NOTICE This material may be protected by copyright law (Title 17 U.S. Code.)

SMULLYAN
Smullyan, Raymond
Gödel's Incompleteness Theorems
N.Y. Oxford U.P. 1992

Chapter I

The General Idea Behind Gödel's Proof

In the next several chapters we will be studying incompleteness proofs for various axiomatizations of arithmetic. Gödel, 1931, carried out his original proof for axiomatic set theory, but the method is equally applicable to axiomatic number theory. The incompleteness of axiomatic number theory is actually a stronger result since it easily yields the incompleteness of axiomatic set theory.

Gödel begins his memorable paper with the following startling words.

> "The development of mathematics in the direction of greater precision has led to large areas of it being formalized, so that proofs can be carried out according to a few mechanical rules. The most comprehensive formal systems to date are, on the one hand, the Principia Mathematica of Whitehead and Russell and, on the other hand, the Zermelo-Fraenkel system of axiomatic set theory. Both systems are so extensive that all methods of proof used in mathematics today can be formalized in them—i.e. can be reduced to a few axioms and rules of inference. It would seem reasonable, therefore, to surmise that these axioms and rules of inference are sufficient to decide *all* mathematical questions which can be formulated in the system concerned. In what follows it will be shown that this is not the case, but rather that, in both of the cited systems, there exist relatively simple problems of the theory of ordinary whole numbers which cannot be decided on the basis of the axioms."

Gödel then goes on to explain that the situation does not depend on the special nature of the two systems under consideration but holds for an extensive class of mathematical systems.

Just what is this "extensive class" of mathematical systems? Various interpretations of this phrase have been given, and Gödel's theorem has accordingly been generalized in several ways. We will consider many such generalizations in the course of this volume. Curiously enough, one of the generalizations that is most direct and most easily accessible to the general reader is also the one that appears to be the least well known. What makes this particularly curious is that the way in question is the very one indicated by Gödel himself in the introductory section of his original paper! We shall shortly turn to this (or rather to a further generalization of it), but before that, we would like the reader to look at the following little puzzles which illustrate Gödel's essential idea in a simple and instructive way.

A Gödelian Puzzle. Let us consider a computing machine that prints out various expressions composed of the following five symbols:

$$\sim P\, N\, (\ )$$

By an *expression*, we mean any finite non-empty string of these five symbols. An expression X is called *printable* if the machine can print it. We assume the machine programmed so that any expression that the machine can print will be printed sooner or later.

By the *norm* of an expression X, we shall mean the expression $X(X)$—e.g. the norm of $P\sim$ is $P\sim(P\sim)$. By a *sentence*, we mean any expression of one of the following four forms (X is any expression):

(1) $P(X)$
(2) $PN(X)$
(3) $\sim P(X)$
(4) $\sim PN(X)$

Informally, P stands for "printable"; N stands for "the norm of" and $\sim$ stands for "not". And so we define $P(X)$ to be *true* if (and only if) X is printable. We define $PN(X)$ to be true if the *norm* of X is printable. We call $\sim P(X)$ true iff (if and only if) X is not printable, and $\sim PN(X)$ is defined to be true iff the norm of X is not printable. [This last sentence we read as "Not printable the norm of X", or, in better English: "The norm of X is not printable".]

We have now given a perfectly precise definition of what it means for a sentence to be true, and we have here an interesting case of self-reference: The machine is printing out various sentences about what the machine can and cannot print, and so it is describing its own behavior! [It somewhat resembles a self-conscious organism, and we can see why such computers are of interest to those working in artificial intelligence.]

We are given that the machine is completely accurate in that all

sentences printed by the machine are true. And so, for example, if the machine ever prints $P(X)$, then X really is printable (X will be printed by the machine sooner or later). Also, if $PN(X)$ is printable, so is $X(X)$ (the norm of X). Now, suppose X is printable. Does it follow that $P(X)$ is printable? Not necessarily. If X is printable, then $P(X)$ is certainly *true*, but we are not given that the machine is capable of printing *all* true sentences but only that the machine never prints any false ones. [Whether the machine can print expressions that are not sentences at all is immaterial. The important thing is that among the *sentences* printable by the machine, all of them are true.]

