An Informal Exposition of Proofs of Godel's Theorems and Church's Theorem

Barkley Rosser

The Journal of Symbolic Logic, Vol. 4, No. 2. (Jun., 1939), pp. 53-60.

Stable URL:
http://links.jstor.org/sici?sici=0022-4812%28193906%294%3A2%3C53%3AAIEOPO%3E2.0.CO%3B2-3

The Journal of Symbolic Logic is currently published by Association for Symbolic Logic.

The Journal of Symbolic Logic
Volume 4, Number 2, June 1939

AN INFORMAL EXPOSITION OF PROOFS OF GÖDEL'S THEOREMS AND CHURCH'S THEOREM

BARKLEY ROSSER

This paper is an attempt to explain as non-technically as possible the principles and devices used in the various proofs of Gödel's Theorems and Church's Theorem.

Roman numerals in references shall refer to the papers in the bibliography.

In the statements of Gödel's Theorems and Church's Theorem, we will employ the phrase "for suitable L." The hidden assumptions which we denote by this phrase have never been put down explicitly in a form intelligible to the average reader.[1] The necessity for thus formulating them has commonly been avoided by proving the theorems for special logics and then remarking that the proofs can be extended to other logics. Hence the conditions necessary for the proofs of Gödel's Theorems and Church's Theorem are at present very indefinite as far as the average reader is concerned. To partly clarify this situation, we will now mention the more prominent of these assumptions.

I. In any proof of Gödel's Theorems or Church's Theorem, two logics are concerned. One serves as the "logic of ordinary discourse" in which the proof is carried out, and the other is a formal logic, L, about which the theorem is proved. The first logic may or may not be formal. However L must be formal. Among other things, this implies that the propositions of L are formulas built according to certain rules of structure. Each formula is to consist of a finite number (counting repetitions) of symbols chosen out of a set (finite or denumerably infinite) which is given at the start; any symbol of the set may be used more than once in any formula. Moreover the symbols have meanings attached, in terms of which propositions of L may be interpreted. The rules of structure of the propositions of L are supposed to be such that the interpretations of the propositions of L will be declarative sentences (not necessarily true) of "ordinary discourse." If A is a proposition of L, and a certain sentence is the interpretation of A, then A is said to be the "expression in L" of that sentence or any sentence equivalent to it. In general, not all sentences can be expressed in L.[2]

Received January 12, 1939.

[1] It is my understanding from conversations with Gödel that an exact formulation of these assumptions was to constitute part of the second part of the paper of which II is the first part. Due to ill health, Gödel has never written this second half. However, in III, Kleene gives an exact statement of a set of assumptions sufficient for his proof of Gödel's First Theorem. Unfortunately they are phrased in terms of general recursive functions, and are illuminating only to someone who is thoroughly familiar with the theory of general recursive functions.

[2] In the L's which receive general attention, no method is apparent of expressing such sentences as "Plato was mortal," "God is good," etc.

II. Amongst the symbols of L must be one, $\sim$, which is interpreted as "not." That is, if A expresses in L a certain sentence, then $\sim A$ expresses in L the contradictory of that sentence.

III. For each positive integer, there must be a particular formula in L which denotes that integer. Also, amongst the symbols of L must be some, called variables, whose mode of interpretation is as follows. If a formula A of L expresses a sentence S and if A contains symbols called variables, $v_1, v_2, \cdots, v_s$, then S contains variables.[3] Moreover, if B is the formula got from A by replacing various of the v_i's of A by other symbols, then the sentence which B expresses is got from S by making corresponding replacements for the variables of S. In particular, if the formula G of L with the symbol v, called a variable, expresses in L the sentence "x has the property Q," with the variable x corresponding to v, and if F is got from G by replacing all the v's of G by the formula denoting the number n, then F expresses in L the sentence "n has the property Q."

IV. Also there must be a process whereby certain of the propositions of L are specified as "provable." The definition of "provable" is always supposed to be made without referring to the meanings of the formulas. However it was always hoped that the set of provable propositions of L would coincide with the set of propositions of L which express true sentences. Gödel's Theorems tell us that such cannot be the case. For Gödel's First Theorem states:

For suitable L, there are undecidable propositions in L; that is, propositions F such that neither F nor $\sim F$ is provable.

