



Zbigniew Tworak

## ANALOGY AND DIAGONAL ARGUMENT

**Abstract.** In this paper, I try to accomplish two goals. The first is to provide a general characterization of a method of proofs called — in mathematics — the diagonal argument. The second is to establish that analogical thinking plays an important role also in mathematical creativity. Namely, mathematical research make use of analogies regarding general strategies of proof. Some of mathematicians, for example George Polya, argued that deductions is impotent without analogy. What I want to show is that there exists a direct line leading from Cantor’s diagonal argument to constructions that underlies the proofs of several important theorems of the mathematical logic (in particular, Church’s theorem concerning the undecidability of formal arithmetic, Gödel’s theorem concerning the incompleteness of formal arithmetic, Tarski’s theorem concerning truth, and Turing’s theorem concerning the Halting Problem), and that the line could be described as an analogical mapping. In other words, Cantor’s diagonal argument and the proofs of the limitative theorems are structurally the same. Hence they can be represented as instances (or special cases) of the same general scheme.

*Keywords:* analogy, diagonal argument, antinomy, limitative theorems, provability, refutability, undecidability, truth.

### 1. Cantor’s diagonal argument

Although the diagonal procedure was invented by Paul Du Bois-Reymond (1831–1889), it found its mature expression in works of George Cantor (1845–1918) devoted to the mathematical theory of infinity. One of the starting points in Cantor’s development of the theory was his discovery that there are different degrees (or kinds) of infinity. Namely, Cantor demonstrated two theorems: first, that the real-number continuum is a *larger* infinity than



the natural numbers (i.e. is not countably infinite), and second, that every set has more subsets than elements. The proofs of the results are based on the diagonal argument.

The *diagonal argument* is the name given to class of arguments, in which so called the *diagonal method* or the *diagonalization* is applied. The essence of the diagonal method is as follows. Given an infinite list of objects of certain kind (numbers, sets, functions etc.) we have a construction which defines a new object of this kind, by systematically destroying the possibility of its identity with each object of the list. We do this by making sure that it is different in at least one place, “along the diagonal”. The new object may be said to “diagonalise out” of the list.

It should be noted that the diagonal method may and has been applied both in *direct* and *indirect* arguments. An indirect diagonal argument is used to prove a certain proposition by showing that assumption of its negation leads to contradiction. Hence it consists of two steps. The first step is a *contradictory assumption*. The second step is to employ the diagonal method to generate a contradiction. This contradiction shows that our (contradictory) assumption must be false. However, a direct diagonal argument is used to prove the existence of certain objects.

Diagonalization not only plays a crucial role both in the proofs of some important theorems, but also give rise to set-theoretical and semantical antinomies. In consequence, the diagonal arguments are sometimes divided onto two classes: (1) leading to proofs of theorems — so called the *good* diagonal arguments, and (2) leading to antinomies — so called the *bad* diagonal arguments (for example, Gumański 1990, pp. 265–270).

### 1.1. The first diagonal argument

A *set* is a collection of objects. We say that two sets,  $A$  and  $B$ , are *the same cardinal size* ( $A =^c B$ ) iff their members can be paired off one for one. In other words,  $A =^c B$  iff there is a bijection between  $A$  and  $B$ . Next, a set  $A$  is *at least as big as*  $B$  ( $A \geq^c B$ ) iff there is a subset of  $A$  which has the same size as  $B$ ; a set  $A$  has *a greater cardinality than*  $B$  ( $A >^c B$ ) iff there is a proper subset of  $A$  which has the same size as  $B$ . Clearly, there are finite sets of increasing size. Also, clearly, the set  $\mathbb{N}$  of natural numbers is larger than any finite set. A set is called *denumerable* (or *countably infinite*) if it is the same size as  $\mathbb{N}$ . In 1891, Cantor showed that there are non-denumerable sets, for example the set of all real numbers of the interval  $[0, 1]$  (the first diagonal argument). It follows that the set is bigger than  $\mathbb{N}$ .

Let  $m$  and  $w$  be two different elements, for example  $m = 0$  and  $w = 1$ . Let  $M$  be a set whose elements  $E$  are infinite sequences  $\langle a_1, a_2, \dots, a_k, \dots \rangle$ , where each of  $a_1, a_2, \dots, a_k, \dots$  is either  $m$  or  $w$ . Cantor showed that the set  $M$  is non-denumerable, i.e.

**THEOREM 1.** *No denumerable sequence of elements of the set  $M$  contain all elements of this set.*

**PROOF.** Let  $S = \langle E_1, E_2, \dots, E_j, \dots \rangle$  be any denumerable infinite sequence of members of the set  $M$ . That is  $S$  is sequence of sequences. The table below is an illustration of this:

	1	2	3	...
$E_1$	$a_{11}$	$a_{12}$	$a_{13}$	...
$E_2$	$a_{21}$	$a_{22}$	$a_{23}$	...
$E_3$	$a_{31}$	$a_{32}$	$a_{33}$	...
$\vdots$	$\vdots$	$\vdots$	$\vdots$	

Each  $a_{jk}$  is either  $m$  or  $w$ .

The proof proceeded in two steps. Step 1 is to suppose (for *reductio ad absurdum*) that the sequence  $S$  is a complete enumeration of all the members of  $M$ . Step 2 is to employ the diagonal method to generate a sequence that belongs to  $M$  but does not appear on the list  $S$ .

To this end we take into consideration the diagonal sequence  $\langle a_{11}, a_{22}, \dots, a_{kk}, \dots \rangle$  and generate a new sequence  $E_0 = \langle b_1, b_2, \dots, b_k, \dots \rangle$ , the anti-diagonal sequence, as follows: if  $a_{kk} = m$  then  $b_k = w$ , and if  $a_{kk} = w$  then  $b_k = m$ . The two tables below picturing this procedure:

The table		1	2	3	...	converts		1	2	3	...
$E_1$		$m$	$a_{12}$	$a_{13}$	...	$E_1$		$w$	$a_{12}$	$a_{13}$	...
$E_2$		$a_{21}$	$w$	$a_{23}$	...	$E_2$		$a_{21}$	$m$	$a_{23}$	...
$E_3$		$a_{31}$	$a_{32}$	$m$	...	$E_3$		$a_{31}$	$a_{32}$	$w$	...
$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$	$\vdots$	

(The table above show why this method is called *the diagonal method*: we have constructed a new sequence by changing each element on the diagonal.)