Is it possible that the machine *can* print all true sentences? The answer is *no* and the problem for the reader is this: Find a true sentence that the machine cannot print. [Hint: Find a sentence that asserts its own non-printability—i.e. one which is true if and only if it is not printable by the machine. The solution is given after the next problem.]

A Variant of the Puzzle. The following variant of the above puzzle will introduce the reader to the notion of *Gödel numbering*.

We now have another machine that prints out expressions composed of the following five symbols:

$$\sim P\, N\, 1\, 0$$

We are representing the natural numbers in binary notation (as strings of 1's and 0's), and for purposes of this problem, we will identify the natural numbers with the binary numerals that represent them.

To each expression we assign a number which we call the *Gödel number* of the expression. We do this according to the following scheme: The individual symbols $\sim, P, N, 1, 0$ are assigned the respective Gödel numbers $10, 100, 1000, 10000, 100000$. Then, the Gödel number of a compound expression is obtained by replacing each symbol by *its* Gödel number—for example, PNP has Gödel number 1001000100. We redefine the *norm* of an expression to be the expression followed by its Gödel number—for example, the norm of PNP is the expression $PNP1001000100$. A *sentence* is now an expression of one of the four forms: PX, PNX, $\sim PX$ and $\sim PNX$, where X is any number (written in binary notation). We call PX true if X is the *Gödel number* of a printable expression. We call PNX true iff X is the Gödel number of an expression whose *norm* is printable. We call $\sim PX$ true if PX is not true (X is not the Gödel number of a printable expression), and we call $\sim PNX$ true iff PNX is not true.

Again we are given that the machine never prints a false sentence. Find a true sentence that the machine cannot print.

Solutions. For the first problem, the sentence is $\sim PN(\sim PN)$. By definition of "true", this sentence is true if and only if the norm of $\sim PN$ is not printable. But the norm of $\sim PN$ is the very sentence $\sim PN(\sim PN)$! And so the sentence is true if and only if it is not printable. This means that either the sentence is true and not printable, or it is printable and not true. The latter alternative violates the given hypothesis that the machine never prints sentences that are not true. Hence the sentence must be true, but the machine cannot print it.

Of course, instead of having talked about a machine that *prints* various expressions in our five symbols, we could have talked about a mathematical system that *proves* various sentences in the same five symbols. We would then reinterpret the letter P to mean *provable* in the system, rather than printable by the machine. Then, given that the system is wholly accurate (in that, false sentences are never provable in it), the sentence $\sim PN(\sim PN)$ would be a sentence that is true but not provable in the system.

Let us further observe that the sentence $PN(\sim PN)$ is false (since its negation is true). Hence it is also not provable in the system (assuming that the system is accurate). And so the sentence

$$PN(\sim PN)$$

is an example of a sentence *undecidable* in a system—i.e. neither it nor its negation is provable in the system.

For the second problem, the solution is $\sim PN101001000$.

Now we shall turn to some incompleteness arguments in a general setting: We consider a very broad notion of a mathematical system and show that if it has certain features, then Gödel's argument goes through. In the chapters that follow, we will look at some particular systems and show that they do indeed possess these features.

I. Abstract Forms of Gödel's and Tarski's Theorems

Each of the languages $\mathcal{L}$ to which Gödel's argument is applicable contains at least the following items.

1. A denumerable set $\mathcal{E}$ whose elements are called the *expressions* of $\mathcal{L}$.
2. A subset $\mathcal{S}$ of $\mathcal{E}$ whose elements are called the *sentences* of $\mathcal{L}$.
3. A subset $\mathcal{P}$ of $\mathcal{S}$ whose elements are called the *provable* sentences of $\mathcal{L}$.
4. A subset $\mathcal{R}$ of $\mathcal{S}$ whose elements are called the *refutable* (sometimes *disprovable*) sentences of $\mathcal{L}$.
5. A set $\mathcal{H}$ of expressions whose elements are called the *predicates* of $\mathcal{L}$. [These were called *class names* in Gödel's introduction. Informally, each predicate H is thought of as being the name of a set of natural numbers.]
6. A function Φ that assigns to every expression E and every natural number n an expression $E(n)$. The function is required to obey the condition that for every predicate H and every natural number n, the expression $H(n)$ is a sentence. [Informally, the sentence $H(n)$ expresses the proposition that the number n belongs to the set named by H.]

