Gödel's Incompleteness Theorems

Gödel's Incompleteness Theorems belong to the most remarkable achievements of the 20^{th} century mathematics, shedding light on the limitations of formal methods and still raising philosophical questions about the nature of human thought, its relations to our brains and to computers, etc.

> *Kurt Gödel's achievement in modern logic is singular and monumental — indeed, it is more than a monument, it is a landmark which will remain visible far in space and time.* [...] *The subject of logic has certainly completely changed its nature and possibilities with Gödel's achievement.* (John von Neumann)

Liar's Paradox and the Paradoxes of Russell and Berry

Consider the following self-referential sentence:

"This sentence is not true."

At least at a glance it look like a proposition, thus is seems legitimate to ask the question: "Is it true or false?" If it is false, then it must be true. Similarly, if it is true, then it cannot be true, hence it must be false. We can conclude that it is true if and only if it is not true. This is the strong version of the famous *Liar's Paradox*; its ancient version, also known as the *Epimenidus Paradox*, consists in the statement

"All the Cretans are liars,"

pronounced by the Cretan Epimenidus (tacitly assuming that liars always lie).

In everyday life we need not to worry too much about the Liar's Paradox. We can do away with it simply by marking that sentence as making no sense and not to care of it any more. However, the situation changes radically if such a self-referential sentence could be formulated within some formal deductive system like, e.g., an axiomatic first order theory. Such a theory would be necessarily inconsistent. This namely happened to the original version of Cantor's "naïve" Set Theory.

Cantor's Set Theory used the unlimited version of the *Comprehension Principle* in forming sets:

> *For any "reasonable" property $P(x)$ one can form the set $\{x\colon P(x)\}$ of all objects x having this property.*

However intuitively appealing this principle might appear, it is fairly hazy, unless we make clear which properties we consider as "reasonable". What's even worse, this principle enables to formulate a set-theoretical version of Liar's Paradox, namely *Russell's Paradox*, named after the British logician and philosopher Bertrand Russell:

Cantor's Comprehension Principle allows us to form the set

$$R = \{x\colon x \text{ is a set and } x \notin x\}$$

of all sets x not belonging to itself.

Then the question: "Does the set R belong to itself?" immediately produces a contradiction. Indeed, we have $R \in R \Leftrightarrow R \notin R$.

Thus the original version of Cantor's Set Theory is inconsistent; the unlimited Comprehension Principle makes it possible to reproduce the Liar's Paradox inside of this theory.

Liar's Paradox can be avoided by restricting Cantor's Comprehension Principle to the following limited form:

For every set M and any "reasonable" property $P(x)$ one can form the set $\{x \in M\colon P(x)\}$ of all objects x from the set M having this property.

Then the previous formation of the set R becomes illegal, and Russell's Paradox disappears. Instead, it is transformed to the following fact:

There is no set of all sets.

Indeed, if there were the set V of all sets, then we could legally form the set

$$R = \{x \in V\colon x \notin x\}$$

of all sets x not belonging to itself, and obtain the contradiction $R \in R \Leftrightarrow R \notin R$ once again.

Berry's Paradox demonstrates the need to clarify the vague concept of a "reasonable property" and that way to make clear which properties can be used even in the limited Comprehension Principle.

Consider the set A of all natural numbers which can be defined by some phrase of English language consisting of less than twenty words. Since the English language has a finite vocabulary, there are just finitely many English phrases consisting of less than twenty words. Hence the set A is finite, and, as the set $\mathbb{N}$ of all natural numbers is infinite, there exist natural numbers not belonging to the set A. In other words the complement $\mathbb{N} \smallsetminus A$ is nonempty, thus, according to the Well Ordering Principle, it contains the smallest element. Then this natural number is defined by the English phrase

"The smallest natural number which cannot be defined by any English phrase consisting of less than twenty words"

which has eighteen words, only. Hence the smallest element of the set $\mathbb{N} \smallsetminus A$ has to belong to the set A, as well. However, this is a contradiction, since $A \cap (\mathbb{N} \smallsetminus A) = \emptyset$.

In Quest for a Way Out of the Crisis

The discovery of paradoxes in Cantor's Set Theory at the turn of the 19^{th} and 20^{th} century threw the mathematics of that time into a deep crisis. Moreover, it happened shortly after Set Theory had become widely accepted and recognized as the universal foundations of the whole of mathematics, providing it with a general common language and a firm ground on which all mathematical branches could be formulated and presented in a uniform way. Therefore the task to find a way out of the crisis became highly acute.

Some mathematicians reacted by refusing completely the conception of actual infinity forming one of the cornerstones of Set Theory (H. Poincaré, L. E. J. Brouwer). Namely Brouwer established the doctrine of *intuitionism*, insisting that the infinity can be treated just as a potential and never completed process of growth or decay beyond any limit. He also proposed a revision of logic, refusing some classical logical laws (e.g., the Law of Excluded Middle $\varphi \vee \neg\varphi$, or the quantifier law $\neg(\forall\, x)\neg\varphi(x) \;\Rightarrow\; (\exists\, x)\varphi(x)$) as inapplicable within the realm of potentially infinite domains. The competing doctrine of *logicism* suggested to develop mathematics as a branch of logic (G. Frege, B. Russell, A. N. Whitehead) and to avoid the self-reference phenomenon, which they found responsible for the contradictions, by means of a fairly complicated hierarchy of the *Theory of Types*. However, none of these conceptions could compete with the approaches offered by the Set Theory making use of the full power of classical logic and, at the same time, avoiding the cumbersome hierarchy of the Theory of Types, along with preserving the conception of actually infinite sets.

The axiomatic system of Set Theory designed by Ernst Zermelo, and later on upgraded by Abraham Fraenkel, became the generally accepted foundations of most of the modern mathematics. The Paradoxes of Russell and Berry (and some similar ones) were avoided by a cautious formulation of the Scheme of Comprehension, allowing to single out new sets just as *subsets of sets given in advance* by means of *properties described by set-theoretical formulas.* Three exceptions of sets, still described by set-theoretical formulas, but not singled out from any in advance given set, are allowed by the Axioms of Pair, Union and Power Set.

No one was able to reproduce the known paradoxes, nor to produce any contradiction within the Zermelo-Fraenkel axiomatic system with the Axiom of Choice ZFC. Unfortunately, this does not exclude the possibility that, all the same, there are some contradictions, hidden deeply under the surface. This raised the task to *prove* the consistency of ZFC or of some other axiomatic system of Set Theory, capable to undertake the role of the foundations of mathematics. The project of proving the consistency of the foundations of mathematics was formulated by David Hilbert, the leading figure of the that time mathematics, who also designed the central notions and methods necessary for that purpose. The project is known under the name *Hilbert's Program.*

> *Hilbert's Program was an ambitious and wide-ranging project in the philosophy and foundations of mathematics. In order to "dispose of the foundational questions in mathematics once and for all", Hilbert proposed a two-pronged approach in 1921: first, classical mathematics should be formalized in axiomatic systems; second, using only restricted, "finitary" means, one should give proofs of the consistency of these axiomatic systems. Although Gödel's Incompleteness Theorems show that the program, as originally conceived, cannot be carried out, it had many partial successes, and generated important advances in logical theory and meta-theory, both at the time and since.*
>
> (Richard Zach, *Hilbert's Program Then and Now*, arXiv:math/0508572)

Gödel's Incompleteness Theorems

Preliminary Accounts

Consider the following self-referential sentence:

"This sentence is unprovable."

tacitly assuming that every sentence which is provable, is necessarily true. Once again we find legitimate to ask the question: "Is that sentence true or false?" If it is false, then it is provable, hence it must be true. This contradiction shows that it cannot be false, hence it is true. That way we have proved that this sentence is true, in other words, *we have proved* the sentence. Thus it is provable, hence, since it declares its own unprovability, it is false. At the same time, provable sentences must be true. It seems that we once again obtained a contradiction, closely related to Liar's Paradox.