The sequence  $E_0$  has follows properties: (a) it belongs to  $M$  (because its each element is either  $m$  or  $w$ ), (b) for a natural number  $k, E_0 \neq E_k$  ( $E_0$  differs from  $E_1$  in the first digit, from  $E_2$  in the second digit, and so on), i.e.  $E_0$  does not appear on



the list  $S$ . Thus the supposition that  $S$  is complete enumeration of all the members of  $M$  is false. In consequence, the set  $M$  is non-denumerable<sup>1</sup>.  $\square$

## 1.2. The second diagonal argument

The next problem is whether there exist a set of largest size. In a generalization of his earlier proof, Cantor showed that there is not. For any set, there is bigger set. This is now called *Cantor's Theorem*. Exactly,

CANTOR'S THEOREM. *The set  $P(A)$  of all subsets of  $A$  has a greater cardinality than  $A$ .*

Equivalently, this theorem asserts that the set of all functions from  $A$  to  $\{0, 1\}$  has a greater cardinality than  $A$  itself. However, Cantor only proved a special case of this theorem (the second diagonal argument) — where  $A$  is the set of real numbers of interval  $[0, 1]$  — but he noted that the proof is quite general. This argument is given in the following famous passage ( $L = [0, 1]$  and  $M = \{0, 1\}^{[0,1]}$ ) :

$M$  does *not* however have the *same* power [cardinal size] as  $L$ . For otherwise  $M$  could be put into unique one-one correspondence to the variables  $z$  [of  $L$ ], and thus  $M$  could be thought of in the form of a single-valued function

$$\varphi(x, z)$$

of the two variables  $x$  and  $z$ , in such a way that through every specification of  $z$  one obtains an element  $f(x) = \varphi(x, z)$  of  $M$  and conversely every element  $f(x)$  of  $M$  could be obtained from  $\varphi(x, z)$  by specifying a certain  $z$ . This however leads to a contradiction. For if we understand by  $g(x)$  that single-valued function of  $x$  which takes only values 0 or 1 and which for every value of  $x$  is different from  $\varphi(x, x)$ , then on the one hand  $g(x)$  is an element of  $M$ , and on the other it cannot be obtained from  $\varphi(x, z)$  by any specification  $z = z_0$ , because  $\varphi(z_0, z_0)$  is different from  $g(z_0)$ . (Hallett 1984, p. 77)

---

<sup>1</sup>Notice that this proof may be easily converted into proof of nondenumerability of the set of real numbers. If we let  $m = 0$  and  $w = 1$ , then each  $E_k$  is the binary expansion of a real number.

Since there is not any one-to-one correspondence between  $M$  and  $L$ , and since  $M$  cannot be smaller than  $L$ , the only possible conclusion is that  $M$  must be larger than  $L$ .

REMARK. Observe that the characterization of  $g$  is impredicative:  $g(x)$  is characterized in term of  $\varphi(x, z)$ , whose definition refers to  $M$  of which  $g$  is an element. Use of the impredicative method shows some non-constructive aspect Cantor's paradise and his mathematical theory.  $\dashv$

PROOF OF CANTOR'S THEOREM. Observe that  $P(A) \geq^c A$ , for any set  $A$ , because we can pair off every member  $x$  of  $A$  one for one with the set  $\{x\}$  in  $P(A)$ . More complicated thing is to show that it is not the case that  $A =^c P(A)$ .

Step 1 is to suppose (for *reductio ad absurdum*) that  $A =^c P(A)$ , i.e. there exist a bijection  $b$  between  $A$  and  $P(A)$ .

We can think of the members of  $A$  as being (well-) ordered, thus we can arrange a list  $S^2$ :

$$S = \langle x_0, x_1, x_2, \dots, x_i, \dots \rangle.$$

Since  $b$  exists, so the corresponding sequence

$$S_b = \langle b(x_0), b(x_1), b(x_2), \dots, b(x_i), \dots \rangle$$

runs through all members of  $P(A)$ . The 0's and 1's in the body of the table below tell us whether the sets in the top row are members of the sets in the side column (we write "1" for " $x_i \in b(x_i)$ ", and "0" for " $x_i \notin b(x_i)$ "):

	$x_0$	$x_1$	$x_2$	$\dots$
$b(x_0)$	1	0	1	$\dots$
$b(x_1)$	0	0	1	$\dots$
$b(x_2)$	1	1	0	$\dots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\dots$

According to the first row,  $x_0$  and  $x_2$  are members of the set  $b(x_0)$ , however  $x_1$  is not. As a result, we can read any sequence of 0's and 1's as a code for a particular set  $b(i)(= z_i)$ . The diagonal is composed of *cells*, given by coordinates  $\langle x_0, b(x_0) \rangle, \langle x_1, b(x_1) \rangle, \langle x_2, b(x_2) \rangle, \dots$  — from top left to bottom right (other configurations of cells are possible). It determine the set  $z^d = \{x \in A : x \in b(x)\}$ .

---

<sup>2</sup> $S$  may be transfinite list and have no last member, but this is no importance.



In the next step we show that  $b$  cannot be *onto* (so  $P(A)$  must have more elements than  $A$ ).

Step 2 is to employ the diagonal method to generate a new set  $z^p$ , the anti-diagonal set, as follows:

$$z^p = \{x \in A : x \notin b(x)\}.$$

Clearly,  $z^p$  is a subset of  $A$ , and so  $z^p$  is a member of  $P(A)$ . Since  $b$  is *onto*, so there is an element  $y \in A$  such that  $z^p = b(y)$ . Consider the question whether  $y \in z^p$ :

$$\begin{aligned} y \in z^p &\text{ iff } y \in \{x \in A : x \notin b(x)\} \\ &\text{ iff } y \notin b(y) && \text{(since } y \in A) \\ &\text{ iff } y \notin z^p. \end{aligned}$$

This is a contradiction. Thus, by *reductio*,  $b$  cannot be *onto*, and ultimately  $P(A) >^c A$ .

