Exercise Sheet n°10

Exercise 1:

This exercise illustrates the idea of the Gödel first incompleteness theorem[1] in an informal fashion.

By an *expression* we mean a non empty finite string on the alphabet

$$\{\neg, P, N, (,)\}.$$

The *norm* of an expression X is defined as the expression $X(X)$. A *sentence* is any expression of one of the following four forms (where X is any expression):

(1) $P(X)$;
(2) $PN(X)$;
(3) $\neg P(X)$;
(4) $\neg PN(X)$;

We now consider a computing machine working as a printer or an enumerator on the alphabet $\{\neg, P, N, (,)\}$. We say that an expression X is *printable* if the machine can print it, that is, if sooner or later X is printed by the machine. We define a sentence of the form $P(X)$ to be *true* if (and only if) X is printable. We define $PN(X)$ to be *true* iff the norm of X is printable. Also the symbol $\neg$ stands for the negation, so that $\neg P(X)$ is said to be *true* iff X is not printable and $\neg PN(X)$ is said to be *true* iff the norm of X is not printable.

Now we suppose that the machine is completely accurate in that all sentences printed by the machine are true. For example, if the machine ever prints the sentence $P(X)$, then X is printable, i.e. X will be printed by the machine sooner or later. Also if $\neg PN(X)$ is printable, then $X(X)$ will not be printed by the machine at any time.

1. Suppose that for some expression X, X is printable. Does it necessarily follow that $P(X)$ is printable?

2. Is it possible that the machine actually prints all true sentences?

Now we turn to a variant of the previous problem. We consider a machine which prints out non empty finite strings on the alphabet $\{\neg, P, N, 1, 0\}$ called *expressions*. To each expressions we assign a number in binary notation which we call the *Gödel number* of the expression. We do this by assigning to each individual symbol $\neg, P, N, 1, 0$ the respective Gödel numbers $10, 100, 1000, 10000, 100000$. Then, the Gödel number of an expression is just the binary expression obtained by concatenation of the Gödel numbers of its symbols. For example, the Gödel number of the expression $PN10$ is 100100010000100000.

We redefine the *norm* of an expression X as the expression obtained by concatenation of X with its Gödel number. For example the norm of $PN10$ is $PN10100100010000100000$. A *sentence* is now defined as any expression of one of the four forms:

$$PX, PNX, \neg PX, \neg PNX,$$

where X is any number in binary notation. We say that a sentence of the form PX is *true* iff 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. Also, $\neg PX$ is said to be *true* exactly when PX is not true (i.e. when X is not the Gödel number of a printable expression) and $\neg PNX$ is said to be *true* exactly when PNX is not true.

Again we suppose that the printable sentences are true, that is the machine never prints false sentences.

3. Find a true sentence that the machine cannot print.

[1] This exercise is taken from R.M. Smullyan, *Gödel's Incompleteness Theorems*, Oxford University Press, 1992.

4. Find a sentence that is not printable and whose negation is not printable.

Exercise 2:

This exercise is a more formal version of the previous exercise[2].

For this exercise, we consider an 'abstract system' $\mathcal{L}$ consisting of the following items.

1. A infinite **countable** set $\mathcal{E}$ whose elements are called *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 $\mathcal{H}$ of $\mathcal{E}$ whose elements are called the *predicates* of $\mathcal{L}$;
5. A function 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.
6. A subset $\mathcal{T}$ of $\mathcal{S}$ whose elements are called the *true* sentences of $\mathcal{L}$.

We say that a predicate H is *true for* a natural number n or that n *satisfies* H, if $H(n)$ is true, i.e. $H(n) \in \mathcal{T}$. By the set *expressed* by a predicate H, we mean the set of all natural numbers n that satisfy H. Thus for any set A of natural 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 of natural numbers is* expressible *or* nameable *in $\mathcal{L}$ if A is expressed by some predicate of $\mathcal{L}$.*

Of course, since the set of expression is countable and the powerset of the natural numbers is not, uncountably many sets of natural numbers are not expressible in $\mathcal{L}$.

Definition. *The abstract system $\mathcal{L}$ is* correct *if all provable sentences are true, i.e. $\mathcal{P}$ is a subset of $\mathcal{T}$.*

We now consider a bijective function g from $\mathcal{E}$ to the natural numbers. For each n, we let E_n be the unique expression such that $g(E_n) = n$. We call $g(E)$ the *Gödel number* of the expression E.

The *diagonalisation* of the expression E_n is defined as the expression $E_n(n)$ as provided by the function described by 5. If E_n is a predicate, then its diagonalisation is a sentence. This sentence is true iff the predicate E_n is true of its own Gödel number n.

We define the *diagonal function* d from the natural numbers to the natural numbers by letting for all n, $d(n) = g(E_n(n))$, that is $d(n)$ is the Gödel number of the diagonalisation of E_n.

For a set A of natural numbers, we let $d^{-1}(A)$ be the set of all natural numbers n such that $d(n) \in A$, i.e. the inverse image of A by the diagonal function. For any subset A of the natural numbers, $A^{\complement}$ denote the complement of A in the set of natural numbers.

We define the set of the Gödel numbers of the provable sentences:

$$P = \{n \in \mathbb{N} \mid E_n \in \mathcal{P}\}.$$

Prove the following:

Theorem (After Gödel with shades of Tarski)**.** *If the set $d^{-1}(P^{\complement})$ is expressible in $\mathcal{L}$ and $\mathcal{L}$ is correct, then there is a true sentence of $\mathcal{L}$ not provable in $\mathcal{L}$.*

A *Gödel sentence* for a set A of natural numbers is defined as a sentence S such that the following condition holds

$$S \in \mathcal{T} \quad \text{iff} \quad g(S) \in A,$$

that is S is true iff $g(S)$ belongs to A.

Prove the following lemma:

Lemma (Diagonal lemma)**.** *For any set A of natural numbers, if $d^{-1}(A)$ is expressible in $\mathcal{L}$, then there is a Gödel sentence for A.*

Finally use this diagonal lemma to give another proof of the previous theorem.

[2]This exercise is taken from R.M. Smullyan, *Gödel's Incompleteness Theorems*, Oxford University Press, 1992.