However, this conclusion can be avoided by making precise the concept of provability. If it means *provability within some formal axiomatic system* (e.g., within some first order theory), then our proof of the above sentence is just an informal intuitive argumentation showing that it is true, and not a proof within that system. Moreover, statements about provability within a given formal system in general *do not belong* to that system, hence the question of their provability within that system makes no sense. Thus it seems that the threatening paradox can be swept away from the very beginning.

All the same, let us admit that some formal systems could perhaps satisfy the following two properties:

(1) There is a sufficiently extensive distinguished class of statements formulated in the language of that system such that all statements from this class which are provable in the system are true in some intuitively appealing meaning of this word.
(2) There is a statement belonging to the above mentioned distinguished class declaring its own unprovability within the system.

Then that system is necessarily incomplete in the following sense:

(3) There are intuitively true statements formulated in the language of the system (and even belonging to that distinguished class) which are unprovable within that system.

Namely the statement belonging to that distinguished class and declaring its own unprovability within the system is an example of an intuitively true statement which is not provable within the system.

Now, the reader probably can hardly suppress the feeling that the existence of such formal axiomatic systems (first order theories) is merely hypothetical, and in fact it should be possible to show that nothing like that can exist. Thus it might be rather surprising to realize what the young Austrian mathematician, logician and philosopher Kurt Gödel (born 1906 in Brno) has proved in 1930. Namely, according to his *First Incompleteness Theorem*, Peano Arithmetic, as well as any first order theory capable to serve as the foundations of a reasonable fragment of mathematics, like ZF or ZFC, provide examples of such axiomatic systems. According to his *Second Incompleteness Theorem*, such systems are capable to formulate a statement declaring their own consistency, nonetheless, if they are consistent, they are unable to prove it, though, in that case, the statement itself is true. As one of the consequences of Gödel's discoveries it became manifest that the goals of Hilbert's Program cannot be achieved.

The First Gödel Incompleteness Theorem

It is worth mentioning that Gödel worked within the intentions of Hilbert's Program and his Incompleteness Theorems appeared surprisingly on the way, without having been planned or anticipated in advance. We will skip almost all technical issues of Gödel's proof and begin with displaying some final results of his coding of formulas and proofs by natural numbers and representation of the provability relation by certain arithmetical predicate. It should be noted that our presentation differs considerable from Gödel's original one.

Informally, the First Gödel Incompleteness Theorem states that any consistent formal system which is sufficiently ample to include Peano Arithmetic is necessarily incomplete, either in the sense that it contains some true propositions about natural numbers which it cannot prove (semantic version), or in the sense that it contains certain arithmetical propositions which it can neither prove nor refute (syntactic version).

Let's begin with introducing some concepts necessary for describing more precisely the variety of first order theories to which Gödel's results apply. A first order theory T in a language with finitely many specific symbols is called *recursively axiomatizable* if it has just finitely many axioms or its axioms can be effectively recognized by some algorithm (e.g., by a computer program). A first order theory T is called *arithmetical* if there is some interpretation of Peano Arithmetic in this theory. This is to say that there are some formulas $\mathrm{Nat}(x)$, $\mathrm{Add}(x, y, z)$, $\mathrm{Mult}(x, y, z)$, $\mathrm{Zero}(x)$, $\mathrm{One}(x)$ in the language of T defining the concept of natural number, the operations of addition and multiplication of natural numbers and the distinguished objects 0 and 1, respectively, in such a way that for the structure of natural numbers thus obtained all the axioms of PA can be proved in T. If T is an arithmetical theory then a formula φ in the language of T is called *arithmetical* if it is built out of the "new" atomic formulas of the form $x = y$, $\mathrm{Add}(x, y, z)$, $\mathrm{Mult}(x, y, z)$, $\mathrm{Zero}(x)$, $\mathrm{One}(x)$ by means of logical connectives and *bounded quantifications* $(\forall\, x)(\mathrm{Nat}(x) \Rightarrow \varphi)$, $(\exists\, x)(\mathrm{Nat}(x) \wedge \varphi)$. An arithmetical theory T is called *arithmetically correct* if all the arithmetical sentences provable in T are satisfied in $(\mathbb{N}; +, \cdot, 0, 1)$.

An obvious example of a recursively axiomatizable arithmetically correct theory is the Peano Arithmetic itself. Other paradigmatic examples of such theories are recursive extensions of PA by axioms which are true in $(\mathbb{N}; +, \cdot, 0, 1)$, as well as various set theories like, e.g., ZF or ZFC.

Given an arithmetical sentence θ we will say that θ is *true* or *valid* or *satisfied* if it is satisfied in the standard model of Peano Arithmetic $(\mathbb{N}; +, \cdot, 0, 1)$. For an arithmetical sentence of the form $\psi(k_1, \ldots, k_n)$, where $\psi(x_1, \ldots, x_n)$ is an arithmetical formula and $k_1, \ldots, k_n$ are concrete natural numbers (constant arithmetical terms), we also use to say that $\psi(k_1, \ldots, k_n)$ *holds* or, simply, $\psi(k_1, \ldots, k_n)$ in that case.

Gödel developed a method of *coding* or *enumeration* by means of which all the arithmetical formulas in the language of an arithmetical theory T with a single free variable

x can be lined up in a sequence $\varphi_0(x), \varphi_1(x), \ldots, \varphi_n(x), \ldots$ in such a way that, for each n, the formula $\varphi_n(x)$ can be effectively constructed (e.g., by a program), and vice versa, for each arithmetical formula $\psi(x)$, its number n such that $\psi(x)$ coincides with $\varphi_n(x)$ can be effectively determined. If T is additionally recursively axiomatizable then also all proofs in T can be lined up in a sequence $\Delta_0, \Delta_1, \ldots, \Delta_k, \ldots$ in such a way that the correspondence $k \leftrightarrow \Delta_k$ can be effectively described (e.g., executed by some programs) in either direction. Moreover, in that case Gödel constructed two effectively decidable ternary arithmetical predicates $P(x, y, z)$ and $R(x, y, z)$ of provability and refutability, respectively, such that for any natural numbers k, m, n the following conditions are satisfied:

$P(m, n, k)$ if and only if Δ_k is a proof of the sentence $\varphi_n(m)$ in T;

$R(m, n, k)$ if and only if Δ_k is a proof of the sentence $\neg\varphi_n(m)$ in T.