   In the first incompleteness proof that we will give for a *particular* system $\mathcal{L}$, we will use a basic concept made precise by Alfred Tarski [1936]—viz. the notion of a *true* sentence (defined quite differently than that of a provable sentence of a system). And so we consider a seventh and final item of our language $\mathcal{L}$.
7. A set $\mathcal{T}$ of sentences whose elements are called the *true* sentences of $\mathcal{L}$.

This concludes our abstract description of the type of systems that we will study in the next several chapters.

Expressibility in $\mathcal{L}$. The notion of *expressibility* in $\mathcal{L}$, which we are about to define, concerns the truth set $\mathcal{T}$ but does not concern either of the sets $\mathcal{P}$ and $\mathcal{R}$.

The word *number* shall mean *natural number* for the rest of this volume. We will say that a predicate H is *true* for a number n or that n *satisfies* H if $H(n)$ is a true sentence (i.e. is an element of $\mathcal{T}$). By the set *expressed* by H, we mean the set of all n that satisfy

H. Thus for any set A of numbers, H expresses A if and only if for every number n:

$$H(n) \in \mathcal{T} \leftrightarrow n \in A.$$

Definition. A set A is called *expressible* or *nameable* in $\mathcal{L}$ if A is expressed by some predicate of $\mathcal{L}$.

Since there are only denumerably many expressions of $\mathcal{L}$, then there are only finitely or denumerably many predicates of $\mathcal{L}$. But by Cantor's well-known theorem, there are non-denumerably many sets of natural numbers. Therefore, not every set of numbers is expressible in $\mathcal{L}$.

Definition. The system $\mathcal{L}$ is called *correct* if every provable sentence is true and every refutable sentence is false (not true). This means that $\mathcal{P}$ is a subset of $\mathcal{T}$ and $\mathcal{R}$ is disjoint from $\mathcal{T}$. We are now interested in sufficient conditions that $\mathcal{L}$, if correct, must contain a true sentence not provable in $\mathcal{L}$.

Gödel Numbering and Diagonalization. We let g be a 1-1 function which assigns to each expression E a natural number $g(E)$ called the *Gödel number* of E. The function g will be constant for the rest of this chapter. [In the concrete systems to be studied in subsequent chapters, a specific Gödel numbering will be given. Our present purely abstract treatment, however, applies to an arbitrary Gödel numbering.] It will be technically convenient to assume that every number is the Gödel number of an expression. [Gödel's original numbering did not have this property, but the Gödel numbering we will use in subsequent chapters will have this property. However, the results of this chapter can, with minor modifications, be proved without this restriction (cf. Ex. 5).] Assuming now that every number n is the Gödel number of a unique expression, we let E_n be that expression whose Gödel number is n. Thus, $g(E_n) = n$.

By the *diagonalization* of E_n we will mean the expression $E_n(n)$. If E_n is a predicate, then its diagonalization is, of course, a sentence; this sentence is true iff the predicate E_n is satisfied by its own Gödel number n. [We write "iff" to mean if and only if; we use "$\leftrightarrow$" synonymously.]

For any n, we let $d(n)$ be the Gödel number of $E_n(n)$. The function $d(x)$ plays a key rôle in all that follows; we call it the *diagonal* function of the system.

We use the term *number-set* to mean set of (natural) numbers. For any number set A, by A^* we shall mean the set of all numbers

n such that $d(n) \in A$. Thus for any n, the equivalence

$$n \in A^* \leftrightarrow d(n) \in A$$

holds by definition of A^*. [A^* could also be written $d^{-1}(A)$, since it is the inverse image of A under the diagonal function $d(x)$.]

