17

Indefinability, Undecidability, Incompleteness

We are now in a position to give a unified treatment of some of the central negative results of logic: Tarski's theorem *on the indefinability of truth,* Church's theorem *on the undecidability of logic, and* Gödel's first incompleteness theorem, *according to which, roughly speaking, any sufficiently strong formal system of arithmetic must be incomplete (if it is consistent). These theorems can all be seen as more or less direct consequences of a single exceedingly ingenious lemma, the* Gödel diagonal lemma. *This lemma, and the various negative results on the limits of logic that follow from it, will be presented in section 17.1. This presentation will be followed by a discussion in section 17.2 of some classic particular examples of sentences that can be neither proved nor disproved in theories of arithmetic like* **Q** *or* **P**. *Further such examples will be presented in the optional section 17.3. According to* Gödel's second incompleteness theorem, *the topic of the next chapter, such examples also include the sentence stating that* **P** *is consistent.*

17.1 The Diagonal Lemma and the Limitative Theorems

By the results in the preceding chapter on the representability of recursive functions, we can 'talk about' such functions within a formal system of arithmetic. By the results of the chapter before that on the arithmetization of syntax, we can 'talk about' sentences and proofs in a formal system of arithmetic in terms of recursive functions. Putting the two together, we can 'talk about' sentences and proofs in a formal system of arithmetic *within the formal system of arithmetic itself*. This is the key to the main lemma of this section, the diagonal lemma.

Until further notice, all formulas, sentences, theories, and so on, will be formulas, sentences, theories, or whatever in the language of arithmetic. Given any expression A of the language of arithmetic, we have introduced in Chapter 15 a code number for A, called the Gödel number of A. If a is this number, then the numeral $\mathbf{a}$ for a, consisting of $\mathbf{0}$ followed by a accents $'$, is naturally called the *Gödel numeral* for A. We write $\ulcorner A \urcorner$ for this code numeral for A. In what follows, $\ulcorner A \urcorner$ will be seen to function somewhat like a name for A.

We define the *diagonalization* of A to be the expression $\exists x(x = \ulcorner A \urcorner \,\&\, A)$. While this notion makes sense for arbitrary expressions, it is of most interest in the case of a formula $A(x)$ with just the one variable x free. Since in general $F(t)$ is equivalent to $\exists x(x = t \,\&\, F(x))$, in case A is such a formula, the diagonalization of A is a sentence equivalent to $A(\ulcorner A \urcorner)$, the result of substituting the code numeral for A itself for the free variable in A. In this case the diagonalization 'says that' A is satisfied by its own Gödel number, or more precisely, the diagonalization will be true in the standard interpretation if and only if A is satisfied by its own Gödel number in the standard interpretation.

17.1 Lemma (Diagonal lemma). Let T be a theory containing **Q**. Then for any formula $B(y)$ there is a sentence G such that $\vdash_T G \leftrightarrow B(\ulcorner G \urcorner)$.

Proof: There is a (primitive) recursive function, diag, such that if a is the Gödel number of an expression A, then diag(a) is the Gödel number of the diagonalization of A. Indeed, we have seen almost exactly the function we want before, in the proof of Corollary 15.6. Recalling that officially $x = y$ is supposed to be written $=(x, y)$, it can be seen to be

$$\text{diag}(y) = \text{exquant}(v, \text{conj}(i * l * v * c * \text{num}(y) * r, y))$$

where v is the code number for the variable, i, l, r, and c for the equals sign, left and right parentheses, and comma, and exquant, conj, and num are as in Proposition 15.1 and Corollary 15.6.

If T is a theory extending **Q**, then diag is representable in T by Theorem 16.16. Let Diag(x, y) be a formula representing diag, so that for any m and n, if diag $(m) = n$, then $\vdash_T \forall y(\text{Diag}(\mathbf{m}, y) \leftrightarrow y = \mathbf{n})$.

Let $A(x)$ be the formula $\exists y(\text{Diag}(x, y) \,\&\, B(y))$. Let a be the Gödel number of $A(x)$, and $\mathbf{a}$ its Gödel numeral. Let G be the sentence $\exists x(x = \mathbf{a} \,\&\, \exists y(\text{Diag}(x, y) \,\&\, B(y)))$.