At the same time the algorithmic decidability of the predicates $P(x, y, z)$ and $R(x, y, z)$ ensures that, for any $m, n, k \in \mathbb{N}$, the satisfaction of any of the statements $P(m, n, k)$, $\neg P(m, n, k)$, $R(m, n, k)$, $\neg R(m, n, k)$, respectively, in $(\mathbb{N}; +, \cdot, 0, 1)$ is equivalent to its provability in PA, henceforth in T. Namely the algorithm deciding whether $P(m, n, k)$ holds or not provides the proof either of the statement $P(m, n, k)$ or of its negation, and similarly for $R(m, n, k)$. Summing up, we have:

Theorem. *Assume that T is a consistent recursively axiomatizable arithmetical theory. Then, for any natural numbers m, n, k, the three conditions in each of the following four rows are equivalent:*

$$\begin{array}{rll}
P(m,n,k) & T \vdash P(m,n,k) & \Delta_k \text{ is a proof of the sentence } \varphi_n(m) \text{ in } T \\
R(m,n,k) & T \vdash R(m,n,k) & \Delta_k \text{ is a proof of the sentence } \neg\varphi_n(m) \text{ in } T \\
\neg P(m,n,k) & T \nvdash P(m,n,k) & T \vdash \neg P(m,n,k) \\
\neg R(m,n,k) & T \nvdash R(m,n,k) & T \vdash \neg R(m,n,k)
\end{array}$$

In particular, T decides both the statements $P(m, n, k)$ and $R(m, n, k)$ for any m, n, k.

Now, we have all the necessary ingredients needed for the formulation of Gödel's results. Consider the formula $\neg(\exists\, z)P(x, x, z)$. It has a single free variable, namely x, hence it occurs in the sequence $\{\varphi_n(x)\}_{n\in\mathbb{N}}$ under some number — let's denote it g. Thus $\varphi_g(x)$ is the above formula, and substituting the natural number g into it for x we obtain the sentence $\varphi_g(g)$, i.e., $\neg(\exists\, z)P(g, g, z)$, saying that, for no $z = k$, Δ_k is the proof of the sentence $\varphi_g(g)$. In other words, the meaning of that sentence is:

$\varphi_g(g)$: *"The sentence $\varphi_g(g)$ is not provable in T."*

Hence $\varphi_g(g)$ is an example of a self-referential sentence in the language of T declaring its own unprovability. On the other hand, the reader should keep in mind that $\varphi_g(g)$ is an arithmetical statement, like, e.g., $\neg(\exists\, x, y, z)\big((x + 1)^2 + (y + 2)^3 = (z + 3)^4\big)$, saying that the diophantic equation $(x+1)^2+(y+2)^3 = (z+3)^4$ has no solution in the domain of all natural numbers.

Thus our preliminary accounts entitle us to state the semantic version of Gödel's First Incompleteness Theorem.

First Gödel Incompleteness Theorem. [Semantic version] *If T is a recursively axiomatizable arithmetically correct first order theory then the Gödel's sentence $\varphi_g(g)$ is true in $(\mathbb{N}; +, \cdot, 0, 1)$, nonetheless, it is unprovable in T. Thus T is incapable to prove all the true arithmetical statements about natural numbers.*

In particular, neither PA nor any of the set theories like, e.g., ZF or ZFC, can prove all the true arithmetical statements about natural numbers.

Since we have no direct access to the infinite domain $\mathbb{N}$ of all natural numbers, the semantic concept of arithmetical truth playing a key role in the semantic version of Gödel's First Incompleteness Theorem "smells of metaphysics" and may evoke some bewilderment in the reader. It relies on our belief that $(\mathbb{N};\, +, \cdot, 0, 1)$ is a model of PA, which, however, can hardly be considered as an obvious or firmly and doubtlessly established fact. Nonetheless, this semantic belief is even stronger than its syntactic counterpart, namely the weaker belief in the consistency of PA, of which we still lack a direct and immediate evidence. Anyway, it will be interesting to see what we can infer from this weaker syntactic assumption.

First Gödel Incompleteness Theorem. [Syntactic version] *Let T be a recursively axiomatizable arithmetical theory.*
(a) *If T is consistent then the Gödel's sentence $\varphi_g(g)$ is unprovable in T.*
(b) *If T is ω-consistent then neither the sentence $\neg\varphi_g(g)$ is provable in T.*
Thus the assumption of ω-consistency of T implies that T is incomplete.

Let us remark that ω-consistency is a technical condition, stronger than mere consistency, which we will formulate in the course of the demonstration of (b).

Demonstration. (a) Assume that the sentence $\varphi_g(g)$ is provable in T. From this point on we can proceed in two different ways. We will present both of them.

First, the provability of $\varphi_g(g)$ means that this sentence has some proof, say Δ_k, in T. Then $P(g, g, k)$ holds, and due to the algorithmic nature of the predicate $P(x, y, z)$, the statement $P(g, g, k)$ is provable in T. It follows that $(\exists\, z)P(g, g, z)$, which is equivalent to $\neg\varphi_g(g)$, is provable in T, as well. Thus we have both $T \vdash \varphi_g(g)$ and $T \vdash \neg\varphi_g(g)$, contradicting the consistency of T.

The second argument starts with realizing the form of $\varphi_g(g)$: in fact we have assumed that $T \vdash \neg(\exists\, z)P(g, g, z)$, hence $T \vdash (\forall\, z)\neg P(g, g, z)$, since the second sentence is equivalent to the first one. It follows that $T \vdash \neg P(g, g, k)$ for each $k \in \mathbb{N}$. Therefore, $\neg P(g, g, k)$ holds for each k, by Theorem… . It means that none of the proofs Δ_k is a proof of the sentence $\varphi_g(g)$ in T, in other words, $\varphi_g(g)$ is unprovable in T. This contradicts our original assumption which is henceforth wrong. Therefore $\varphi_g(g)$ is unprovable in T.

(b) Assume that the sentence $\neg\varphi_g(g)$, which is equivalent to $(\exists\, z)P(g, g, z)$, is provable in T. If there were some $k \in \mathbb{N}$ such that $P(g, g, k)$, we could infer that Δ_k is a proof of $\varphi_g(g)$ in T. Then both $\varphi_g(g)$ as well as $\neg\varphi_g(g)$ were provable in T, and we could refute our initial assumption that $T \vdash \neg\varphi_g(g)$ as contradicting the mere consistency of T (and get through without the assumption of its ω-consistency).

So does the provability of the arithmetical sentence $(\exists\, z)P(g, g, z)$ imply that there is indeed some $k \in \mathbb{N}$ such that $P(g, g, k)$? The positive answer to this question seems obvious at a glance. If there were a constructive proof of the statement $(\exists\, z)P(g,g,z)$, it would give us some concrete k such that $P(g,g,k)$. Unfortunately, we cannot exclude that the proof of the statement $(\exists\, z)P(g,g,z)$ proceeds in an indirect nonconstructive way, just deriving a contradiction from the assumption $\neg(\exists\, z)P(g,g,z)$, and giving not even a hint how the k such that $P(g,g,k)$ could be found. To conclude, our optimism was precocious, and our original idea of demonstration doesn't work. To get through we need something more.

An arithmetical theory T is called *ω-consistent* if, for no arithmetical formula $\psi(z)$, all the sentences $(\exists\, z)\psi(z)$, $\neg\psi(0)$, $\neg\psi(1)$, ... , $\neg\psi(k)$, ... are provable in T. Obviously, any ω-consistent arithmetical theory must be consistent.