An Abstract Form of Gödel's Theorem. We let P be the set of Gödel numbers of all the provable sentences. For any number set A, by its complement $\tilde{A}$, we mean the complement of A relative to the set N of natural numbers—i.e. $\tilde{A}$ is the set of all natural numbers not in A.

Theorem (GT)—After Gödel with shades of Tarski. *If the set $\tilde{P}^*$ is expressible in $\mathcal{L}$ and $\mathcal{L}$ is correct, then there is a true sentence of $\mathcal{L}$ not provable in $\mathcal{L}$.*

Proof. Suppose $\mathcal{L}$ is correct and $\tilde{P}^*$ is expressible in $\mathcal{L}$. Let H be a predicate that expresses $\tilde{P}^*$ in $\mathcal{L}$, and let h be the Gödel number of H. Let G be the diagonalization of H (i.e. the sentence $H(h)$). We will show that G is true but not provable in $\mathcal{L}$.

Since H expresses $\tilde{P}^*$ in $\mathcal{L}$, then for any number n, $H(n)$ is *true* $\leftrightarrow$ $n \in \tilde{P}^*$. Since this equivalence holds for *every* n, then it holds in particular for n the number h. So we take h for n (and this is the part of the argument called *diagonalizing*) and we have the equivalence: $H(h)$ is *true* $\leftrightarrow h \in \tilde{P}^*$. Now,

$$h \in \tilde{P}^* \leftrightarrow d(h) \in \tilde{P} \leftrightarrow d(h) \notin P.$$

But $d(h)$ is the Gödel number of $H(h)$ (since h is the Gödel number of H) and so $d(h) \in P \leftrightarrow H(h)$ is provable in $\mathcal{L}$ and $d(h) \notin P \leftrightarrow H(h)$ is not provable in $\mathcal{L}$. And so we have

1. $H(h)$ is *true* $\leftrightarrow$ $H(h)$ is not provable in $\mathcal{L}$. This means that $H(h)$ is either true and not provable in $\mathcal{L}$ or false but provable in $\mathcal{L}$. The latter alternative violates the hypothesis that $\mathcal{L}$ is correct. Hence it must be that $H(h)$ is true but not provable in $\mathcal{L}$.

When it comes to the particular languages $\mathcal{L}$ that we will study, we will verify the hypothesis that $\tilde{P}^*$ is expressible in $\mathcal{L}$ by separately verifying the following three conditions.

G_1: For any set A expressible in $\mathcal{L}$, the set A^* is expressible in $\mathcal{L}$.
G_2: For any set A expressible in $\mathcal{L}$, the set $\tilde{A}$ is expressible in $\mathcal{L}$.
G_3: The set P is expressible in $\mathcal{L}$.

Conditions G_1 and G_2, of course, imply that for any set A expressible in $\mathcal{L}$, the set $\tilde{A}^*$ is expressible in $\mathcal{L}$. Hence if P is expressible in $\mathcal{L}$, then so is $\tilde{P}^*$.

We might remark that the verification of G_1 will turn out to be relatively simple; the verification of G_2 will be completely trivial; but the verification of G_3 will turn out to be extremely elaborate.

Gödel Sentences. Woven into the proof of Theorem GT is a very important principle which was made explicit by Rudolf Carnap [1934] and which is closely related to Tarski's theorem, to which we will soon turn.

Call a sentence E_n a *Gödel* sentence for a number set A if either E_n is true and its Gödel number n lies in A, or E_n is false and its Gödel number lies outside A. Thus, E_n is a Gödel sentence for A iff the following condition holds:

$$E_n \in \mathcal{T} \leftrightarrow n \in A.$$

[Informally, a Gödel sentence for A can be thought of as a sentence asserting that its own Gödel number lies in A. If the sentence is true, then its Gödel number does lie in A. If the sentence if false, then its Gödel number does not lie in A.]

The following lemma and theorem pertains only to the set $\mathcal{T}$. The sets $\mathcal{P}$ and $\mathcal{R}$ are irrelevant.