As F and $\sim F$ express contradictory sentences, one of them must express a true sentence. So there will be a proposition of L which expresses a true sentence, but nevertheless is not provable. This still leaves open the possibility that all provable propositions of L may express true sentences. As the notion of "truth of a sentence" is vague, it is usual to deal with weaker but more precise notions. For instance, L is said to be "simply consistent" if there is no proposition F such that both F and $\sim F$ are provable. Clearly, if L is not simply consistent, then some provable proposition of L must express a false sentence. However, some provable propositions of L may express false sentences even if L is simply consistent. Tarski[4] showed this by constructing a logic L which was simply consistent but in which one could prove the propositions expressing each sentence of the following infinite set (with Q properly chosen):

Not all positive integers have property Q.
1 has property Q.
2 has property Q.
3 has property Q.
. . . .

[3] I am purposely overlooking the complications due to the use of "apparent variables" as being irrelevant to the present discussion.

[4] Alfred Tarski, *Einige Betrachtungen über die Begriffe der ω-Widerspruchsfreiheit und der ω-Vollständigkeit*, ***Monatshefte für Mathematik und Physik***, vol. 40 (1933), pp. 97–112.

A logic L in which this latter situation does not occur for any property Q is said to be ω-consistent.

V. If F and $\sim F$ are both provable in L, then all propositions of L are provable. So if L is not simply consistent, it is not ω-consistent. So ω-consistency implies simple consistency. In fact, the non-provability of any formula whatever of L implies the simple consistency of L.

VI. There is a symbol, $\supset$, of L such that if the formula A expresses the sentence S and the formula B expresses the sentence T, then $A \supset B$ expresses the sentence "If S, then T." Also the definition of "provable" shall be such that if A and $A \supset B$ are provable then so is B.

This completes our list. The list was compiled for expository purposes only. Hence the list suffers the double defect of not containing absolutely all necessary assumptions, and of containing some assumptions which may not be necessary. Also the assumptions are not always stated with strict accuracy, on the ground that readers who know of cases not covered by our simplified versions of the assumptions will know the corrections that need to be made, and that readers who do not know of such cases will not fall into error thereby.

Three proofs of Gödel's First Theorem (see above) will be considered in this paper, namely Gödel's proof (II, Satz VI), Rosser's proof (IV, Thm. II), and Kleene's proof (III, Thm. XIII). These proofs will be referred to as GG_1, RG_1, and KG_1 respectively. All three use the general assumptions listed above. In addition, GG_1 assumes that L is ω-consistent, RG_1 assumes that L is simply consistent, and KG_1 assumes a more complicated type of consistency, roughly equivalent to ω-consistency.

Gödel's Second Theorem states:

For suitable L, the simple consistency of L cannot be proved in L.

Gödel proves this statement (II Satz XI) with the special assumption of simple consistency. His proof will be referred to as GG_2.

Church's Theorem states:

For suitable L, there exists no effective method of deciding which propositions of L are provable.

The statement is proved by Church (I, last paragraph) with the special assumption of ω-consistency, and by Rosser (IV, Thm. III) with the special assumption of simple consistency. These proofs will be referred to as CC and RC respectively.

Clearly the existence of CC or RC presupposes a precise definition of "effective." "Effective method" is here used in the rather special sense of a method each step of which is precisely predetermined and which is certain to produce the answer in a finite number of steps. With this special meaning, three different precise definitions have been given to date.[5] The simplest of these to state

[5] One definition is given by Church in I. Another definition is due to Jacques Herbrand and Kurt Gödel. It is stated in I, footnote 3, p. 346. The third definition was given independently in two slightly different forms by E. L. Post, *Finite combinatory processes—*

(due to Post and Turing) says essentially that an effective method of solving a certain set of problems exists if one can build a machine which will then solve any problem of the set with no human intervention beyond inserting the question and (later) reading the answer. All three definitions are equivalent, so it does not matter which one is used. Moreover, the fact that all three are equivalent is a very strong argument for the correctness of any one.

All the proofs GG_1, KG_1, RG_1, GG_2, CC, and RC use Gödel's device, which we now describe, for numbering formulas. First assign numbers to the symbols of L in any way that seems suitable. For instance Gödel discusses a logic involving the symbols

$$\sim, \mathbf{v}, \Pi, 0, f, (,),$$

and an infinite set of variables in each of an infinite set of types. He assigns numbers to these symbols as follows: 1 to 0, 3 to f, 5 to $\sim$, 7 to $\mathbf{v}$, 9 to Π, 11 to (, 13 to), and p_i^n (where the p_i's are primes greater than 13) to variables of type n.