Observe that the set  $z^p$  is formed in following way: 0's occupying cells of the diagonal we change into 1's and *vice versa* — 1's we change into 0's. It is easy to check that for every  $x_i \in A$ ,  $z^p \neq b(x_i)$ . For the illustration:  $x_0 \in b(x_0)$ , so  $x_0 \notin z^p$ , and  $x_1 \notin b(x_1)$ , so  $x_1 \in z^p$ . This ensures that  $z^p$  is different from  $b(x_0)$ , and from  $b(x_1)$ . And so on. Therefore  $z^p$  does not appear on the list  $S$ .  $\square$

REMARK. Every set  $A \subseteq U$  can be represented by characteristic function  $Ch_A: U \rightarrow \{0, 1\}$  such that for any  $x \in U$ :

$$Ch_A(x) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A. \end{cases}$$

Sets and their characteristic functions are inter-definable, i.e. if we have a set, then we have a characteristic function and if we have a characteristic function, then we can single out the set of elements where its value is 1. Thus we can transform the proof above as follows. Let  $c_x = Ch_{b(x)}$  be a function such that

$$c_x(x) = \begin{cases} 1, & \text{if } y \in b(x), \\ 0, & \text{if } y \notin b(x). \end{cases}$$

Now we define a new characteristic function  $c^p$ , the anti-diagonal function, as follows:

$$c^p(x) = 1 - c_x(x).$$

$c^p$  is the characteristic function of some subset  $w$  of  $A$ . Since  $c^p$  is different from each  $c_x$  (because  $c^p(x) = 1 - c_x(x) \neq c_x(x)$ ), so  $w$  is different from each  $b(x)$ . Of course,  $w \in P(A)$ . Thus  $b$  cannot be onto and  $P(A) >^c A$ . Moreover

$$w = \{x \in A : c^p(x) = 1\} = \{x \in A : c_x(x) = 0\} = \{x \in A : x \notin b(x)\} = z^p. \dashv$$

## 2. General characterization of the diagonal method

There is analogy between Cantor's diagonal proof that a set has a smaller cardinality than its power set and proofs some classical metamathematical theorems. In other words, they share the same general scheme by which a contradiction results from an enlargement of a set by *constructed-from-the-outside* entities.

### 2.1. The structure of diagonal arguments

As we saw in the previous, all Cantor's diagonal arguments have a common underlying structure, that produces a contradiction. The structure we can describe as follows<sup>3</sup>. Let  $X, Y, Z$  be non-empty sets and let  $\tau, \beta, \delta, \alpha$  be relations having the following properties:

- $\tau \subseteq X \times Y \times Z$  and
  - (i)  $\forall x \in X \forall y \in Y \exists z \in Z \tau(x, y, z)$ .
  - (ii)  $\forall x \in X \forall y \in Y \forall z_1, z_2 \in Z [\tau(x, y, z_1) \wedge \tau(x, y, z_2) \rightarrow z_1 = z_2]$ ,
  - (iii)  $\forall z \in Z \exists x \in X \exists y \in Y \tau(x, y, z)$ .

This is,  $\tau$  is surjective (or: is a function from  $X \times Y$  onto  $Z$ )<sup>4</sup>.

- $\beta \subseteq X \times Y$  and
  - (iv)  $\forall x \in X \exists y \in Y \beta(x, y)$ ,
  - (v)  $\forall x \in X \forall y_1, y_2 \in Y [\beta(x, y_1) \wedge \beta(x, y_2) \rightarrow y_1 = y_2]$ ,
  - (vi)  $\forall x_1, x_2 \in X \forall y \in Y [\beta(x_1, y) \wedge \beta(x_2, y) \rightarrow x_1 = x_2]$ ,
  - (vii)  $\forall y \in Y \exists x \in X \beta(x, y)$ .

This is,  $\beta$  is a bijection (or: is *one-to-one* and *onto*) between  $X$  and  $Y$ .

<sup>3</sup> Simmons (1990) gives a construction very similar to this approach.

<sup>4</sup>If  $f$  is an  $n$ -place function we usually write  $f(x_1, \dots, x_n) = y$  for  $f(x_1, \dots, x_n, y)$ .



- $\delta \subseteq X \times Y \times Z$  and

$$(viii) \quad \forall x \in X \forall y \in Y \forall z \in Z [\delta(x, y, z) \equiv \beta(x, y) \wedge \tau(x, y, z)].$$

This is,  $\delta \subset \tau$ .

- $\alpha \subseteq X \times Y \times Z$  and

$$(ix) \quad \forall x \in X \forall y \in Y \exists z \in Z \alpha(x, y, z),$$

$$(x) \quad \forall x \in X \forall y \in Y \forall z_1, z_2 \in Z [\alpha(x, y, z_1) \wedge \alpha(x, y, z_2) \rightarrow z_1 = z_2],$$

$$(xi) \quad \forall x \in X \forall y \in Y [\exists z \in Z \alpha(x, y, z) \equiv \beta(x, y)],$$

$$(xii) \quad \forall x \in X \forall y \in Y \forall z \in Z [\alpha(x, y, z) \rightarrow \neg \tau(x, y, z)].$$

This is,  $\alpha$  is a function such that for even  $x \in X$ ,  $\alpha(x, \beta(x)) \neq \tau(x, \beta(x))$ . Any function  $\delta$  that satisfies postulate (viii) will be called a *diagonaliser* (with respect to  $\tau$ ). Then  $\alpha$  to “diagonalise out” of  $\tau$ .

REMARK. The mapping  $\alpha$  is unambiguously given by the set  $\{z_{x_i y_k}\}_{(x_i y_k) \in X \times Y}$ . We say that  $\alpha$  is an *array* on  $X$  and  $Y$  (two-dimensional array) because it can be illustrated by

$$\begin{bmatrix} z_{x_1 y_1} & z_{x_1 y_2} & \cdots & z_{x_1 y_i} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ z_{x_i y_1} & z_{x_i y_2} & \cdots & z_{x_i y_i} & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}$$

The pairs  $\langle x_i, y_k \rangle \in X \times Y$  are called *coordinates*. The sequence of members of the set  $Z$  that

- are given by coordinates  $\langle x_i, y_1 \rangle, \langle x_i, y_2 \rangle, \dots, \langle x_i, y_k \rangle, \dots$  is called *i-th row*;
- are given by coordinates  $\langle x_1, y_i \rangle, \langle x_2, y_i \rangle, \dots, \langle x_k, y_i \rangle, \dots$  is called *i-th column*;
- are given by coordinates  $\langle x_1, y_1 \rangle, \langle x_2, y_2 \rangle, \dots, \langle x_i, y_i \rangle, \dots$  is called the *diagonal*. ⊣

Led  $\tau$  be an array on  $X$  and  $Y$ , where  $X$  and  $Y$  are fixed sets, and  $\varphi = \delta$  or  $\alpha$ . We say that  $\varphi$  occurs as a row of  $\tau$  iff

$$\exists w \in X \forall x \in X \forall y \in Y \forall z \in Z [\varphi(x, y, z) \rightarrow \tau(w, y, z)].$$

The next theorem states that  $\tau$  is *open* (or  $\tau$  is *not closed*).