Now assuming that T is ω-consistent and $T \vdash (\exists\, z)P(g,g,z)$, we can conclude that $T \nvdash \neg P(g,g,k)$ for some k. Then $P(g,g,k)$ holds for this k by Theorem... . From this point on the original argument can be applied.

The following Example illustrates the difference between a purely existential and a constructive proof of an existential statement.

Example. We will prove the theorem:

"There exist irrational numbers $a, b > 0$ such that the number a^b is rational."

Proof. It is known (and easy to show) that $\sqrt{2}$ is an irrational number. Then the number $\sqrt{2}^{\sqrt{2}}$ is either rational or irrational. If it is rational, we are done by taking $a = b = \sqrt{2}$. If $\sqrt{2}^{\sqrt{2}}$ is irrational, we put $a = \sqrt{2}^{\sqrt{2}}$ and $b = \sqrt{2}$. Then both a, b are irrational and

$$a^b = \left(\sqrt{2}^{\sqrt{2}}\right)^{\sqrt{2}} = \left(\sqrt{2}\right)^{\left(\sqrt{2}\,\sqrt{2}\right)} = \left(\sqrt{2}\right)^2 = 2$$

Since $a^b = 2$ is obviously rational, we are done again.

The reader should realize that our proof is purely existential, making use of the Law of Excluded Middle. We do not know whether the number $\sqrt{2}^{\sqrt{2}}$ is rational or irrational, therefore we do not know which one of the couple of possibilities really works. From the intuitionistic or constructivist viewpoint such proofs are unacceptable. A constructive proof would require to decide whether $\sqrt{2}^{\sqrt{2}}$ is rational or irrational and provide an explicit unambiguous choice of the pair a, b.

In fact it is known, but not so easy to show, that the number $\sqrt{2}^{\sqrt{2}}$ is irrational (hence the second possibility takes place in the proof above).

Later on B. Rosser formulated a modification of Gödel's statement $\neg(\exists\, z)P(g,g,z)$ making possible to avoid the assumption of ω-consistency and to prove the incompleteness of recursively axiomatizable arithmetic theories assuming their mere consistency.

Consider the arithmetical formula $(\forall\, z)(P(x,x,z) \Rightarrow (\exists\, u \le z)R(x,x,u))$. Let us denote by r its number in the list $\{\varphi_n(x)\}_{n\in\mathbb{N}}$. Substituting r for x into the formula $\varphi_r(x)$ we obtain the self-referential sentence $\varphi_r(r)$, i.e.,

$$(\forall\, z)(P(r,r,z) \Rightarrow (\exists\, u \le z)R(r,r,u))$$

Its meaning can be deciphered as follows:

$\varphi_r(r)$: *"If the sentence $\varphi_r(r)$ is provable in T by some proof of a given number then, among the proofs of at most that number, there is a proof of its negation $\neg\varphi_r(r)$ in T."*

Needless to say, the following version of the First Incompleteness Theorem is of syntactic nature.

Gödel-Rosser Incompleteness Theorem. *Let T be any recursively axiomatizable arithmetical theory. If T is consistent then neither the Rosser sentence $\varphi_r(r)$ nor its negation $\neg\varphi_r(r)$ are provable in T. Hence, if T is consistent then it is incomplete.*

Demonstration. Assume that $\varphi_r(r)$ is provable and Δ_k is its proof in T. Then both the statements $P(r,r,k)$ and $(\exists\, u \le k)R(r,r,u)$ are provable in T, as well. The latter is equivalent to the alternative

$$R(r,r,0) \vee R(r,r,1) \vee \ldots \vee R(r,r,k)$$

Then, however, it suffices to check the proofs Δ_0, Δ_1, ... , Δ_k and it is guaranteed that one from among them is a proof of the sentence $\neg\varphi_r(r)$ in T, contradicting its consistency.

Now, assume that $\neg\varphi_r(r)$ is provable in T by a proof Δ_l. Then we have $R(r,r,l)$, and the algorithm verifying this fact provides a proof of $R(r,r,l)$ in T. Realizing that $\neg\varphi_r(r)$ is equivalent to the sentence

$$(\exists\, z)\big(P(r,r,z) \wedge (\forall\, u)(R(r,r,u) \Rightarrow z < u)\big)$$

we can infer that the statement $(\exists\, z < l)P(r,r,z)$ is provable in T. Then necessarily $l > 0$, and the last statement is equivalent to the alternative

$$P(r,r,0) \vee P(r,r,1) \vee \ldots \vee P(r,r,l-1)$$

Hence among the proofs Δ_0, Δ_1, ... , Δ_{l-1}, there is a proof of the sentence $\varphi_r(r)$ in T, contradicting the consistency of T, again.

The Second Gödel Incompleteness Theorem

Informally, the Second Gödel Incompleteness Theorem states that any formal system which is sufficiently ample to include Peano Arithmetic cannot prove its own consistency. However, it should be realized that the statement that some formal system is consistent is a statement *about the system* which is not even formulated in the language of the system, thus the question of its provability within the system makes no sense. Hence it is a highly important fact that some formal systems, in particular, all recursively axiomatizable arithmetical first order theories, indeed allow for formulation of such statements.

Given an arithmetical theory T, we say that an arithmetical sentence θ is a *consistency statement* for T if the consistency of T is equivalent to the validity of θ in $(\mathbb{N};\, +, \cdot, 0, 1)$.

If T is additionally recursively axiomatizable then there are several possibilities how to formulate the consistency statement for T.

(1) Let us recall that a theory T in a first order language L is inconsistent if there is a sentence ψ in the language L such that both ψ and $\neg\psi$ are provable in T. Accordingly, the consistency statement can be formulated in following fairly suggestive way:

$$\mathrm{Cons}_1(T)\colon \neg(\exists\, x, y, z, w)(P(x, y, z) \wedge R(x, y, w))$$

excluding the existence of any sentence of the form $\varphi_n(m)$ such that both $\varphi_n(m)$, $\neg\varphi_m(m)$ were provable in T.

(2) Equivalently, T is inconsistent if and only if *every* L-sentence is provable in T. Thus T is consistent if and only if there is at least one L-sentence ψ not provable in T. We have a considerable freedom of choice for this sentence. In particular, we can follow the “way of economy” suggested by John von Neumann and take Gödel’s statement $\varphi_g(g)$ for that purpose. Indeed, if $\varphi_g(g)$ is provable in T then, as we already have seen, T is inconsistent. The other way round, if $\varphi_g(g)$ is unprovable in T then, of course, T is consistent. Thus T is consistent if and only if $\varphi_g(g)$ is not provable in T. This gives us the consistency statement

$$\mathrm{Cons}_2(T)\colon \neg(\exists\, z)P(g, g, z),$$

which coincides with the formerly introduced Gödel’s statement $\varphi_g(g)$.

(3) Last but not least, we can take some logical axiom or some axiom of PA; then the requirement of unprovability of its negation is clearly equivalent to the consistency of T. In particular, let $s \in \mathbb{N}$ be the number of the formula $x \neq x$. Then $\varphi_s(0)$ is the sentence $0 \neq 0$. That way we obtain yet another consistency statement:

$$\mathrm{Cons}_3(T)\colon \neg(\exists\, z)P(0, s, z)$$

expressing the unprovability of the sentence $0 \neq 0$ in T.