Having assigned numbers to symbols, we next assign numbers to formulas as follows. Let $n_1, n_2, \cdots, n_s$ be the numbers of the symbols of a formula F in the order in which they occur in F. Let $p_1, p_2, \cdots, p_s$ be the first s primes in order of increasing magnitude (counting 2 as the first prime). Then the number assigned to F will be $p_1^{n_1} \cdot p_2^{n_2} \cdot \cdots \cdot p_s^{n_s}$. For example, one of the provable formulas of the logic which Gödel used is

$$\sim(x\Pi((\sim(x(fy)))\mathbf{v}(x(0)))),$$

(x and y being variables of types 2 and 1 respectively). The numbers of the symbols of this formula are successively 5, 11, 289, 9, 11, 11, 5, 11, 289, 11, 3, 17, 13, 13, 13, 7, 11, 289, 11, 1, 13, 13, 13, 13. So the number of the formula itself is $2^5 \cdot 3^{11} \cdot 5^{289} \cdot 7^9 \cdot 11^{11} \cdot 13^{11} \cdot 17^5 \cdot 19^{11} \cdot 23^{289} \cdot 29^{11} \cdot 31^3 \cdot 37^{17} \cdot 41^{13} \cdot 43^{13} \cdot 47^{13} \cdot 53^7 \cdot 59^{11} \cdot 61^{289} \cdot 67^{11} \cdot 71 \cdot 73^{13} \cdot 79^{13} \cdot 83^{13} \cdot 89^{13}$.

We see that for every formula, a number is assigned. However, not all numbers are assigned to formulas.[6] If a number is assigned to a formula, the formula can always be found as follows. Factor the number into its prime factors. Then the number of 2's occurring in the factorization is the number of the first symbol of the formula, the number of 3's occurring in the factorization is the number of the second symbol of the formula, the number of 5's occurring in the factorization is the number of the third symbol of the formula, etc.

When numbers have been assigned to formulas, statements about formulas can be replaced by statements about numbers. That is, if P is a property of formulas, we can find a property of numbers, Q, such that the formula A has the

formulation 1, this JOURNAL, vol. 1 (1936), pp. 103–105, and A. M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, ***Proceedings of the London Mathematical Society***, ser. 2 vol. 42 (1937), pp. 230–265 (see also the correction to the above, in the same journal, vol. 43 (1937), pp. 544–546). The first two definitions are proved equivalent in I. The third is proved equivalent to the first two by A. M. Turing, *Computability and λ-definability*, this JOURNAL, vol. 2 (1937), pp. 153–163.

[6] The number 4 is not assigned to any formula; for $4=2^2$, and so the first and only symbol of the formula must have the number 2 assigned to it, and no symbol has 2 assigned to it.

property P if and only if the number of A has the property Q. Throughout the rest of the paper, P will signify a property of formulas, and Q will signify the corresponding property of numbers. That is, Q will be the property of numbers such that we can use the statements "A has property P" and "the number of A has property Q" interchangeably.

Many statements about numbers can be expressed in L, even though all cannot. In particular, if P is properly chosen, we can often express "x has the property Q" in L. If x is taken to be the number of a formula of L, we are then expressing in L a statement about a formula of L. This element of circularity is capitalized in the following basic lemma:[7]

LEMMA 1. *Let "x has the property Q" be expressible in L. Then for suitable L, there can be found a formula F of L, with a number n, such that F expresses "n has the property Q." That is, F expresses "F has the property P."*

We now call attention to an extra assumption implicit in the "for suitable L" of Lemma 1,[8] namely that "$z=\phi(x, x)$" be expressible in L, where $\phi(x, y)$ is the function described below.

DEFINITION. $\phi(x, y)$ is the number of the formula got by taking the formula with the number x and replacing all occurrences of v in it by the formula of L which denotes the number of y.

We now give the proof of Lemma 1. Assume "x has the property Q" and "$z=\phi(x, x)$" are expressible in L. Then "$\phi(x, x)$ has the property Q" is expressible in L.[9] Let G be the formula of L which expresses "$\phi(x, x)$ has the property Q." G has a number, n. Now get F from G by replacing all v's of G by the formula of L which denotes n. Then F denotes "$\phi(n, n)$ has the property Q" (cf. Assumption III). However (cf. the definition of $\phi(x, y)$), $\phi(n, n)$ is the number of F, because F was got by taking the formula with the number n and replacing all occurrences of v in it by the formula of L which denotes n. So F expresses "the number of F has the property Q," that is "F has the property P."