THE DIAGONALIZATION LEMMA. Let  $\tau, \beta, \delta$  and  $\alpha$  be relations defined by postulates (i)–(xii) (see above). Then  $\alpha$  does not occur as a row of  $\tau$ , i.e.

$$\neg \exists w \in X \forall x \in X \forall y \in Y \forall z \in Z [\alpha(x, y, z) \rightarrow \tau(w, y, z)].$$

PROOF (SIMMONS 1990). We suppose for *reductio ad absurdum* that  $\alpha$  occurs as a row of  $\tau$ , i.e.

$$(1) \quad \exists w \in X \forall x \in X \forall y \in Y \forall z \in Z [\alpha(x, y, z) \rightarrow \tau(w, y, z)].$$

Taking  $a$  as an instantiation value for the initial existential quantifier in (1) we have:

$$(2) \quad \forall x \in X \forall y \in Y \forall z \in Z [\alpha(x, y, z) \rightarrow \tau(a, y, z)],$$

and taking  $a, b$  and  $c$  as instantiation values for the universal quantifiers, i.e., taking  $a$  for  $x, b$  for  $y$  and  $c$  for  $z$ , we have

$$(3) \quad \alpha(a, b, c) \rightarrow \tau(a, b, c).$$

From postulate (xii), instantiating the universal quantifiers with  $a, b$  and  $c$ , i.e., by taking  $a$  for  $x, b$  for  $y$  and  $c$  for  $z$ , we have:

$$(4) \quad \alpha(a, b, c) \rightarrow \neg \tau(a, b, c),$$

which together with (3), by standard laws, yields

$$(5) \quad \alpha(a, b, c) \rightarrow [\tau(a, b, c) \wedge \neg \tau(a, b, c)].$$

From postulate (xi), instantiating the universal quantifiers with  $a$  and  $b$ , i.e., by taking  $a$  for  $x$  and  $b$  for  $y$ , we have:

$$(6) \quad \exists z \in Z \alpha(a, b, z) \equiv \beta(a, b).$$

Next, from postulate (vii), instantiating the universal quantifiers with  $b$ , i.e., by *taking*  $b$  for  $y$ , we have:

$$(7) \quad \exists x \in X \beta(x, b),$$

and instantiating the existential quantifier with  $a$ , i.e., by *taking*  $a$  for  $x$ , we have

$$(8) \quad \beta(a, b).$$



From (6) and (8) by equivalence elimination we get

$$(9) \quad \exists z \in Z\alpha(a, b, z).$$

Taking  $c$  as an instantiation value for the existential quantifier in (9), we have:

$$(10) \quad \alpha(a, b, c).$$

Finally, from (5) and (10) by *modus ponens* we get

$$(11) \quad \tau(a, b, c) \wedge \neg\tau(a, b, c).$$

which is a contradiction. Thus, the hypothesis that  $\alpha$  occurs as a row of  $\tau$  must be false.  $\square$

Recall, the diagonal method can be applied both in *direct* and *indirect* arguments. In both cases we specify (in set-theoretical terms) an two-dimensional array  $\tau$  and a diagonal sequences  $\delta$  and  $\alpha$ . An indirect diagonal argument proves a proposition that  $\tau$  is not closed or limited (or more specifically,  $\alpha$  does not occur as a row of  $\tau$ ) by showing that the assumption of its negation leads to a contradiction. This contradiction is achieved by a proof of Diagonalization Lemma, for the appropriate interpretation. Next, the direct diagonal argument proves that  $\tau$  is not closed by showing that there exists a certain object which does not belong to  $X$ , although it has the property that puts it in  $X$ . Here, the final result is some interpretation of Diagonalization Lemma.

## 2.2. ... and its three simple interpretations

### 2.2.1. Russell's interpretation

Consider the *Russell's Antinomy*<sup>5</sup>. Call a set  $x$  *ordinary* if it is not member of itself, i.e.  $x \notin x$ . We suppose that there is a set, say  $R$ , such that for any set  $y$ ,  $y \in R$  iff  $y$  is ordinary. Assume that  $X$  and  $Y$  are each the proper class of all sets, and  $Z = \{0, 1\}$ . Then the array  $\alpha$  is given by

$$\tau(x, y) = \begin{cases} 1, & \text{if } y \in x, \\ 0, & \text{if } y \notin x. \end{cases}$$

---

<sup>5</sup>Russell's Antinomy is a stripped-down version of Cantor's Antinomy.

The function  $\beta$  is simply the identity function. The function  $\delta$  is given by

$$\delta(x, x) = \begin{cases} 1, & \text{if } \tau(x, x) = 1, \text{ i.e. } x \in x, \\ 0, & \text{if } \tau(x, x) = 0, \text{ i.e. } x \notin x, \end{cases}$$

and the function  $\alpha$  is given by

$$\alpha(x, x) = \begin{cases} 1, & \text{if } \tau(x, x) = 0, \text{ i.e. } x \notin x, \\ 0, & \text{if } \tau(x, x) = 1, \text{ i.e. } x \in x. \end{cases}$$

The supposition that the set  $R$  (of all ordinary sets) exists, i.e.  $R \in X$  and  $\alpha$  occurs as a row of  $\tau$ , generates — *via* the proof of Diagonalization Lemma — a contradiction<sup>6</sup>. In more detail, from the initial supposition, we have:

$$\tau(R, x) = \begin{cases} 1, & \text{if } \tau(x, x) = 0, \\ 0, & \text{if } \tau(x, x) = 1. \end{cases}$$

A contradiction is clearly manifested once we put  $x = R$  in the equation above:

$$\tau(R, R) = 1 \text{ iff } \tau(R, R) = 0.$$

From this equivalence, we have

$$R \in R \text{ iff } R \notin R.$$

Assuming that  $A \equiv \neg A/A \wedge \neg A$  is a valid scheme of the underlying logic, a contradiction results. This is just the Russell's Antinomy.

Since the supposition that the set  $R$  exists is arbitrary (i.e. its acceptance is only a result of our decision), we can make this assumption for *reductio ad absurdum*. By Diagonalization Lemma,  $\alpha$  does not occur as a row of  $\tau$ . Then the diagonal argument is indirect and leads to the conclusion that the set  $R$  of all ordinary sets does not exist.