Lemma (D)—A Diagonal Lemma. (a) For any set A, if A^* is expressible in $\mathcal{L}$, then there is a Gödel sentence for A.

(b) If $\mathcal{L}$ satisfies condition G_1, then for any set A expressible in $\mathcal{L}$, there is a Gödel sentence for A.

Proof.

(a) Suppose H is a predicate that expresses A^* in $\mathcal{L}$; let h be its Gödel number. Then $d(h)$ is the Gödel number of $H(h)$. For any number n, $H(n)$ is ***true*** $\leftrightarrow n \in A^*$, therefore, $H(h)$ is ***true*** $\leftrightarrow h \in A^*$. And $h \in A^* \leftrightarrow d(h) \in A$. Therefore, $H(h)$ is ***true*** $\leftrightarrow d(h) \in A$, and since $d(h)$ is the Gödel number of $H(h)$, then $H(h)$ is a Gödel sentence for A.

(b) Immediate from (a).

Let us note that if we had first proved Lemma D, we would have had the following swift proof of Theorem GT: Since $\tilde{P}^*$ is nameable in $\mathcal{L}$, then by lemma D, there is a Gödel sentence G for $\tilde{P}$. A Gödel

sentence for $\tilde{P}$ is nothing more nor less than a sentence which is true if and only if it is not provable (in $\mathcal{L}$). And for any *correct* system $\mathcal{L}$, a Gödel sentence for $\tilde{P}$ is a sentence which is true but not provable in $\mathcal{L}$. [Such a sentence can be thought of as asserting its own non-provability in $\mathcal{L}$.]

An Abstract Form of Tarski's Theorem. Lemma D has another important consequence: Let T be the set of Gödel numbers of the ***true*** sentences of $\mathcal{L}$. Then the following theorem holds.

Theorem (T) (After Tarski).

1. *The set $\tilde{T}^*$ is not nameable in $\mathcal{L}$.*
2. *If condition G_1 holds, then $\tilde{T}$ is not nameable in $\mathcal{L}$.*
3. *If conditions G_1 and G_2 both hold, then the set T is not nameable in $\mathcal{L}$.*

Proof. To begin with, there cannot possibly be a Gödel sentence for the set $\tilde{T}$ because such a sentence would be true if and only if its Gödel number was *not* the Gödel number of a true sentence, and this is absurd.

1. If $\tilde{T}^*$ were nameable in $\mathcal{L}$, then by (a) of Lemma D, there would be a Gödel sentence for the set $\tilde{T}$, which we have just shown is impossible. Therefore, $\tilde{T}^*$ is not nameable in $\mathcal{L}$.
2. Suppose condition G_1 holds. Then if $\tilde{T}$ were nameable in $\mathcal{L}$, the set $\tilde{T}^*$ would be nameable in $\mathcal{L}$, violating (1).
3. If G_2 also holds, then if T were nameable in $\mathcal{L}$, then $\tilde{T}$ would also be nameable in $\mathcal{L}$, violating (2).

Remarks.

1. Conclusion (3) above is sometimes paraphrased: For systems of sufficient strength, truth within the system is not definable within the system. The phrase "sufficient strength" has been interpreted in several ways. We would like to point out that conditions G_1 and G_2 suffice for this "sufficient strength."
2. Gödel (1931) likens his proof to the famous paradox of the Cretan who says that all Cretans are liars.[1] An analogy that comes closer to Gödel's theorem is this: Imagine a land in which every inhabitant either always tells the truth or always lies. Some of the inhabitants are Athenians and some are Cretans. It is given

[1] Actually, the liar paradox is more closely related to Tarski's theorem than to Gödel's.

that all the Athenians of the land always tell the truth and all the Cretans of the land always lie. What statement could an inhabitant make that would convince you that he always tells the truth but that he is not an Athenian?

All he would need to say is: "I am not an Athenian." A liar couldn't make that claim (because a liar is really *not* an Athenian; only truth-tellers are Athenian). Therefore, he must be truthful. Hence his statement was true, which means that he is really not an Athenian. So he is a truth teller but not an Athenian.