Exercise. (a) When dealing with the syntactic version of the First Gödel’s Incompleteness Theorem, we have shown that from the provability of Gödel’s sentence $\varphi_g(g)$ in T there follows the provability of its negation $\neg\varphi_g(g)$ in T. Taking for granted that the implication

$$(\exists\, z)P(g, g, z) \Rightarrow (\exists\, u)R(g, g, u)$$

formalizing that account is provable in T, show that the implication

$$\neg(\exists\, x, y, z, w)(P(x, y, z) \wedge R(x, y, w)) \Rightarrow \neg(\exists\, z)P(g, g, z)$$

is provable in T, as well. Therefore the provability of the consistency statement $\mathrm{Cons}_1(T)$ from (1) in T implies the same for Gödel’s sentence $\varphi_g(g)$.

(b) Similarly as in (a), we can infer that from the provability of Gödel’s sentence $\varphi_g(g)$ there follows the provability of the sentence $0 \neq 0$ in T. Take for granted that the implication

$$(\exists\, z)P(g, g, z) \Rightarrow (\exists\, u)P(0, s, u)$$

formalizing this account is provable in T and show that the provability of the consistency statement $\mathrm{Cons}_3(T)$ from (3) in T implies the same for Gödel's sentence $\varphi_g(g)$, again.

Thus for a recursively axiomatizable arithmetical theory T with the provability predicate $P(x,y,z)$ and, possibly, with the refutability predicate $R(z,y,z)$, the consistency statement $\mathrm{Cons}(T)$ can be formulated within its language by any of the three sentences $\mathrm{Cons}_1(T)$, $\mathrm{Cons}_2(T)$, $\mathrm{Cons}_3(T)$ mentioned above (as well as by many more ones). At the same time, the assumption of provability of any of these statements in T yields the provability of Gödel's sentence $\varphi_g(g)$ in T. Summing up we have:

Second Gödel Incompleteness Theorem. *Let T be a recursively axiomatizable arithmetical theory. Then T allows for the formulation of its own consistency statement* $\mathrm{Cons}(T)$*. However, if T is consistent then any of the consistency statements* $\mathrm{Cons}_1(T)$*,* $\mathrm{Cons}_2(T)$*,* $\mathrm{Cons}_3(T)$ *from the above list is unprovable in T.*

If T is a consistent recursively axiomatizable arithmetical theory then, by Gödel's Second Incompleteness Theorem, the statement $\mathrm{Cons}(T)$ is not provable in T, thus, due to the Theorem on Proof by Contradiction, its extension $T \cup \{\neg\,\mathrm{Cons}(T)\}$ is consistent, as well. However, as T is consistent, the axiom $\neg\,\mathrm{Cons}(T)$ is not satisfied in $(\mathbb{N};\, +, \cdot, 0, 1)$, hence $T \cup \{\neg\,\mathrm{Cons}(T)\}$ cannot be arithmetically correct even if T is. Next we denote, for definiteness' sake, by $\mathrm{Cons}(T)$ the Gödel's statement $\varphi_g(g)$. Then the statement $(\exists\, z)P(g,g,z)$, being logically equivalent to $\neg\,\mathrm{Cons}(T)$, is provable in $T \cup \{\neg\,\mathrm{Cons}(T)\}$. However, since T is consistent, none of the proofs Δ_k is a proof of the sentence $\varphi_g(g)$, i.e., of $\mathrm{Cons}(T)$, in T, therefore all the statements $\neg P(g,g,k)$, for $k \in \mathbb{N}$, are true in $(\mathbb{N};\, +, \cdot, 0, 1)$, hence provable in T and the more in $T \cup \{\neg\,\mathrm{Cons}(T)\}$. That way $T \cup \{\neg\,\mathrm{Cons}(T)\}$ is an example of a consistent theory which is not ω-consistent. On the other hand, if T is arithmetically correct then so is $T \cup \{\mathrm{Cons}(T)\}$.

In the following two exercises T denotes a recursively axiomatizable arithmetical theory with the provability predicate $P(x,y,z)$ and refutability predicate $R(x,y,z)$.

Exercise. The initial account in (2) suggests the following formalization of the consistency statement for T:

$$\mathrm{Cons}_4(T)\colon\ (\exists\, x,y)(\forall\, z)\neg P(x,y,z)$$

declaring the existence of some sentence $\varphi_n(m)$ unprovable in T. However, the fact that its syntactic complexity (due to the quantifier prefix $\exists\forall$) is one step higher than that of the previous three consistency statements causes that it is not so easy to derive any conclusions from the assumption of the provability of $\mathrm{Cons}_4(T)$ in T.

(a) Show that $\mathrm{Cons}_4(T)$ is a consistency statement for T.

(b) Examine the provability status of the consistency statement $\mathrm{Cons}_4(T)$ in T. Realize that the mere assumption that T is consistent still does not allow us to show neither that $\mathrm{Cons}_4(T)$ is provable nor that it is unprovable in T. Observe that the implication $\varphi_g(g) \Rightarrow \mathrm{Cons}_4(T)$ is logically valid. Next, show that if T is ω-consistent then the negation $\neg\,\mathrm{Cons}_4(T)$ is not provable in T.

Exercise. (a) Show that the Rosser formula $\varphi_r(r)$ is not a consistency statement for T. What about its negation $\neg\varphi_r(r)$?

(b) Show that both the sentences $\neg(\exists\, z)P(r,r,z), \neg(\exists\, u)R(r,r,u)$ are consistency statements for T. What is their provability status?

In view of Gödel's results it is perhaps surprising but anyway worthwhile to mention that S. Feferman in 1960, at the cost of higher complexity, constructed a consistency statement $\mathrm{Cons}^*(\mathrm{PA})$ for Peano Arithmetic which, nevertheless, is *provable* in PA. However, for such a consistency statement neither the equivalence $\mathrm{Cons}^*(\mathrm{PA}) \Leftrightarrow \mathrm{Cons}_i(\mathrm{PA})$ nor even the implication $\mathrm{Cons}^*(\mathrm{PA}) \Rightarrow \mathrm{Cons}_i(\mathrm{PA})$, for any $i = 1, 2, 3$, is provable in PA (unless PA is inconsistent).