### 2.2.2. Grelling's interpretation

Russell's Antinomy is about sets (or at least, about our ideas about sets). An analogous argument gives us theorem stating that the *set ordinary* formulas is not definable. Let  $M = \langle U, D \rangle$  be an interpretation for the language  $L$ . Assume that every formula of  $L$  is a member of  $U$  and every member of  $U$  is

---

<sup>6</sup> In general, the supposition that  $\alpha$  occurs as a row of  $\tau$  is *assumed* before a paradox argument.



named by some closed term of  $L$ . With respect to  $M$ , the formula  $A(x)$  of  $L$  (with just  $x$  free) defines the set  $\{t^M : t \text{ is closed term of } L \text{ and } A(t) \text{ is true in } M\}$ . We call a formula *ordinary* if it does not belong to the set that it defines (just as in the Russell's Antinomy we called a set ordinary if it does not belong to itself). For example, if  $A(x)$  is  $Formula(x)$ , then it defines the set  $A^S$  of formulas of  $L$ ; and since  $A(x) \in A^S$ , so  $A(x)$  is not ordinary. Then: (1) if  $A(x)$  is defining formula,  $A^S$  is the set that it defines; (2) a defining formula  $A(x)$  is ordinary if  $A(x) \notin A^S$ . We suppose (for *reductio ad absurdum*) that the set of ordinary formulas is definable in  $L$ , i.e. there is a formula that defines the set of all ordinary formulas. Assume that  $X$  and  $Y$  are each the set of defining formulas. Then the array is given by

$$\tau(x, y) = \begin{cases} 1, & \text{if } y \in x^S, \\ 0, & \text{if } y \notin x^S. \end{cases}$$

The function  $\beta$  is the identity function; and the function  $\alpha$  is given by

$$\alpha(x, x) = \begin{cases} 1, & \text{if } \tau(x, x) = 0, \text{ i.e. } x \notin x^S, \\ 0, & \text{if } \tau(x, x) = 1, \text{ i.e. } x \in x^S. \end{cases}$$

The supposition that the set of ordinary formulas is definable in  $L$ , i.e.  $\alpha$  occurs as a row of  $\tau$ , generates — *via* the proof of Diagonalization Lemma — a contradiction: the formula that defines the set of all ordinary formulas is ordinary itself iff it is not ordinary. But we can make this assumption for *reductio ad absurdum*. By Diagonalization Lemma,  $\alpha$  does not occur as a row of  $\tau$ . Then the diagonal argument is indirect and leads to the conclusion that there is no formula of  $L$  that defines the set of all ordinary formulas.

### 2.2.3. The catalogue

We suppose that there is a catalogue, say  $C$ , such that, for any catalogue  $y$ ,  $y$  is catalogued in  $C$  iff  $y$  is not catalogued in  $y$ . Let  $X$  and  $Y$  be each the set of catalogues, and  $Z = \{0, 1\}$ . Then the array  $\tau$  is given by

$$\tau(x, y) = \begin{cases} 1, & \text{if } y \text{ is catalogued in } x, \\ 0, & \text{if } y \text{ is not catalogued in } x. \end{cases}$$

The function  $\beta$  is simply the identity function; and the function  $\alpha$  is given by

$$\alpha(x, x) = \begin{cases} 1, & \text{if } \tau(x, x) = 0, \text{ i.e. } x \text{ is not catalogued in } x, \\ 0, & \text{if } \tau(x, x) = 1, \text{ i.e. } y \text{ is catalogued in } x. \end{cases}$$

The supposition that the catalogue  $C$  exists, i.e.  $\alpha$  occurs as a row of  $\tau$ , generates — *via* the proof of Diagonalization Lemma — the contradiction:  $C$  is catalogued in  $C$  and  $C$  is not catalogued in  $C$ . But we can make this assumption for *reductio ad absurdum*. Then by Diagonalization Lemma, we obtain the conclusion that the catalogue  $C$  does not exist.

### 3. The limitative theorems and the diagonal argument

David Hilbert, who in 1900 at the 2<sup>nd</sup> International Congress of Mathematicians (held in Paris) set the agenda for much of 20<sup>th</sup> century mathematics, asserted that mathematicians should seek to express mathematics in the form of a consistent, complete and decidable formal systems. The system is called *consistent* if no sentence is both provable and refutable in the system; and *inconsistent* otherwise. The system is called *complete* if every sentence is either provable or refutable in the system; and *incomplete* otherwise. The system is *decidable* if there is an effective method or algorithm for telling, of each mathematical sentence, whether or not the sentence is provable in the system; and *undecidable* otherwise. Hilbert's point was that if we came to possess such a formal system, then our ignorance would be banished from mathematics forever. Given any mathematical assertion, we would be able to tell whether the assertion is true or false by determining whether or not it is provable in the system. That the formal system be decidable was an important property: an undecidable system could not serve fully to banish ignorance, since we could not always be sure of being able to determine whether or not the assertion in question is provable in the system. Similarly, an incomplete system would be unsatisfactory, because the assertion in question might be true and yet unprovable in the system.

In 1931, Kurt Gödel proved that Hilbert's ideal is impossible to satisfy, even in the case of simple arithmetic. There can be no consistent and complete formal system of arithmetic. This result is known as "Gödel's first incompleteness theorem". The theorem says nothing about decidability, however. That aspect was considered by Alonzo Church and Alan Turing. In 1936, they showed independently, that no consistent formal system of arithmetic is decidable — there are well-defined mathematical problems that cannot be solved by effective methods<sup>7</sup>. Church's result is known as

---

<sup>7</sup>Turing (in the paper *On computable numbers, with an application to the Entscheidungsproblem*) took on the task of determining the exact nature of the computation. He described the architecture of a general computing device, now known as a *Turing machine*.



“undecidability theorem”, however Turing’s result — as “theorem about the unsolvability of the Halting Problem”<sup>8</sup>. In consequence, the Hilbertian dream lay in ruin.

At the same time, Alfred Tarski showed that arithmetical truth is not even arithmetically definable (this result can be generalized beyond the language of arithmetic, to any interpreted language).

In proofs of these theorems the diagonal method plays a central role. A few definitions are needed at this stage.

### 3.1. Some basic notions

#### 3.1.1. Representability

Let  $\Sigma$  be a consistent extension of Peano Arithmetic (PA) and let  $M$  be a intended interpretation of the language  $L_\Sigma$ . Each natural number  $k$  is associated with a certain term  $\bar{k}$ , called the *numeral* for  $k$ .

Let  $T$  be a subset of the set of sentences of  $L_\Sigma$  that are *true* in  $M$  (i.e. the set of *truths* of  $\Sigma$ ) and let  $P$  be a subset of the set of sentences of  $L_\Sigma$  that are *provable* in  $\Sigma$  (i.e. the set of *theorems* of  $\Sigma$ ). By  $R$  we shall denote a set of sentences of  $L_\Sigma$  called *refutable* (or *disprovable*) sentences.

