" is expressible.

As a result, there is a formula $Pr(x, y)$ of $\mathcal{L}$ such that for any two natural numbers $m, n \in \mathbb{N}$, we have $\mathfrak{N} \models Pr(\mathbf{m}, \mathbf{n})$ iff m is the Gödel number of a proof of a formula of $\mathcal{L}$ whose Gödel number is n.

Exercise 37 For a formula φ of $\mathcal{L}$, show that $(\exists x) Pr(x, \ulcorner \varphi \urcorner)$ expresses that φ is a theorem of **PA**.

5.2 The diagonalization lemma

Next we examine the diagonalization lemma, the key ingredient of Gödel's incompleteness theorems. The diagonalization lemma states that for each formula $\varphi(x)$ of $\mathcal{L}$, with the only free variable x, there exists a sentence ψ of $\mathcal{L}$ such that $\mathbf{PA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$.

Clearly if $\mathbf{PA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$, then $\mathfrak{N} \models \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$. For our slightly informal proof of the first incompleteness theorem, this weaker version of the diagonalization lemma suffices. In order to see why it holds, let $\varphi(x)$ be a formula of $\mathcal{L}$ with the only free variable x. Set $m = \ulcorner \varphi(x) \urcorner$ and $q = \ulcorner \varphi(\mathbf{m}) \urcorner$; that is, q is the Gödel number of $\varphi(\mathbf{m})$, where m is the Gödel number of $\varphi(x)$. This property of "being the Gödel number of ($\varphi(\mathbf{m})$, where m is) the Gödel number (of $\varphi(x)$)" is expressible in $\mathcal{L}$. Let $D(x, y)$ be the formula (with two free variables) expressing this property.

Now let m be the Gödel number of $(\forall x)(D(y, x) \rightarrow \varphi(x))$ and set

$$\psi = (\forall x)(D(\mathbf{m}, x) \rightarrow \varphi(x)).$$

Clearly ψ is a sentence. Let q be the Gödel number of ψ. Then $\mathfrak{N} \models D(\mathbf{m}, \mathbf{q})$.

Exercise 38 Reason why.

It is left to be shown that $\mathfrak{N} \models \psi \rightarrow \varphi(\mathbf{q})$ and $\mathfrak{N} \models \varphi(\mathbf{q}) \rightarrow \psi$.

Exercise 39 Show that $\mathfrak{N} \models \psi \rightarrow \varphi(\mathbf{q})$. Hint: Assume that $\mathfrak{N} \models \psi$. Then $\mathfrak{N} \models D(\mathbf{m}, \mathbf{q}) \rightarrow \varphi(\mathbf{q})$. Since $\mathfrak{N} \models D(\mathbf{m}, \mathbf{q})$, conclude that $\mathfrak{N} \models \varphi(\mathbf{q})$. Therefore, $\mathfrak{N} \models \psi \rightarrow \varphi(\mathbf{q})$.

Exercise 40 Show that $\mathfrak{N} \models \varphi(\mathbf{q}) \rightarrow \psi$. Hint: Assume that $\mathfrak{N} \models \varphi(\mathbf{q})$. Then $\mathfrak{N} \models D(\mathbf{m}, \mathbf{q}) \rightarrow \varphi(\mathbf{q})$. Moreover, if $n \neq q$, then $\mathfrak{N} \not\models D(\mathbf{m}, \mathbf{n})$. Therefore, $\mathfrak{N} \models (\forall x)(D(\mathbf{m}, x) \rightarrow \varphi(x))$. Thus, $\mathfrak{N} \models \psi$, and so $\mathfrak{N} \models \varphi(\mathbf{q}) \rightarrow \psi$.

5.3 Gödel's first incompleteness theorem

We are finally ready to give a (slightly informal) proof of Gödel's first incompleteness theorem. Recall that our goal is to find a sentence φ of $\mathcal{L}$ such that $\mathfrak{N} \models \varphi$, but neither φ nor $\neg\varphi$ is provable in **PA**. Also recall that $Pr(\mathbf{m}, \mathbf{n})$ is the formula of $\mathcal{L}$ expressing that m is the Gödel number of a proof in **PA** of a formula whose Gödel number is n. Let

$$\varphi(x) = (\forall x)\neg Pr(x, y).$$

Then $\varphi(\ulcorner \alpha \urcorner) = (\forall x)\neg Pr(x, \ulcorner \alpha \urcorner)$ expresses that the formula α is not provable in **PA**.