Thus G is $\exists x(x = \mathbf{a} \,\&\, A(x))$, and is logically equivalent to $A(\mathbf{a})$ or $\exists y(\text{Diag}(\mathbf{a}, y) \,\&\, B(y))$. The biconditional $G \leftrightarrow \exists y(\text{Diag}(\mathbf{a}, y) \,\&\, B(y))$ is therefore valid, and as such provable in any theory, so we have

$$\vdash_T G \leftrightarrow \exists y(\text{Diag}(\mathbf{a}, y) \,\&\, B(y)).$$

Let g be the Gödel number of G, and $\mathbf{g}$ its Gödel numeral. Since G is the diagonalization of $A(x)$, diag$(a) = g$ and so we have

$$\vdash_T \forall y(\text{Diag}(\mathbf{a}, y) \leftrightarrow y = \mathbf{g}).$$

It follows that

$$\vdash_T G \leftrightarrow \exists y(y = \mathbf{g} \,\&\, B(y)).$$

Since $\exists y(y = \mathbf{g} \,\&\, B(y))$ is logically equivalent to $B(\mathbf{g})$, we have

$$\vdash_T \exists y(y = \mathbf{g} \,\&\, B(y)) \leftrightarrow B(\mathbf{g}).$$

It follows that

$$\vdash_T G \leftrightarrow B(\mathbf{g})$$

or in other words, $\vdash_T G \leftrightarrow B(\ulcorner G \urcorner)$, as required.

17.2 Lemma. Let T be a consistent theory extending **Q**. Then the set of Gödel numbers of theorems of T is not definable in T.

Proof: Let T be an extension of **Q**. Suppose $\theta(y)$ defines the set Θ of Gödel numbers of sentences in T. By the diagonal lemma there is a sentence G such that

$$\vdash_T G \leftrightarrow {\sim}\theta(\ulcorner G \urcorner).$$

In other words, letting g be the Gödel number of G, and $\mathbf{g}$ its Gödel numeral, we have

$$\vdash_T G \leftrightarrow {\sim}\theta(\mathbf{g}).$$

Then G is a theorem of T. For if we assume G is not a theorem of T, then g is not in Θ, and since $\theta(y)$ defines Θ, we have $\vdash_T {\sim}\theta(\mathbf{g})$; but then since $\vdash_T G \leftrightarrow {\sim}\theta(\mathbf{g})$, we have $\vdash_T G$ and G is a theorem of T after all. But since G is a theorem of T, g is in Θ, and so we have $\vdash_T \theta(\mathbf{g})$; but then, since $\vdash_T G \leftrightarrow {\sim}\theta(\mathbf{g})$, we have $\vdash_T {\sim}G$, and T is inconsistent.

Now the 'limitative theorems' come tumbling out in rapid succession.

17.3 Theorem (Tarski's theorem). The set of Gödel numbers of sentences of the language of arithmetic that are correct, or true in the standard interpretation, is not arithmetically definable.

Proof: The set T in question is the theory we have been calling true arithmetic. It is a consistent extension of **Q**, and arithmetic definability is simply definability in this theory, so the theorem is immediate from Lemma 17.2.

17.4 Theorem (Undecidability of arithmetic). The set of Gödel numbers of sentences of the language of arithmetic that are correct, or true in the standard interpretation, is not recursive.

Proof: This follows from Theorem 17.3 and the fact that all recursive sets are definable in arithmetic.

Assuming Church's thesis, this means that the set in question is not effectively decidable: there are no rules—of a kind requiring only diligence and persistence, not ingenuity and insight, to execute—with the property that applied to any sentence of the language of arithmetic they will eventually tell one whether or not it is correct.

17.5 Theorem (Essential undecidability theorem). No consistent extension of **Q** is decidable (and in particular, **Q** itself is undecidable).

Proof: Suppose T is a consistent extension of **Q** (in particular, T could just be **Q** itself). Then by Lemma 17.2, the set Θ of Gödel numbers of theorems of T is not definable in T. Now, again as in the proof of Theorem 17.4, we invoke the fact that every recursive set is definable in T. So the set Θ is not recursive, which is to say T is not decidable.

17.6 Theorem (Church's theorem). The set of valid sentences is not decidable.

Proof: Let C be the conjunction of all the axioms of **Q**. Then a sentence A is a theorem of **Q** if and only if A is a consequence of C, hence if and only if $(\sim C \vee A)$ is valid. The function f taking the Gödel number of A to that of $(\sim C \vee A)$ is recursive [it being simply $f(y) = \text{disj}(\text{neg}(c), y)$, in the notation of the proof of Proposition 15.1]. If the set Λ of logically valid sentences were recursive, the set K of Gödel numbers of theorems of **Q** would be obtainable from it by substitution of the recursive function f, since a is in K if and only if $f(n)$ is in Λ, and so would be recursive, which it is not by Theorem 17.4.

The sets of valid sentences, and of theorems of any axiomatizable theory, are semirecursive by Corollaries 15.4 and 15.5, and intuitively, of course, both are *positively* effectively decidable: in principle, if not in practice, just by searching through all demonstrations (or all proofs from the axioms of the theory), *if* a sentence is valid (or a theorem of the theory), one will eventually find that out. But Theorems 17.5 and 17.6 tell us these sets are not recursive, and so by Church's thesis are not effectively decidable.