A set  $S$  of natural numbers is *T-representable* (or *definable*) in  $L_\Sigma$  iff there exists a formula  $A(x)$  of the language  $L_\Sigma$  with  $x$  just free such that for any number  $k$  the following equivalence holds

$$k \in S \text{ iff } A(\bar{k}) \in T.$$

---

A Turing machine is a very simple device. It consist of an infinitely long tape with symbols (chosen from some finite set). A pointer marks the current position and the machine is in one of a finite set of internal states. At each step the machine reads the symbol at the current position on the tape. For each combination of current state and symbol read, a program specifies the new state and either a symbol to write to the tape or a direction to move the pointer (left or right) or halt. A function is computable if a Turing machine can compute it. If a Turing machine cannot be programmed to solve a particular problem, then we can conclude that the problem in question does not have a computable solution.

<sup>8</sup>Hilbert listed 23 unsolved problems of mathematics, among which the tenth problem concerned diophantine equations and was a decision problem: Specify a procedure which, in a finite number steps, enables one to determine whether or not a given polynomial diophantine equation with an arbitrary number of indeterminates and with integer coefficients has a solution in integers. A positive solution of the problem consists of furnishing the required procedure or algorithm. A negative solution consists of proving that no such procedure or algorithm can exist and in this case we declare the problem “unsolvable”. The undecidability result is here singularly important because the Hilbert’s tenth problem could be solved iff the Turing’s halting problem could be.

We say that  $S$  is  $T$ -represented in  $L_\Sigma$  by  $A(x)$  and  $A(x)$   $T$ -represents  $S$  in  $L_\Sigma$ . For example, a formula  $\exists x(x + s(\bar{0}) = y)$   $T$ -represents (defines) the set of positive integers  $\{1, 2, 3, \dots\}$ . Observe that if  $A(x)$   $T$ -represents  $S$ , then its negation  $\neg A(x)$   $T$ -represents the complement  $S'$  of  $S$ .

$T$ -representability is a semantic notion, however  $P$ -representability and  $R$ -representability are syntactic ones. A set  $S$  is  $P$ -representable (or *weakly representable*) in  $\Sigma$  iff there exists a formula  $A(x)$  of the language  $L_\Sigma$  with just  $x$  free such that for any number  $k$  the following equivalence holds

$$k \in S \text{ iff } A(\bar{k}) \in P.$$

We say that  $S$  is  $P$ -represented in  $\Sigma$  by  $A(x)$  and  $A(x)$   $P$ -represents  $S$  in  $\Sigma$ .

A set  $S$  is  $R$ -representable (or *contrarepresentable*) in  $\Sigma$  iff there exists a formula  $A(x)$  of the language  $L_\Sigma$  with just  $x$  free such that for any number  $k$  the following equivalence holds

$$k \in S \text{ iff } A(\bar{k}) \in R \text{ iff } \neg A(\bar{k}) \in P.$$

We say that  $S$  is  $R$ -represented in  $\Sigma$  by  $A(x)$  and  $A(x)$   $R$ -represents  $S$  in  $\Sigma$ . Observe that every contrarepresentable set is weakly representable.

We say that a set  $S$  is *strongly representable* in  $\Sigma$  iff there exists a formula  $A(x)$  of the language  $L_\Sigma$  that weakly represents  $S$  (i.e. for any number  $k, k \in S$  iff  $A(\bar{k}) \in P$ ) and cotrerepresents the complement  $S'$  of  $S$  (i.e. for any number  $k, k \notin S$  iff  $\neg A(\bar{k}) \in P$ ). Similarly for relations among the natural numbers.<sup>9</sup>

### 3.1.2. Computability, recursiveness, and representability

Certain functions are *calculable* or *computable* in the following sense. Suppose that we have a function  $f$  from numbers to numbers. We do not require that  $f$  be total, i.e.  $f$  may be partial<sup>10</sup>. So, informally,  $f$  would be considered computable if we had an algorithm or decision procedure such that, for each number  $k$ , gives us the value of  $f(k)$  if  $k$  is in the domain of  $f$ , and runs forever if  $k$  is not in the domain of  $f$ . Sets (of natural numbers)

---

<sup>9</sup>I omit here the definitions of these notions regarding number functions. We have connection: a set (a relation) is strongly representable (definable) in  $\Sigma$  iff its characteristic function is representable (definable) in  $\Sigma$ .

<sup>10</sup>A function  $f: A \rightarrow B$  is called *partial* if  $f$  assigns one member of  $B$  to some (but not all) members of  $A$ ; the members to which  $f$  assigns a value constitute its domain. The function is left undefined for some members of  $A$ .



and relations (among them), may — in the same sense — be computable; in particular, a set or a relation is said to be computable if its characteristic function is computable.

By *semi-decision procedure* for membership in a set (or relation) we mean an algorithm that will never give incorrect answer (though it may not answer some questions at all), and will report correctly that something is a member of the set when in fact it is. Let  $graph(f) = \{\langle n, k \rangle : n \in domain(f) \text{ and } k = f(n)\}$ . Suppose that  $f$  is intuitively computable. Then we have *semi-decision* procedure for the set  $graph(f)$ , along the following lines. Suppose that we want to test whether the pair  $\langle n, k \rangle \in graph(f)$ . Start the algorithm for  $f$  working, to compute  $f(n)$ . If it ever halts, and has  $k$  as answer, then  $\langle n, k \rangle \in graph(f)$ . So we have a method of discovering the answer is “yes” if it really is. On the other hand, if  $n$  is not in  $domain(f)$ , the algorithm will never halts, so we only have a semi-decision procedure. This works conversely as well.

Let  $S$  be a set or relation. Observe that if  $x \in S$ , and  $S$  is semi-decidable (i.e. there is a semi-decision procedure for  $S$ ), we can discover that  $x$  is a member, but if  $x \notin S$ , a semi-decision procedure will tell us nothing. A *decision procedure* should be able to answer “yes” or “no”, depending, and not just “yes”. But the notion of a decision procedure can be reduced to that of a semi-decision procedure in the following way:  $S$  has a decision procedure (i.e.  $S$  is decidable) iff it and its complement both have semi-decision procedures.

There is a developed mathematical theory of computability, founded by Alonzo Church, Alan Turing and Kurt Gödel, called *recursion theory*. This theory provides a rigorous definition of the notion of a *recursive function* that makes rigorous the intuitive notion of a computable function. Famous **Church’s Thesis** says that the computable functions are the recursive ones.

We say that a set  $S$  is *recursively enumerable* (or *semi-computable*) if there is a recursive relation  $R(x, y)$  such that for any natural number  $k$ ,

$$k \in S \text{ iff } \exists y R(k, y).$$

For example, the set of theorems of PA is recursively enumerable ( $A$  is theorem of PA iff there exists a proof of  $A$ ). Similarly for relations.