To use Lemma 1, one must know that "$z=\phi(x, x)$" is expressible in L. Gödel proves this for a large class of L's by proving:

(a) $\phi(x, y)$ is "rekursiv" (II, pp. 179–188).

(b) If $\psi(x_1, x_2, \cdots, x_s)$ is "rekursiv," then "$z=\psi(x_1, x_2, \cdots, x_s)$" is expressible in L (II, Satz V).

The proofs of both (a) and (b) are very complicated and technical, and will not even be sketched here.

Lemma 1 is basic in GG_1, GG_2, RG_1, and RC, but is not used in CC or KG_1.

We now outline GG_1, GG_2, RG_1, and RC. Each proof depends on the choice of a suitable property P to be used in Lemma 1. Gödel chooses for P the property of not being provable in L. So if we denote (as Gödel does) "the

[7] This lemma is due to Gödel. On pp. 187–188 of II, he proves it for a particular Q. However he does not state the lemma explicitly.

[8] This assumption was not included in the original list because at that point the idea of the number of a formula had not yet been explained.

[9] Note that the statement in question is equivalent to "there is a z such that $z=\phi(x,x)$ and z has the property Q."

formula with the number x is provable in L" by "Bew(x)," then "x has property Q" is equivalent to "not-Bew(x)."

By an extensive argument involving "rekursiv" functions, Gödel shows that for a large class of L's:

(c) "Bew(x)" (and hence "not-Bew(x)") is expressible in L.

(d) If L is ω-consistent and if the formula expressing "Bew(x)" is provable, then "Bew(x)" is true.

(e) If "Bew(x)" is true, then the formula expressing "Bew(x)" is provable.

Now (Lemma 1), let us find a formula F with the number n, such that F expresses "not-Bew(n)."[10]

LEMMA 2. *If L is simply consistent, then F is not provable in L.*

For suppose F to be provable. That is, the formula with the number n is provable. That is, Bew(n). So by (e), the formula which expresses "Bew(n)" is provable. However, F expresses "not-Bew(n)," and so $\sim$F expresses "Bew(n)" (Assumption II). So $\sim$F is provable. However we assumed F provable, so that L is not simply consistent. So if L had been simply consistent, F would not have been provable.

LEMMA 3. *If L is ω-consistent, then ∼F is not provable in L.*

For suppose L to be ω-consistent and pretend that $\sim F$ is provable. $\sim F$ expresses "Bew(n)." So by (d), Bew(n). That is, F is provable. So L is not simply consistent. However ω-consistency implies simple consistency (Assumption V), so our pretense that $\sim$F could be provable has to be false.

As ω-consistency implies simple consistency, Lemma 2 and Lemma 3 together give GG_1 (which assumed ω-consistency).

GG_2 runs as follows. Let A be a provable proposition of L, and let m be the number of $\sim$A. If Bew(m), then both A and $\sim A$ are provable, and L is not simply consistent. On the other hand, if L is not simply consistent, all propositions of L are provable, including $\sim A$, so that Bew(m). Hence "not-Bew(m)" and "L is simply consistent" are equivalent. So Lemma 2 is equivalent to

"If not-Bew(m), then not-Bew(n),"

since n is the number of F. Let Wid be the formula of L which expresses "not-Bew(m)." F is the formula of L which expresses "not-Bew(n)." So

$$\text{Wid} \supset F$$

expresses Lemma 2 in L (Assumption VI). Now the proof of Lemma 2 can be carried out in a great many logics, so that in those logics

$$\text{Wid} \supset F$$

is provable. Then if Wid were provable, F would be provable (Assumption VI). So Wid is not provable if L is simply consistent (by Lemma 2), which is what Gödel's Second Theorem states.

---

[10] That is, F expresses "F is not provable." Naturally one would expect F to have certain peculiarities.

For RG_1, Rosser chooses a property, Prov(x), which differs very slightly from Bew(x).[11] By an argument involving a generalization of "rekursiv" functions, Rosser proved that for a large class of logics:

(f) Prov(x) is expressible in L.

(g) If L is simply consistent, then:

(1) If the formula expressing "Prov(x)" is provable, then "not-Bew(Neg(x))" is true.[12]

(2) If $\sim$(the formula expressing "Prov(x)") is provable, then "Bew(x)" is false.