17.7 Theorem (Gödel's first incompleteness theorem). There is no consistent, complete, axiomatizable extension of **Q**.

Proof: Any complete axiomatizable theory is decidable by Corollary 15.7, but no consistent extension of **Q** is decidable by Theorem 17.5 above.

The import of Gödel's first incompleteness theorem is sometimes expressed in the words 'any sufficiently strong formal system of arithmetic (or mathematics) is incomplete, unless it is inconsistent'. Here by 'formal system' is meant a theory whose theorems are derivable by the rules of logical derivation from a set of axioms that is effectively decidable, and hence (assuming Church's thesis) recursive. So 'formal system' amounts to 'axiomatizable theory', and 'formal system of arithmetic' to 'axiomatizable theory in the language of arithmetic'. Gödel's first incompleteness theorem in the version in which we have given it indicates a sufficient condition for being 'sufficiently strong', namely, being an extension of **Q**. Since **Q** is a comparatively weak theory, this version of Gödel's first incompleteness theorem is a correspondingly strong result.

Now a formal system of mathematics might well be such that the domain of its intended interpretation was a more inclusive set than the set of natural numbers, and it might well be such that it did not have symbols specifically for 'less than' and the other items for which there are symbols in the language of arithmetic. So the principle that any two natural numbers are comparable as to ordering might not be expressed, as it is in the axioms of **Q**, by the sentence

$$\forall x \forall y (x < y \vee x = y \vee y < x).$$

Still, it is reasonable to understand 'sufficiently strong' as implying that this principle can be *somehow* expressed in the language of the theory, perhaps by a sentence

$$\forall x(N(x) \to \forall y(N(y) \to (L(x, y) \lor x = y \lor L(y, x))))$$

where $N(x)$ appropriately expresses 'x is a natural number' and $L(x, y)$ appropriately expresses 'x is less than y'. Moreover, the sentence that thus 'translates' this or any axiom of **Q** should be a theorem of the theory. Such is the case, for instance, with the formal systems considered in works on set theory, such as the one known as **ZFC**, which are adequate for formalizing essentially all accepted mathematical proofs. When the notion of 'translation' is made precise, it can be shown that any 'sufficiently strong' formal system of mathematics in the sense we have been indicating is still subject to the limitative theorems of this chapter. In particular, if consistent, it will be incomplete.

Perhaps the most important implication of the incompleteness theorem is what is says about the notions of *truth* (in the standard interpretation) and *provability* (in a formal system): *that they are in no sense the same.*

17.2 Undecidable Sentences

A sentence in the language of a theory T is said to be *disprovable in T* if its negation is provable in T, and is said to be *undecidable in* or *by* or *for T* if it is neither provable nor disprovable in T. (Do not confuse the notion of an undecidable sentence with that of an undecidable theory. True arithmetic, for example, is an undecidable theory with no undecidable sentences: the sentences of its language that are true in the standard interpretation all being provable, and those that are false all being disprovable.) If T is a theory in the language of arithmetic that is consistent, axiomatizable, and an extension of **Q**, then T is an undecidable theory by Theorem 17.4, and there exist undecidable sentences for T by Theorem 17.7. Our proof of the latter theorem did not, however, exhibit any explicit example of a sentence that is undecidable for T. An immediate question is: can we find any such specific examples?

In order to do so, we use the fact that the set of sentences that are provable and the set of sentences that are disprovable from any recursive set of axioms is semirecursive, and that all recursive sets are definable by ∃-rudimentary formulas. It follows that there are formulas $\mathrm{Prv}_T(x)$ and $\mathrm{Disprv}_T(x)$ of forms $\exists y\, \mathrm{Prf}_T(x, y)$ and $\exists y\, \mathrm{Disprf}_T(x, y)$ respectively, with Prf and Disprf rudimentary, such that $\vdash_T A$ if and only if the sentence $\mathrm{Prv}_T(\ulcorner A \urcorner)$ is correct or true in the standard interpretation, and hence if and only if for some b the sentence $\mathrm{Prf}_T(\ulcorner A \urcorner, \mathbf{b})$ is correct or—what is equivalent for rudimentary sentences—provable in **Q** and in T; and similarly for disprovability. $\mathrm{Prf}_T(x, y)$ could be read 'y is a witness to the provability of x in T'.

By the diagonalization lemma, there is a sentence G_T such that

$$\vdash_T G_T \leftrightarrow \sim\exists y\ \mathrm{Prf}_T(\ulcorner G_T \urcorner, y)$$

and a sentence R_T such that

$$\vdash_T R_T \leftrightarrow \forall y(\mathrm{Prf}_T(\ulcorner R_T \urcorner, y) \to \exists z < y\ \mathrm{Disprf}(\ulcorner R_T \urcorner, z)).$$