The main theorems on recursively sets (or relations) states:

- (1) A set  $S$  is recursively enumerable iff  $S$  is the domain of some partial recursive function  $f$ , and iff  $S$  is the range of some partial recursive



function  $f$ , and iff either  $S = \emptyset$  or  $S$  is the range of some total recursive function  $f$ .

- (2) Subsets of an enumerable set are enumerable.
- (3) A set  $S$  is recursive (or *computable*) iff both  $S$  and its complement  $S'$  are recursively enumerable.
- (4) Representability theorems: (a) Every recursive (computable) functions is representable in any theory  $\Sigma$  that extends PA. (b) Every recursive set is (strongly) representable in any theory  $\Sigma$  than extends PA; (c) Every recursively enumerable set is weakly representable in any theory  $\Sigma$  than extends PA.

### 3.1.3. Gödel numbering

If we have a formal theory  $\Sigma$  extends PA, then in this theory we can also talk about the language and theorems of  $\Sigma$  itself. Terms, formulas, proofs and other syntactical objects are finite sequences of symbols. Since the nature of symbols is irrelevant, these objects can be identified with finite sequences of natural numbers. Any finite sequence of natural numbers can be coded by single natural number. Thus formulas, proofs and other syntactical object can be identified with natural numbers, and properties of formulas can be identified with properties of natural numbers. For example, a *proof predicate* in  $\Sigma$  of the formula with code  $y$ . Using the proof predicate, we can define a new predicator, called a *provability predicate*:  $Theorem(y) := \exists x Proof(x, y)$ , which says that  $y$  is the code of a formula provable in  $\Sigma$  or  $y$  is the code of a theorem of  $\Sigma$ . Finally, we can interpret certain formulas in  $\Sigma$  as assertions about formulas and proofs.

Without going into details, by *Gödel numbering* we call any function  $g$  from the set of expressions of  $L_\Sigma$  into the natural numbers if and only if it meet these three conditions: (a) different Gödel numbers are assigned to different expressions; (b) it is effectively calculable what the Gödel number of any expression is; (c) it is effectively decidable whether a number is the Gödel number of some expression in the set and, if so, effectively calculable which expression it is the Gödel number of. If  $A$  is any formula, we will let  $g(A)$  be the Gödel number of  $A$ , and given any number  $k$ ,  $\bar{k}$  will be its numeral. Thus  $\bar{g(A)}$  is the numeral of the Gödel number of  $A$ . We write  $\ulcorner A \urcorner$  for  $\bar{g(A)}$ .



We say that the formula  $A(x)$  of  $L_\Sigma$   $X$ -represents a set of formulas if  $A(x)$   $X$ -represents a set of Gödel numbers of those formulas.

### 3.1.4. Diagonalization

For any formula  $A_k(x)$  containing just the variable  $x$  free ( $k$  being the Gödel number of  $A_k$ ) by its *diagonalization* we mean the sentence  $A_k(x/\bar{k})$ . Intuitively, the sentence is supposed to express the proposition that the Gödel number of  $A_k(x)$  lies in the set represented (defined) by  $A_k(x)$ . We write  $d(k)$  for the Gödel number of  $A_k(x/\bar{k})$ .

Since  $A(a) \equiv \exists x(x = a \wedge A(x))$  is a theorem of first-order logic, we can also introduce the notion of the diagonalization of a formula  $A_k(x)$ , as follows: the diagonalization of  $A_k(x)$  is the sentence  $\exists x(x = \bar{k} \wedge A_k(x))$ . Intuitively, if  $A_k$  is a formula with just  $x$  free, the diagonalization of  $A_k$  says that  $A_k$  is true of its own Gödel number.

We say that a formula  $A_k(x)$  of the language  $L_\Sigma$  is *T-ordinary* iff either: (a)  $A_k(x/\bar{k}) \notin T$ , (b)  $\neg A_k(x/\bar{k}) \in T$ . Equivalently: iff  $k$  is not in the set  $T$ -represented (defined) by  $A_k$ . We can replace truth (in given interpretation) by provability (in given theory). But then (a) and (b) give us two alternatives, that need not be equivalent. Namely, (c) a formula  $A_k(x)$  is *P-ordinary* iff  $A_k(x/\bar{k}) \notin P$  (equivalently: iff  $k$  is not in the set  $P$ -represented in  $\Sigma$  by  $A_k$ ), (d) a formula  $A_k(x)$  is *R-ordinary* iff  $\neg A_k(x/\bar{k}) \in P$  (equivalently: iff  $k$  is in the set  $R$ -represented in  $\Sigma$  by  $A_k$ ). It is easy to see that if  $\Sigma$  is consistent, then any formula that is *R-ordinary* must also be *P-ordinary*.

## 3.2. The limitative theorems

Assume now that

- (i)  $\Sigma$  is a consistent extension of the Peano Arithmetic (or more generally,  $\Sigma$  is a consistent theory which can represent all recursive functions), and
- (ii) the set of (Gödel numbers of) theorems of  $\Sigma$  is recursively enumerable (so, there is a formula of  $L_\Sigma$  that weakly represents it in  $\Sigma$ ).

### 3.2.1. Church's undecidability theorem

CHURCH'S UNDECIDABILITY THEOREM.  $\Sigma$  is undecidable, i.e. there is no effective procedure for deciding whether or not a given sentence is theorem

in  $\Sigma$ .<sup>11</sup>

In more general form, the theorem says that if  $\Sigma$  is any sufficiently strong formal theory in which every recursively enumerable set is weakly representable, then the set  $P$  of (Gödel numbers of) theorems of  $\Sigma$  is recursively enumerable but not recursive<sup>12</sup>.

PROOF. Suppose otherwise. Let  $\Sigma$  be a consistent formal theory in which every recursively enumerable set is weakly representable, and suppose that the set  $g(P)$  of Gödel numbers of theorems (provable sentences) of  $\Sigma$  is recursive. By  $S_k$  we denote the set weakly represented by  $A_k(x)$ . Under assumption that the set  $g(P)$  is recursive, the (sub)set  $D = \{k : A_k(\bar{k}) \in P\} \subseteq g(P)$  also is recursive. Hence  $D$  and its complement  $D' = \{k : A_k(\bar{k}) \notin P\}$  are recursively enumerable, and — from representability theorem — they are weakly representable in  $\Sigma$ . The proof that  $D'$  is not weakly representable in  $\Sigma$ , and thus is not recursively enumerable, is a diagonal argument.