Exercise 41 Reason why.

By the diagonalization lemma, there exists a sentence $\mathcal{G}$ such that

$$\mathfrak{N} \models \mathcal{G} \leftrightarrow (\forall x)\neg Pr(x, \ulcorner \mathcal{G} \urcorner).$$

Exercise 42 Reason that $\mathcal{G}$ asserts that $\mathcal{G}$ is not provable in **PA**. In other words, $\mathcal{G}$ asserts its own unprovability in **PA**!

We call $\mathcal{G}$ a *Gödel sentence* for **PA**.

Exercise 43 Show that $\mathbf{PA} \nvdash \mathcal{G}$. Hint: Assume that $\mathbf{PA} \vdash \mathcal{G}$. Then $\mathfrak{N} \models \mathcal{G}$. By the diagonalization lemma, $\mathfrak{N} \models \mathcal{G} \leftrightarrow (\forall x)\neg Pr(x, \ulcorner \mathcal{G} \urcorner)$. Therefore, $\mathfrak{N} \models (\forall x)\neg Pr(x, \ulcorner \mathcal{G} \urcorner)$. But this sentence asserts that $\mathcal{G}$ is not provable in **PA**, contradicting our assumption.

Exercise 44 Show that $\mathbf{PA} \nvdash \neg\mathcal{G}$. Hint: Assume that $\mathbf{PA} \vdash \neg\mathcal{G}$. Then $\mathfrak{N} \models \neg\mathcal{G}$. Since $\mathfrak{N} \models \neg\mathcal{G} \leftrightarrow \neg(\forall x)\neg Pr(x, \ulcorner \mathcal{G} \urcorner)$, we obtain $\mathfrak{N} \models \neg(\forall x)\neg Pr(x, \ulcorner \mathcal{G} \urcorner)$, hence $\mathfrak{N} \models (\exists x) Pr(x, \ulcorner \mathcal{G} \urcorner)$. Therefore, there exists $n \in \mathbb{N}$ such that $\mathfrak{N} \models Pr(\mathbf{n}, \ulcorner \mathcal{G} \urcorner)$. But this sentence asserts that $\mathcal{G}$ is provable in **PA**, contradicting our assumption.

Exercise 45 Show that $\mathfrak{N} \models \mathcal{G}$. Hint: Show that $\mathfrak{N} \models (\forall x)\neg Pr(x, \ulcorner \mathcal{G} \urcorner)$. Then apply the diagonalization lemma.

As a result, we found a sentence $\mathcal{G}$ such that $\mathcal{G}$ is true, but it is undecidable in **PA**. We can, of course, add $\mathcal{G}$ to **PA** and consider a new theory $T = \mathbf{PA} \cup \{\mathcal{G}\}$. But if we repeat the Gödel reasoning for T, then we arrive at a new sentence $\mathcal{H}$ which is true, but undecidable in T. Adding $\mathcal{H}$ to T will not help either because the Gödel argument will still apply to $T \cup \{\mathcal{H}\}$. Thus, if $\mathrm{Th}(\mathfrak{N})$ denotes the set of all sentences satisfiable in $\mathfrak{N}$, then $\mathrm{Th}(\mathfrak{N})$ is neither effectively axiomatizable nor decidable.

Exercise 46 Reason why.

Note that the Gödel argument works for any consistent effectively axiomatizable first-order theory which is strong enough to prove Peano's postulates for natural numbers.

6 The second incompleteness theorem

After learning Gödel's first incompleteness theorem, we are ready for Gödel's second incompleteness theorem. In a sense, its consequences are even more remarkable than those of the first incompleteness theorem. By a sufficiently strong theory we mean a first-order theory which can prove Peano's postulates for natural numbers. The second incompleteness theorem then states that if an effectively axiomatizable sufficiently strong first-order theory is consistent, then it cannot prove its own consistency.