If we think of the Athenians as playing the rôle of the sentences of $\mathcal{L}$, which are not only true but provable in $\mathcal{L}$, then any inhabitant who claims he is not Athenian plays the rôle of Gödel's sentence G, which asserts its own non-provability in $\mathcal{L}$. [The Cretans, of course, play the rôle of the refutable sentences of $\mathcal{L}$, but their function won't emerge till a bit later.]

II. Undecidable Sentences of $\mathcal{L}$

So far, the set $\mathcal{R}$ of *refutable* sentences has played no rôle. Now it shall play a key one.

$\mathcal{L}$ is called *consistent* if no sentence is both provable and refutable in $\mathcal{L}$ (i.e. the sets $\mathcal{P}$ and $\mathcal{R}$ are disjoint) and *inconsistent* otherwise. The definition of consistency refers only to the sets $\mathcal{P}$ and $\mathcal{R}$, not to the set $\mathcal{T}$. Nevertheless, if $\mathcal{L}$ is correct, then it is automatically consistent (because if $\mathcal{P}$ is a subset of $\mathcal{T}$ and $\mathcal{T}$ is disjoint from $\mathcal{R}$, then $\mathcal{P}$ must be disjoint from $\mathcal{R}$). The converse is not necessarily true (we will later consider some systems that are consistent but not correct).

A sentence X is called *decidable* in $\mathcal{L}$ if it is either provable or refutable in $\mathcal{L}$ and *undecidable* in $\mathcal{L}$ otherwise. The system $\mathcal{L}$ is called *complete* if every sentence is decidable in $\mathcal{L}$ and *incomplete* if some sentence is undecidable in $\mathcal{L}$.

Suppose now $\mathcal{L}$ satisfies the hypothesis of Theorem GT. Then some sentence G is true but not provable in $\mathcal{L}$. Since G is true, it is not refutable in $\mathcal{L}$ either (by the assumption of correctness). Hence G is undecidable in $\mathcal{L}$. And so we at once have

Theorem 1. If $\mathcal{L}$ is correct and if the set $\widetilde{P}^*$ is expressible in $\mathcal{L}$, then $\mathcal{L}$ is incomplete.

A Dual of Theorem 1. In T.F.S. (Theory of Formal Systems, 1961) we introduced what might aptly be called a "dual form" of Gödel's argument, which we will first explain informally. Instead of constructing a sentence that says "I am not provable," we will construct a sentence that says "I *am* refutable." As we are about to see, such a sentence must also be undecidable in $\mathcal{L}$ (if $\mathcal{L}$ is correct).

We have defined P to be the set of Gödel numbers of the provable sentences. We now define R to be the set of Gödel numbers of the refutable sentences.

Theorem (1°)—(A Dual of Theorem 1). If $\mathcal{L}$ is correct and the set R^* is expressible in $\mathcal{L}$, then $\mathcal{L}$ is incomplete. More specifically, if $\mathcal{L}$ is correct and K is a predicate that expresses the set R^*, then its diagonalization $K(k)$ is undecidable in $\mathcal{L}$ (k is the Gödel number of K).

Proof. Assume hypothesis. Since K expresses R^*, then by the proof of (a) of Lemma D, the sentence $K(k)$ is a Gödel sentence for the set R. Thus, $K(k)$ is true iff its Gödel number is in R, or, what is the same thing, $K(k)$ is true iff $K(k)$ is in $\mathcal{R}$, so $K(k)$ is true iff $K(k)$ is refutable in $\mathcal{L}$. This means that $K(k)$ is either true and refutable or false but not refutable. By the assumption of correctness, $K(k)$ cannot be true and refutable. Hence it is false but not refutable. Since the sentence is false, it is not provable either (again by the assumption that $\mathcal{L}$ is correct). Hence $K(k)$ is neither provable nor refutable in $\mathcal{L}$.