Now (Lemma 1) let us find a formula F with the number n, such that F expresses "not-Prov(n)." Assume that L is simply consistent. Now $\sim F$ expresses "Prov(n)." So if $\sim F$ is provable, then, by (g)(1), $\sim F$ is not provable. Likewise, as F expresses "not-Prov(n)," F is $\sim$(the formula expressing "Prov(x)"). So if F is provable, then, by (g)(2), F is not provable. This completes RG_1.

For RC, we start out by assuming that L is simply consistent and that there is an effective method of deciding which propositions of L are provable. As the Herbrand-Gödel definition of "effective method" involves a generalization of "rekursiv," Rosser was able to prove for a large class of logics, by use of this generalization of "rekursiv," that there must be a property of numbers, Prov(x), such that:

(h) "Prov(x)" is expressible in L.

(i) The formula expressing "Prov(x)" in L is provable in L if and only if the formula with the number x is provable in L.

(j) Either the formula expressing "Prov(x)" or the formula expressing "not-Prov(x)" is provable in L.

Now (Lemma 1) let us choose F with the number n, so that F expresses "not-Prov(n)." By (j), either F or $\sim F$ is provable. If F is provable, then, by (i), $\sim F$ is provable, contradicting our assumption of simple consistency. If $\sim F$ is provable, then, by (i), F is provable.

The proofs KG_1 and CC do not involve Lemma 1.

We now outline KG_1. In III, Kleene shows how general recursive functions (generalizations of "rekursiv" functions) can be defined by positive integers. He further shows that in a large class of logics, "y defines a general recursive function" can be expressed. Let L be one of these logics. Then one can find a general recursive function $f(x)$ such that:

(k) As x runs over the positive integers $f(x)$ runs over those values of y such that the expression of "y is a general recursive function" is a provable formula of L.

That is, $f(x)$ enumerates a certain class of numbers which define general recursive functions, and therefore enumerates a class of (general recursive) functions. To these the diagonal process is applied to get a new function. Explicitly, Kleene defines $g(x)$ as $1+$(the value, for the argument x, of the general recursive function defined by $f(x)$). Then $g(x)$ is a general recursive

---

[11] Bew(x) and Prov(x) are equivalent if L is simply consistent and only then.

[12] If x is the number of a formula A then Neg(x) is the number of the formula $\sim A$.

function and is defined by an integer m. Let F express "m defines a general recursive function." Then F expresses a true statement, and $\sim F$ a false one, so that $\sim F$ cannot be provable if suitable consistency assumptions are made. If F were provable in L, then by (k) there would be an integer n such that $f(n)=m$. Then $g(n)=1+$(the value, for the argument n, of the general recursive function defined by $f(n))=1+$(the value, for the argument n, of $g(x))=1+g(n)$. This contradiction shows that F cannot be provable.

KG_1 may be contrasted with GG_1 and RG_1 by saying that GG_1 and RG_1 resemble the Epimenides paradox, whereas KG_1 resembles the Richard paradox.

We now outline CC. In I, Church proves of a certain set of sequences that there is no effective method of solving the problem: Given a sequence of the set, does 2 occur in it or not?

Now for a large class of logics, "2 occurs in the sequence s of the set" can be expressed in L by a formula H and moreover, if L is ω-consistent, then H is provable if and only if what it expresses is true. So an effective method of deciding whether a given formula is provable would allow one to decide effectively whether or not H is provable, and hence to solve the problem of whether 2 occurs in a given sequence.

I wish to express here my gratitude to various members of the Department of Mathematics of Cornell University who obliged me by critically reading various drafts of this paper.

Bibliography

I. Alonzo Church, *An unsolvable problem of elementary number theory*, ***American journal of mathematics***, vol. 58 (1936), pp. 345-363.

II. Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, ***Monatshefte für Mathematik und Physik***, vol. 38 (1931), pp. 173-198. (A less difficult exposition of Gödel's work is to be found in Carnap's ***The logical syntax of language***.)

III. S. C. Kleene, *General recursive functions of natural numbers*, ***Mathematische Annalen***, vol. 112 (1936), pp. 727-742.

IV. Barkley Rosser, *Extensions of some theorems of Gödel and Church*, this JOURNAL, vol. 1 (1936), pp. 87-91.

CORNELL UNIVERSITY