Attempts at Completion

There naturally arises the question whether Peano Arithmetic cannot be completed by adding to it some new axioms of which we know that they are satisfied in the standard model $(\mathbb{N}; +, \cdot, 0, 1)$. One possible candidate could be recursively constructed as follows: Let T_0 be the theory PA itself. Given the theory T_q, for $q \in \mathbb{N}$, we construct the sequence $\Delta_0^q, \Delta_1^q, \dots, \Delta_k^q, \dots$ of all proofs in T_q and the provability predicate $P_q(x,y,z)$ for T_q such that, for any $k, m, n \in \mathbb{N}$,

$$P_q(m,n,k) \quad \text{if and only if} \quad \Delta_k^q \text{ is a proof of the sentence } \varphi_n(m) \text{ in } T_q$$

Then we put

$$T_{q+1} = T_q \cup \{\mathrm{Cons}(T_q)\}$$

where $\mathrm{Cons}(T_q)$ is any of the consistency statements $\mathrm{Cons}_i(T_q)$ for fixed $i = 1, \dots, 3$. In other words, T_{q+1} is the extension of T_q by the consistency axiom $\mathrm{Cons}(T_q)$ for T_q. Obviously, every T_q is a recursively axiomatizable arithmetically correct theory. Thus putting

$$\widehat{T} = \bigcup_{q \in \mathbb{N}} T_q$$

we get an arithmetically correct theory in which all the consistency statements $\mathrm{Cons}(T_q)$ can be proved. However, $\widehat{T}$ is still recursively axiomatizable, hence all the previous incompleteness results apply to it. In particular, $\widehat{T}$ is incomplete, it can formulate its own consistency statement $\mathrm{Cons}\big(\widehat{T}\,\big)$ which, nevertheless, it is incapable to prove. Moreover, as shown by Alan Turing, Peano Arithmetic cannot be completed even by transfinite iteration of the procedure of extending it by adding consecutive consistency statements to it.

One of the aspects of the incompleteness of PA and related theories can be more specifically identified as the phenomenon of *ω-incompleteness*, i.e., a kind of "nonuniformity" of provability in them. For instance, if $\psi(x)$ is a formula in the language of PA then the provability in PA of all the sentences $\psi(m)$ for any $m \in \mathbb{N}$ still does not imply the provability of its universal closure $(\forall\, x)\psi(x)$ in PA. It can namely happen that the particular proofs of the individual instances $\psi(0)$, $\psi(1)$, ... , $\psi(m)$, ... differ to such an extent that it is impossible to compose a uniform proof of the universally quantified statement $(\forall x)\psi(x)$ out of them. An arithmetical theory T is called *ω-complete* if this cannot happen, i.e., if, for every arithmetical formula $\psi(x)$, the provability in T of all the particular instances $\psi(m)$ for all $m \in \mathbb{N}$ already implies the provability of its universal closure $(\forall x)\psi(x)$ in T.

A construction of a complete extension of PA based on the removal of the ω-incompleteness phenomenon was proposed by S. Feferman in 1962. However, in order to extend PA to both a complete and ω-complete theory he had to sacrifice the condition of recursive axiomatization. By transfinite recursion over the ordinal numbers less than certain limit ordinal $\zeta \le \omega^{\omega^\omega}$ he constructed a sequence of arithmetical theories $\{T_\alpha\}_{\alpha<\zeta}$ and a sequence of provability predicates $\big\{P_\alpha(x,y,z)\big\}_{\alpha<\zeta}$ for these theories such that

$$\begin{aligned} T_0 &= \mathrm{PA} \\ T_{\alpha+1} &= T_\alpha \cup \big\{(\forall x)(\exists z)P_\alpha(x,n,z) \Rightarrow (\forall x)\varphi_n(x) \colon n \in \mathbb{N}\big\} \quad \text{for each } \alpha < \zeta \\ T_\lambda &= \bigcup_{\alpha<\lambda} T_\alpha \quad \text{for any limit ordinal } \lambda < \zeta \end{aligned}$$

Adding the new axioms

$$(\forall x)(\exists z)P_\alpha(x,n,z) \Rightarrow (\forall x)\varphi_n(x)$$

for $n \in \mathbb{N}$ to the axioms of T_α guarantees the provability of every universally quantified statement $(\forall x)\varphi_n(x)$ in $T_{\alpha+1}$, once all its particular instances $\varphi_n(m)$ for $m \in \mathbb{N}$ are provable in T_α.

Finally, it can be shown that the arithmetical theory

$$\widetilde{T} = \bigcup_{\alpha<\zeta} T_\alpha$$

is not only ω-complete but also complete and ω-consistent. However, in order to derive at this conclusion we have to assume that PA is consistent. Assuming that PA is arithmetically correct, we can infer that so is $\widetilde{T}$.

The Theorems of Tarski and Church-Turing

To complete the picture we formulate two further incompleteness results by A. Tarski, and A. Church and A. Turing, respectively. *Tarski's Theorem on Undefinability of Truth* states informally that the property of arithmetical sentences "to be true" cannot be defined by any formula in the language of those sentences. More precisely, it says that the *satisfaction relation* for arithmetical formulas in the standard model $(\mathbb{N}; +, \cdot, 0, 1)$ cannot be expressed by any arithmetical formula. In the theorems below we once again refer to the sequence $\{\varphi_n(x)\}_{n\in\mathbb{N}}$ of arithmetical formulas.

Tarski's Theorem on Undefinability of Truth. *Let T be any arithmetical first order theory. Then there is no arithmetical formula $\sigma(x, y)$ in the language of T such that for any $m, n \in \mathbb{N}$ we have*

$$(\mathbb{N}; +, \cdot, 0, 1) \vDash \varphi_n(m) \Leftrightarrow \sigma(m, n)$$

Demonstration. Admit that such a formula $\sigma(x, y)$ exists. Then $\neg\sigma(x, x)$ is an arithmetical formula with a single free variable x, thus it can be found in the sequence $\{\varphi_n(x)\}_{n\in\mathbb{N}}$ under some index $t \in \mathbb{N}$. Then the following statements are equivalent in $(\mathbb{N}; +, \cdot, 0, 1)$: $\sigma(t, t)$, $\varphi_t(t)$, $\neg\sigma(t, t)$. Hence

$$(\mathbb{N}; +, \cdot, 0, 1) \vDash \sigma(t, t) \Leftrightarrow \neg\sigma(t, t)$$

which is contradiction.

Tarski's Theorem, which is of semantic nature, imposes severe limitations on the possibility of self-representation of arithmetical theories. In order to be able to define a satisfaction formula $\sigma(x, y)$ for T it is necessary to extend T to a first order theory T' in a "metalanguage" whose expressive power goes beyond that of T. For example, a satisfaction formula for Peano Arithmetic can be defined in the Second Order Arithmetic or in the Zermelo-Fraenkel Set Theory.

Dealing with decidability questions both A. Church and A. Turing were heavily influenced by the work of K. Gödel on completeness of the First Order Logic and even more by his work on incompleteness of Peano Arithmetic and related theories. While Church developed the so called λ-*calculus* and used it as a paradigmatic model of general computations, Turing designed ideal models of computing devices which became known as *Turing machines*. Soon it became clear that both approaches are equivalent. The proof of their Undecidability Theorem is beyond the scope of our course.

Church-Turing Undecidability Theorem. *Let T be any consistent recursively axiomatizable arithmetical theory. Then there is no algorithm which could decide whether any given arithmetical sentence in the language of T is provable in T. In particular, there is no algorithm which could decide the question of provability in T of the sentence $\varphi_n(m)$ for every input $(m, n) \in \mathbb{N} \times \mathbb{N}$.*

Church also proved that there is no algorithm which could decide whether a sentence in a first order language L with at least one binary relational symbol or at least two operation symbols is a "first order tautology", i.e., whether it is satisfied in all L-structures, (or, which is the same, whether is provable just from the logical axioms). Thus there is a striking difference between the First Order Logic and the Propositional Calculus in which the tautologies can be effectively recognized by the truth table algorithm.