Let T be an effectively axiomatizable sufficiently strong first-order theory. We recall that $Pr(\mathbf{m}, \mathbf{n})$ expresses that m is the Gödel number of a proof in T of a formula whose Gödel number is n.

Exercise 47 Reason that $Pr(\mathbf{m}, \ulcorner \bot \urcorner)$ expresses that T is inconsistent. Also reason that $(\forall x)\neg Pr(x, \ulcorner \bot \urcorner)$ expresses that T is consistent.

Let Con_T denote the sentence $(\forall x)\neg Pr(x, \ulcorner \bot \urcorner)$. Then Con_T is a sentence of T expressing the consistency of T. The second incompleteness theorem then states that $T \nvdash Con_T$. To see why, let $\mathcal{G}$ be a Gödel sentence for T, and consider the sentence $Con_T \rightarrow \mathcal{G}$.

Exercise 48 Reason that the sentence $Con_T \rightarrow \mathcal{G}$ states that if T is consistent, then $\mathcal{G}$ is not provable in T.

Exercise 49 Show that $T \vdash Con_T \to \mathcal{G}$. Hint: Recall that the first incompleteness theorem (formulated for T) asserts that if T is consistent, then $\mathcal{G}$ is undecidable in T. In particular, $\mathcal{G}$ is not provable in T. But this is exactly what $Con_T \to \mathcal{G}$ states. Thus, by the first incompleteness theorem, $T \vdash Con_T \to \mathcal{G}$.

Exercise 50 Deduce that $T \nvdash Con_T$. Hint: Use Exercise 49 and Gödel's first incompleteness theorem.

As a result, we obtain that if T is consistent, then T cannot prove its own consistency, thus arriving at the second incompleteness theorem.

7 Notes to the Instructor

The project is designed for an upper level undergraduate course in mathematical logic. It is more suited for a second semester course in mathematical logic. The project assumes that students are familiar with basics of first-order logic, including its syntax and semantics, completeness, Löwenheim-Skolem, and compactness theorems. The whole course may be designed around the project. Since Gödel's incompleteness theorems are rather challenging, the project is designed so that it first explains the informal ideas behind self-referential statements and paradoxes, then treats Peano Arithmetic on formal basis, and finally tackles Gödel's incompleteness theorems. The proofs of incompleteness theorems are also mostly done from an informal point of view. Instructors may wish to discuss how to convert them into purely formal statements. After the completion of the project, instructors may wish to discuss Tarski's theorem on undefinability of truth, as well as Henkin sentences and Löb's theorem. If there is enough time left, it would also be fitting to cover the Church-Turing theorem on undecidability of first-order logic. There are plenty of exercises in the project, some of them quite challenging. Instructors may wish to pick and choose the exercises they find relevant for their needs. They may also want to spend some class time on guiding students through some of them.

References

[1] Janet Barnett, *Applications of Boolean algebra: Claude Shannon and circuit design*, Available from the webpage `http://www.cs.nmsu.edu/historical-projects/`.

[2] _____, *Boolean algebra as an abstract structure: Edward V. Huntington and axiomatization*, `http://www.cs.nmsu.edu/historical-projects/`.

[3] _____, *Origins of Boolean algebra in the logic of classes: George Boole, John Venn and C. S. Peirce*, `http://www.cs.nmsu.edu/historical-projects/`.

[4] Janet Barnett, Guram Bezhanishvili, Hing Leung, Jerry Lodder, David Pengelley, and Desh Ranjan, *Historical projects in discrete mathematics and computer science*, Resources for Teaching Discrete Mathematics, B. Hopkins, editor, MAA, 2009, pp. 165–274.

[5] Guram Bezhanishvili, *Henkin's method and the completeness theorem*, Available from the webpage `http://www.cs.nmsu.edu/historical-projects/`.

[6] ———, *Peano arithmetic*, `http://www.cs.nmsu.edu/historical-projects/`.

[7] Guram Bezhanishvili and Wesley Fussner, *Introduction to symbolic logic*, Available from the webpage `http://www.cs.nmsu.edu/historical-projects/`.