Remarks. Just as the Gödel sentence $H(h)$ can be thought of as saying: "I am not provable in $\mathcal{L}$," we can think of $K(k)$ as saying: "I am refutable in $\mathcal{L}$." Going back to our analogy of Athenians and Cretans, just as $H(h)$ corresponds to an inhabitant who claims that he is not an Athenian, the sentence $K(k)$ corresponds to an inhabitant who claims that he *is* a Cretan. He must be a liar but not a Cretan. Hence (like an inhabitant who claims he is not an Athenian) he must be neither an Athenian nor a Cretan.

Suppose now we have a correct system $\mathcal{L}$ satisfying the following two conditions:

G_1: For any expressible set A, the set A^* is expressible

$G_3{}'$: The set R is expressible

Then, of course, the set R^* is expressible. So by Theorem 1°, $\mathcal{L}$

is inconsistent or incomplete. We note that the complementation condition G_2 is not required in this proof.

The first exercise below contains an interesting variant of Theorem 1°.

Exercise 1. Suppose $\mathcal{L}$ is a correct system such that the following two conditions hold.

1. The set P^* is expressible in $\mathcal{L}$.
2. For any predicate H, there is a predicate H' such that for every n, the sentence $H'(n)$ is provable in $\mathcal{L}$ if and only if $H(n)$ is refutable in $\mathcal{L}$.

Prove that $\mathcal{L}$ is incomplete.

Exercise 2. We say that a predicate H *represents* a set A in $\mathcal{L}$ if for every number n, the sentence $H(n)$ is *provable* in $\mathcal{L}$ if and only if $n \in A$. [Note that this definition makes no reference to the truth set $\mathcal{T}$ but only to the provability set $\mathcal{P}$.]

Show that if the set R^* is representable in $\mathcal{L}$, then $\mathcal{L}$, if consistent, is incomplete.

Exercise 3. Show that if some superset of R^* disjoint from P^* is representable in $\mathcal{L}$, then $\mathcal{L}$ is incomplete. [We call B a *superset* of A if A is a subset of B.]

Exercise 4. Let us say that a predicate H *contrarepresents* a set A in $\mathcal{L}$ if for every number n, the sentence $H(n)$ is *refutable* in $\mathcal{L}$ iff $n \in A$. Show that if the set P^* is contrarepresentable in $\mathcal{L}$ and $\mathcal{L}$ is consistent, then $\mathcal{L}$ is incomplete. [This result and the result of Exercise 2 will be expanded in Chapter 5; the result of Exercise 3 is related to Rosser's incompleteness proof, which we will study in Chapter 6.]

Exercise 5. Suppose we have a Gödel numbering g such that it is not the case that every number is a Gödel number. Then we define a function $d(x)$ to be *a* diagonal function (rather than *the* diagonal function) if it has the property that for any number e, *if* e is the Gödel number of an expression E, then $d(e)$ is the Gödel number of $E(e)$. Prove that for any diagonal function $d(x)$, if $d^{-1}(A)$ is expressible in $\mathcal{L}$, then there is a Gödel sentence for A.

Exercise 6. Is it necessarily true that for any set A, the set $\widetilde{A}^*$ is the same as the set $\widetilde{A^*}$?

Exercise 7. To emphasize the wholly constructive nature of Gödel's

proof, suppose $\mathcal{L}$ is a correct system such that the following three conditions hold.

1. E_7 is a predicate that expresses the set P.
2. For any number n, if E_n is a predicate, then so is E_{3n}, and E_{3n} expresses the complement of the set expressed by E_n.
3. For any number n, if E_n is a predicate, then E_{3n+1} is a predicate, and if A is the set expressed by E_n, then A^* is the set expressed by E_{3n+1}.

(a) Find numbers a and b (either the same or different) such that $E_a(b)$ is a true sentence not provable in $\mathcal{L}$. [There are two solutions in which a and b are both less than 100. Can the reader find them both?]

(b) Show that there are infinitely many pairs (a, b) such that $E_a(b)$ is true but not provable in $\mathcal{L}$.

(c) Given that E_{10} is a predicate, find numbers c and d such that $E_c(d)$ is a Gödel sentence for the set expressed by E_{10}.