Compared with Tarski's Theorem, Church-Turing Theorem is of syntactic character. While Tarski's Theorem imposes some limits to what can be *expressed* by formal languages, Church-Turing Theorem sets up some limits to what can be *computed* by any mechanical or electronic device or *effectively decided* by means of an algorithmic computational procedure. However, they both, together with Gödel's Incompleteness Theorems, of course, raise various questions about the relation of computers, human brains and human mind or spirit.

Goodstein Sequences:

An Example of a True Arithmetical Statement Unprovable in PA

It can be objected that the true statements unprovable in PA constructed by Gödel and Rosser, like $\varphi_g(g)$, $\varphi_r(r)$, Cons(PA) (no matter which possibility we choose), are highly artificial and deprived of proper mathematical meaning and content. However, there are indeed several known arithmetical theorems of combinatorial or number theoretic character which, nonetheless, are unprovable in PA. As a rule, they illustrate the ω-incompleteness phenomenon at the same time. Some examples of universally quantified statements of the form $(\forall\, x)\psi(x)$ unprovable in PA, nonetheless true in $(\mathbb{N}; +, \cdot, 0, 1)$ in the sense that all the particular instances $\psi(m)$ for each $m \in \mathbb{N}$ are even provable in PA, are provided by the *Paris-Harrington* strengthening of *Ramsey's Theorem* or by the *Goodstein sequences*. Both these results are in fact equivalent to the consistency of Peano Arithmetic. We will briefly explain the nature of the latter example.

Given a natural number $b \geq 2$, the *hereditary base b expansion* of any natural number m is obtained from the its usual base b expansion by expanding all its exponents at the base b, again, doing the same with the exponents of exponents, and repeating this procedure until all the numbers bigger than b are eliminated from this expression. For instance, the hereditary base 2 expansion of the number $m = 357$ reads as follows:

$$\begin{aligned} 357 &= 2^8 + 2^6 + 2^5 + 2^2 + 1 = 2^{2^3} + 2^{2^2+2} + 2^{2^2+1} + 1 \\ &= 2^{2^{2+1}} + 2^{2^2+2} + 2^{2^{2+1}} + 2^{2^2+1} + 1 \end{aligned}$$

Its hereditary base 3 expansion is

$$357 = 3^5 + 3^4 + 3^3 + 2 \cdot 3 = 3^{3+2} + 3^{3+1} + 3^3 + 2 \cdot 3$$

Similarly, the hereditary base 2 expansion of the number $m = 1\,000$ is

$$\begin{aligned} 1\,000 &= 2^9 + 2^8 + 2^7 + 2^6 + 2^5 + 2^3 \\ &= 2^{2^3+1} + 2^{2^3} + 2^{2^2+2+1} + 2^{2^2+2} + 2^{2^2+1} + 2^{2+1} \\ &= 2^{2^{2+1}+1} + 2^{2^{2+1}} + 2^{2^2+2+1} + 2^{2^2+2} + 2^{2^2+1} + 2^{2+1} \end{aligned}$$

On the other hand, its hereditary base 5 expansion coincides with its plane base 5 expansion:

$$1\,000 = 5^4 + 3 \cdot 5^3$$

For any natural number m we construct the *Goodstein sequence* of natural numbers

$$G(m,0), G(m,1), G(m,2), \ldots, G(m,n), G(m,n+1), \ldots$$

corresponding to m, which starts with $G(m,0) = m$, and having arrived at the number $G(m,n)$, if $G(m,n) > 0$ then the next item $G(m,n+1)$ is obtained by replacing every occurrence of the number $n+2$ in the hereditary base $n+2$ expansion of $G(m,n)$ by the number $n+3$ and subtracting 1 from the result; if $G(m,n) = 0$ then $G(m,n+1) = 0$, as well. For example, for $m = 29$, we get

$$\begin{aligned}
G(m,0) &= 2^4 + 2^3 + 2^2 + 1 = 2^{2^2} + 2^{2+1} + 2^2 + 1 \\
G(m,1) &= 3^{3^3} + 3^{3+1} + 3^3 + 1 - 1 = 3^{3^3} + 3^{3+1} + 3^3 = 7\,625\,597\,485\,095 \\
G(m,2) &= 4^{4^4} + 4^{4+1} + 4^4 - 1 = 4^{4^4} + 4^{4+1} + 3 \cdot 4^3 + 3 \cdot 4^2 + 3 \cdot 4 + 3 \approx 1.340 \cdot 10^{154} \\
G(m,3) &= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 3 - 1 \\
&= 5^{5^5} + 5^{5+1} + 3 \cdot 5^3 + 3 \cdot 5^2 + 3 \cdot 5 + 2 \sim 10^{2\,200} \\
G(m,4) &= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 2 - 1 \\
&= 6^{6^6} + 6^{6+1} + 3 \cdot 6^3 + 3 \cdot 6^2 + 3 \cdot 6 + 1 \sim 10^{36\,305} \\
G(m,5) &= 7^{7^7} + 7^{7+1} + 3 \cdot 7^3 + 3 \cdot 7^2 + 3 \cdot 7 + 1 - 1 \\
&= 7^{7^7} + 7^{7+1} + 3 \cdot 7^3 + 3 \cdot 7^2 + 3 \cdot 7 \sim 10^{696\,000} \\
&\ldots\ldots
\end{aligned}$$

The above computations indicate that the Goodstein sequences $\{G(m,n)\}_{n=0}^{\infty}$ grow rapidly for any m, and not just for the particular value $m = 29$. Thus the following result is highly surprising and unexpected.

Goodstein's Theorem [1944]**.** *For every natural number m there exists a natural number n such that $G(m,n) = 0$.*

In fact, for $m \leq 3$, the sequence $\{G(m,n)\}_{n=0}^{\infty}$ assumes the value 0 fairly quickly. The reader can easily verify that $G(0,n) = 0$, for each n, $G(1,0) = 1$, $G(1,n) = 0$ for $n \geq 1$, $G(2,0) = G(2,1) = 2$, $G(2,2) = 1$ and $G(2,n) = 0$ for $n \geq 3$. For $m = 3$ we have

$$\begin{array}{lll}
G(3,0) = 3 = 2 + 1 & G(3,1) = 3 + 1 - 1 = 3 & G(3,2) = 4 - 1 = 3 \\
G(3,3) = 3 - 1 = 2 & G(3,4) = 2 - 1 = 1 & G(3,5) = 0 = G(3,n) \text{ for } n > 5
\end{array}$$

For $m = 4$ the first n such that $G(4,n) = 0$ equals the immense value $3 \cdot \left(2^{402\,653\,211} - 1\right)$.

Formally, the proof of Goodstein's Theorem uses transfinite induction over the countable well-ordered set of all ordinal numbers less than the ordinal

$$\varepsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}$$

i.e., the first ordinal α satisfying $\omega^\alpha = \alpha$. However, the main idea of this proof can be explained easily. It consists in dominating every sequence $\{G(m,n)\}_{n=0}^\infty$, with m fixed, by a sequence $\{\Gamma(m,n)\}_{n=0}^\infty$ of ordinal numbers $\Gamma(m,n) < \varepsilon_0$ such that $G(m,n) \le \Gamma(m,n)$ and $\Gamma(m,n) > \Gamma(m,n+1)$ whenever $\Gamma(m,n) > 0$, for each n. Since the set of all ordinals $< \varepsilon_0$ is well-ordered by the relation $<$, it cannot contain any infinite strictly decreasing sequence. Hence each of the sequences $\{\Gamma(m,n)\}_{n=0}^\infty$ must eventually stabilize at the value $\Gamma(m,n) = 0$ for some n. Then $G(m,n) = 0$, as well.