[8] Martin Davis, *The undecidable. Basic papers on undecidable propositions, unsolvable problems and computable functions*, Edited by Martin Davis, Raven Press, Hewlett, N.Y., 1965.

[9] John W. Dawson, Jr., *Logical dilemmas: The life and work of Kurt Gödel*, A. K. Peters Ltd., Wellesley, MA, 1997.

[10] ———, *Max Dehn, Kurt Gödel, and the trans-Siberian escape route*, Notices Amer. Math. Soc. **49** (2002), no. 9, 1068–1075.

[11] Richard Dedekind, *Was sind und was sollen die Zahlen?*, Vieweg, Braunschweig, 1888.

[12] Torkel Franzén, *Gödel's theorem: An incomplete guide to its use and abuse*, A. K. Peters Ltd., Wellesley, MA, 2005.

[13] Gottlob Frege, *Die Grundlagen der Arithmetik*, Köbner, Breslau, 1884.

[14] Kurt Gödel, *Die Vollständigkeit der Axiome des logischen Funktionenkalküls*, Monatsh. Math. Phys. **37** (1930), 349–360.

[15] ———, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatsh. Math. Phys. **38** (1931), 173–198.

[16] ———, *On undecidable propositions of formal mathematical systems*, Mimeographed lecture notes by S. C. Kleene and J. B. Rosser, Institute for Advanced Study, Princeton, N.J., 1934.

[17] ———, *The Consistency of the Continuum Hypothesis*, Annals of Mathematics Studies, no. 3, Princeton University Press, Princeton, N. J., 1940.

[18] Rebecca Goldstein, *Incompleteness: The proof and paradox of Kurt Gödel*, Great Discoveries, Norton & Company, New York, 2005.

[19] Hermann Grassmann, *Lehrbuch der Arithmetik*, Enslin, Berlin, 1861.

[20] I. Grattan-Guinness, *The search for mathematical roots, 1870–1940: Logics, set theories and the foundations of mathematics from Cantor through Russell to Gödel*, Princeton Paperbacks, Princeton University Press, Princeton, NJ, 2000.

[21] David Hilbert and Wilhelm Ackermann, *Grundzüge der theoretischen Logik*, Springer–Verlag, Berlin, 1928.

[22] Douglas R. Hofstadter, *Gödel, Escher, Bach: an eternal golden braid*, Basic Books Inc. Publishers, New York, 1979.

[23] Jerry Lodder, *Deduction through the ages: A history of truth*, Available from the webpage `http://www.cs.nmsu.edu/historical-projects/`.

[24] Ernest Nagel and James R. Newman, *Gödel's proof*, New York University Press, New York, 1958.

[25] Giuseppe Peano, *Arithmetices principia, nova methodo exposita*, Bocca, Torino, 1889.

[26] Bertrand Russell and Alfred North Whitehead, *Principia Mathematica*, Cambridge University Press, Cambridge, England, 1910 (Vol. 1), 1912 (Vol. 2), 1913 (Vol. 3).

[27] Raymond M. Smullyan, *Gödel's incompleteness theorems*, Oxford Logic Guides, vol. 19, The Clarendon Press Oxford University Press, New York, 1992.

[28] ———, *Forever undecided*, Oxford University Press, New York, 2000.

[29] Jean van Heijenoort, *From Frege to Gödel. A source book in mathematical logic, 1879–1931*, Harvard University Press, Cambridge, Mass., 1967.

[30] Hao Wang, *Reflections on Kurt Gödel*, A Bradford Book, MIT Press, Cambridge, MA, 1987.

[31] ———, *A logical journey: From Gödel to philosophy. Final editing and with an addition to the preface by Palle Yourgrau and Leigh Cauman*, Representation and Mind, MIT Press, Cambridge, MA, 1996.

[32] Palle Yourgrau, *Gödel meets Einstein: Time travel in the Gödel universe. Revised edition of the disappearance of time [Cambridge Univ. Press, Cambridge, 1991]*, Open Court Publishing Co., Chicago, IL, 1999.

[33] ———, *A world without time: The forgotten legacy of Gödel and Einstein*, Basic Books, New York, 2005.