We now introduce the notion of the *diagonalization* of an expression A. Recall from the last chapter that $\mathbf{n}$ is the result of writing n occurrences of $'$ immediately after $\mathbf{0}$. If A is an expression with gödel number n, we define $\ulcorner A \urcorner$ to be the expression $\mathbf{n}$. In what follows $\ulcorner A \urcorner$ will be seen to behave rather like a name for the expression A. The *diagonalization* of A is the expression

$$\exists x (x = \ulcorner A \urcorner \,\&\, A).$$

If A is a formula in the language of arithmetic that contains just the variable x free, then the diagonalization of A will be a sentence that 'says that' A is true of its own gödel number – or, more precisely, the diagonalization will be true (in the standard interpretation $\mathcal{N}$) if and only if A is true (in $\mathcal{N}$) of its own gödel number.

Lemma 1

There is a recursive function, diag, such that if n is the gödel number of an expression A, diag(n) is the gödel number of the diagonalization of A.

Proof. Let lh (' length') be defined by $\mathrm{lh}(n) = \mu m(0 < m \,\&\, n < 10^m)$. Since every natural number n is less than 10^m for some positive m, and as exponentiation and less than are recursive, lh is a recursive function. $\mathrm{lh}(n)$ is the number of digits in the usual arabic numeral for n. Thus $\mathrm{lh}(1879) = 4$; $\mathrm{lh}(0) = \mathrm{lh}(9) = 1$.

Let $*$ be defined by: $m * n = m \cdot 10^{\mathrm{lh}(n)} + n$. $*$ is recursive. If $m \neq 0$, $m * n$ is the number denoted by the arabic numeral formed by writing the arabic numeral for m immediately before the arabic numeral for n. Thus $2 * 3 = 23$.

Let num be defined by: $\mathrm{num}(0) = 6$; $\mathrm{num}(n+1) = \mathrm{num}(n) * 68$ (all n). num is recursive. As $\mathrm{gn}(\mathbf{0}) = 6$ and $\mathrm{gn}(') = 68$, $\mathrm{num}(n)$ is the gödel number of $\mathbf{n}$.

As no arabic numeral for a gödel number contains the digit '0', diag(n) may be taken to $= 4515788 * (\mathrm{num}(n) * (3 * (n * 2)))$. diag is then recursive. $\exists x\, (x =$

We'll now reserve the word 'theory' for those theories whose variables are $x_0, x_1, \ldots$, and whose names, n-place function signs, sentence letters, and n-place predicate letters are some (or all) of $f_0^0, f_1^0, \ldots, f_0^n, f_1^n, \ldots, A_0^0, A_0^1, \ldots, A_0^n, A_1^n, \ldots$, respectively. As in the last chapter, we assume that $\mathbf{0}$ and $'$ occur in the language of all theories. We assume further that the set of gödel numbers of symbols of the language of the theory is recursive,

i.e. (by Church's thesis), that there is an effective procedure for deciding whether a given symbol may occur in some sentence in the language of the theory.

Here's the diagonal lemma:

Lemma 2

Let T be a theory in which diag is representable. Then for any formula $B(y)$ (of the language of T, containing just the variable y free), there is a sentence G such that

$$\vdash_T G \leftrightarrow B(\ulcorner G \urcorner).$$

Proof. Let $A(x, y)$ represent diag in T. Then for any n, k, if diag$(n) = k$, $\vdash_T \forall y (A(\mathbf{n}, y) \leftrightarrow y = \mathbf{k})$.

Let F be the expression $\exists y (A(x,y) \,\&\, B(y))$. F is a formula of the language of T that contains just the variable x free.

Let n be the gödel number of F.

Let G be the expression $\exists x (x = \mathbf{n} \,\&\, \exists y (A(x,y) \,\&\, B(y)))$. As $\mathbf{n} = \ulcorner F \urcorner$, G is the diagonalization of F and a sentence of the language of T. Since G is logically equivalent to $\exists y (A(\mathbf{n}, y) \,\&\, B(y))$, we have

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

Let k be the gödel number of G. Then

$$\mathrm{diag}(n) = k, \quad \text{and} \quad \mathbf{k} = \ulcorner G \urcorner.$$

So $\quad \vdash_T \forall y (A(\mathbf{n}, y) \leftrightarrow y = \mathbf{k})$.

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

So $\quad \vdash_T G \leftrightarrow B(\mathbf{k})$, i.e., $\vdash_T G \leftrightarrow B(\ulcorner G \urcorner)$.

A theory is called *consistent* if there is no theorem of the theory whose negation is also a theorem. Equivalently, a theory is consistent iff there is some sentence in its language that is not a theorem, iff the theory is satisfiable.

A set θ of natural numbers is said to be *definable in* theory T if there is a formula $B(x)$ of the language of T such that for any number k, if $k \in \theta$, then $\vdash_T B(\mathbf{k})$, and if $k \notin \theta$, then $\vdash_T -B(\mathbf{k})$. The formula $B(x)$ is said to define θ in T. A two-place relation R of natural numbers is likewise definable in T if there is a formula $C(x, y)$ of the language of T such that for any numbers k, n, if kRn, then $\vdash_T C(\mathbf{k}, \mathbf{n})$, and if $k \not R n$, then $\vdash_T -C(\mathbf{k}, \mathbf{n})$, and $C(x, y)$ is then said to define R in T. (A perfectly analogous definition