The ordinal number $\Gamma(m,n)$ is obtained by replacing each occurrence of the term $n+2$ in the hereditary base $n+2$ expansion of the number $G(m,n)$ by the ordinal ω. In the particular case $m = 29$ we have

$$G(m,0) = 2^{2^2} + 2^{2+1} + 2^2 + 1 < \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega + 1 = \Gamma(m,0)$$

$$G(m,1) = 3^{3^3} + 3^{3+1} + 3^3 < \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega = \Gamma(m,1)$$

$$G(m,2) = 4^{4^4} + 4^{4+1} + 3\cdot 4^3 + 3\cdot 4^2 + 3\cdot 4 + 3 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 3 = \Gamma(m,2)$$

$$G(m,3) = 5^{5^5} + 5^{5+1} + 3\cdot 5^3 + 3\cdot 5^2 + 3\cdot 5 + 2 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 2 = \Gamma(m,3)$$

$$G(m,4) = 6^{6^6} + 6^{6+1} + 3\cdot 6^3 + 3\cdot 6^2 + 3\cdot 6 + 1 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 1 = \Gamma(m,4)$$

$$G(m,5) = 7^{7^7} + 7^{7+1} + 3\cdot 7^3 + 3\cdot 7^2 + 3\cdot 7 < \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega = \Gamma(m,5)$$

$$\dots\dots$$

Then the sequence of ordinals

$$\begin{aligned}
&\Gamma(m,0) = \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega + 1\\
&\quad > \Gamma(m,1) = \omega^{\omega^\omega} + \omega^{\omega+1} + \omega^\omega\\
&\qquad > \Gamma(m,2) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 3\\
&\qquad\quad > \Gamma(m,3) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 2\\
&\qquad\qquad > \Gamma(m,4) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega + 1\\
&\qquad\qquad\quad > \Gamma(m,5) = \omega^{\omega^\omega} + \omega^{\omega+1} + 3\cdot\omega^3 + 3\cdot\omega^2 + 3\cdot\omega\\
&\qquad\qquad\qquad > \dots
\end{aligned}$$

cannot decrease for ever, hence it must eventually stabilize at the value $\Gamma(m,n) = 0$ for some unimaginably huge value of n. For that n also $G(m,n) = 0$.

As shown by J. Paris and L. Kirby, Goodstein's Theorem cannot be proved just by means of the Peano Arithmetic alone.

Paris-Kirby Theorem [1982]**.** *In* PA *it is provable that Goodstein's Theorem implies the consistency statement* Cons(PA)*. As a consequence, if* PA *is consistent then Goodstein's Theorem is not provable in* PA.

On the other hand, for any fixed $m \in \mathbb{N}$, the existential statement $(\exists\, y)(G(m,y) = 0)$ is provable in PA. We know this though already for rather small values of m we not only do not know the precise value of such a $y = n$ but we even do not dispose of any explicit proof of that statement in PA. We only know that the primitive step-by-step computation must eventually produce the result. However, this computation will not terminate within the existence not only of the mankind but of the entire universe. At the same time, as an illustration of the ω-incompleteness phenomenon mentioned in connection with Feferman's construction, it should be realized that within PA it is impossible to extract any general common idea out of those particular proofs and convert them into a single proof of the universal-existential sentence $(\forall x)(\exists y)(G(x, y) = 0)$.

Philosophical Consequences

Themes for an Essay

There is a vast literature dealing with mathematical, philosophical, metaphysical and others extra-mathematical consequences of Gödel's Incompleteness Theorems and some related results. Let us confine to a brief list of some traditionally inferred conclusions:

(1) Human knowledge is necessarily incomplete and we never can be sure that it is free of contradictions.

(2) Human knowledge cannot be reduced to any formal system. By realizing the incompleteness phenomena inherent for such systems we are capable to transcend their limitations.

(3) Computers can compute and prove just within the scope of some formal system. Humans, however, are able to seize and reveal some truths unprovable within any formal system. It follows that human brain — in spite of the fact that with respect to some parameters (as, e.g., the speed of computation) it is far behind the computers — still possesses some capabilities making it superior to any computer.

It is extremely interesting to present some Gödel's ideas upon these issues here. Gödel namely went a step farther beyond (3). According to him, we all probably agree that computers can compute and prove just within the scope of some formal system given in advance. Similarly, the activity of human brain can in principle be simulated by certain computer (though we do not dispose of such computers at present). However, human beings are capable of viewing or grasping even some truths unprovable within any formal system. It follows that human mind or human intellect or human spirit, however we call it, is endowed not only with some capabilities which make it superior to any computer but also with some faculties which cannot be explained as a mere manifestation of the activity and functioning of human brain.[1]

Try to ponder over the above quoted conclusions and opinions. To which degree you agree or disagree with any of them and why? To which degree can the above conclusions be justified by the incompleteness results we have been dealing with? Discuss those points and try to make them more precise, finally arriving at some formulations you can agree with. To which degree follow your conclusions from the results of Gödel, Rosser, Tarski, Church and Turing?

[1] Freely quoted according to Hao Wang [9].

Suggestions for Further Reading

[1] Martin Davis, *Pragmatic Platonism, Mathematics and the Infinite*; https://www.researchgate.net/publication/329449494

[2] Martin Davis, *What Did Gödel Believe and When Did He Believe It?* Bull. Symbolic Logic **11** (2005), 194-206; https://www.jstor.org/stable/1556749?seq=1

[3] Solomon Feferman, *The Impact of the Incompleteness Theorems on Mathematics*, Notices Amer. Math. Soc. **53** (April 2006), 434–439; https://www.ams.org/notices/200604/fea-feferman.pdf

[4] Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse*, A. K. Peters, Wellesley, 2005.

[5] Haim Gaifman, *What Gödel's Incompleteness Result Does and Does Not Show*; https://pdfs.semanticscholar.org/1efb/5896951d8c20984f558c1e4ab4d6d029143d.pdf

[6] Panu Raatikainen, *Gödel's Incompleteness Theorems*, Stanford Encyclopedia of Philosophy; https://plato.stanford.edu/entries/goedel-incompleteness/

[7] Peter Smith, *An Introduction to Gödel's Theorems*, Cambridge University Press, Cambridge, 2007.

[8] Raymond Smullyan, *Gödel's Incompleteness Theorems*, Oxford University Press, Oxford, 1992.

[9] Hao Wang, *A Logical Journey, From Gödel to Philosophy*, MIT Press, Cambridge, Massachusetts–London, England, 1996.

[10] Richard Zach, *Hilbert's Program*, Stanford Encyclopedia of Philosophy; http://plato.stanford.edu/archives/fall2003/entries/hilbert-program/(2003)

[11] Pavol Zlatoš, *Ani matematika si nemôže byť istá sama sebou, Úvahy o množinách, nekonečne, paradoxoch a Gödelových vetách*, IRIS, Bratislava, 1995; http://thales.doa.fmph.uniba.sk/zlatos/animat/animat.pdf