Let  $X$  be the set of all 1-place weakly representing ( $P$ -representing) formulas of  $L_\Sigma$ , and let  $Y$  be the set of natural numbers. Define the array  $\tau$  by

$$\tau(A_k, n) = \begin{cases} 1, & \text{if } A_k(x/\bar{n}) \in P, \text{ i.e. } n \in S_k, \\ 0, & \text{if } A_k(x/\bar{n}) \notin P, \text{ i.e. } n \notin S_k. \end{cases}$$

Let  $\beta$  be a Gödel numbering (a coding function) with the domain  $X$ .<sup>13</sup> Then  $\delta$  is given by

$$\delta(A_k, k) = \begin{cases} 1, & \text{if } \tau(A_k, k) = 1, \text{ i.e. } A_k(x/\bar{k}) \in P, \text{ i.e. } k \in S_k, \\ 0, & \text{if } \tau(A_k, k) = 0, \text{ i.e. } A_k(x/\bar{k}) \notin P, \text{ i.e. } k \notin S_k, \end{cases}$$

and  $\alpha$  is given by

$$\alpha(A_k, k) = \begin{cases} 1, & \text{if } \tau(A_k, k) = 0, \text{ i.e. } A_k(x/\bar{k}) \notin P, \text{ i.e. } k \notin S_k, \\ 0, & \text{if } \tau(A_k, k) = 1, \text{ i.e. } A_k(x/\bar{k}) \in P, \text{ i.e. } k \in S_k. \end{cases}$$

By Diagonalization Lemma,  $\alpha$  does not occur as a row of  $\tau$ , i.e. there is no a formula of  $L_\Sigma$ , that weakly represents  $D'$  in  $\Sigma$ . Therefore,  $D'$  cannot be recursively enumerable, so that  $D$  and hence  $g(P)$  cannot be recursive.  $\square$

---

<sup>11</sup>The God of mathematics must love mathematicians, in fact all who can think logically and with imagination, because he made no mechanical substitute for them.

<sup>12</sup>The theory is called *decidable* if its set of theorems is recursive; and *undecidable* otherwise.

<sup>13</sup>Alternatively, both  $X$  and  $Y$  are a set of Gödel numbers of 1-place weakly representing formulas, and  $\beta$  is the identity function.



Since  $D'$  is the set of Gödel numbers of all  $P$ -ordinary formulas of  $L_\Sigma$ , the following theorem holds:

**THE GRELLING'S (HETEROLOGICAL) PARADOX. THE FIRST VERSION.** *Let  $\Sigma$  be conservative extension of PA. If there exists a formula of  $L_\Sigma$  that weakly represents the set of (Gödel numbers of)  $P$ -ordinary formulas of  $L_\Sigma$  in  $\Sigma$ , then  $\Sigma$  is inconsistent<sup>14</sup>.*

**PROOF.** The proof of this theorem is the appropriate interpretation the proof of Diagonalization Lemma. We claim that there is a formula of  $L_\Sigma$  with one variable, say  $Ordinary(x)$ , that weakly represents  $D'$  in  $\Sigma$ . Thus  $Ordinary(x) \in X$  (= the set of all 1-place representing formulas of  $L_\Sigma$ ), and  $\alpha$  occurs as a row of  $\tau$ . But then

$$\tau(Ordinary(x), k) = \begin{cases} 1, & \text{if } \tau(A_k, k) = 0, \text{ i.e. } A_k(x/\bar{k}) \notin P, \\ 0, & \text{if } \tau(A_k, k) = 1, \text{ i.e. } A_k(x/\bar{k}) \in P. \end{cases}$$

From the equation above, by taking  $g(Ordinary(x))$  for  $k$ , we have:

$$\tau(Ordinary(x), g(Ordinary(x))) = 1 \text{ iff } \tau(Ordinary(x), g(Ordinary(x))) = 0,$$

which is a contradiction. This contradiction can be expressed by the following equivalence:

$$Ordinary(\ulcorner Ordinary \urcorner) \in P \text{ iff } Ordinary(\ulcorner Ordinary \urcorner) \notin P.$$

The last equivalence is a certain version of the heterological paradox.  $\square$

### 3.2.2. The unsolvability of the Halting Problem (for Turing machines)

The Halting Problem is a decision problem which can be informally stated as follows.

Recall, one of Hilbert's requirements for formal mathematical systems was that there be an objective criterion for deciding whether a proof written in the language of the system is valid or not. In other words, there must be an algorithm or a program for a Turing machine, for checking proofs. Using both Cantor's diagonal argument and coding techniques similar to Gödel's, Alan Turing showed that there is no algorithm for deciding whether or not a program ever halts. In more details, the Halting Problem concerns a

---

<sup>14</sup>In more general form, the theorem says that if the set  $P$  of all theorems of  $\Sigma$  (or the set  $g(P)$ ) is recursive, then  $\Sigma$  is inconsistent.

Turing machine program  $h$  which takes one argument and which satisfies the following specification:  $h(n)$  halts iff it has constructed a correct proof that the one-argument program defined by  $n$  does not halt when presented with data  $n$ . The argument (input)  $n$  is the source code for another program, and  $h$  concerns itself with proving properties of this second program. Turing's halting theorem is obtained when  $h$  is asked to prove a theorem about itself, i.e., when  $n = g(h)$ . There is a contradiction unless  $h(g(h))$  does not halt; this result is just the Turing's halting theorem. In consequence, the Halting Problem is unsolvable.

Armed with the general definition of a formal axiomatic system  $\Sigma$  as an recursively enumerable set of provably sentences of  $L_\Sigma$ , one can immediately obtain (a version of) Gödel's first incompleteness theorem from Turing's theorem (which, historically, came second).

The Halting Problem for Turing machines is a manifestation of undecidability because a Turing computation is equivalent to the computation of a partial recursive function, which is only defined for a subset of the set of natural numbers. If the set is undecidable, one cannot always tell in advance whether the Turing machine program will halt (i.e., whether the input is in the domain of the partial recursive function) or not (when the input is not in the domain). Hence, we can express the unsolvability of the Halting Problem as follows.

1. Let  $f_k$  be a (possible) partial recursive function with a Gödel number  $k$  (or a Turing machine program with a single natural input and a single natural output). For any natural number  $n$ , if  $f_k(n)$  is defined (the program halts), we write  $f_k(n) \downarrow$ , however if  $f_k(n)$  is undefined (the program does not halt), we write  $f_k(n) \uparrow$ .
2. Then the set  $S_k = \{n : n \in \text{domain}(f_k)\} = \{n : f_k(n) \downarrow\}$ , where  $f_k$  is a partial recursive function, is a recursively enumerable set.
3. The Turing's halting theorem concerns the set  $H = \{k : f_k(k) \downarrow\} = \{k : k \in S_k\}$  (referred to as the "Halting Problem"). The Halting Problem asks: Is the set  $H$  recursive? If the set  $H$  is recursive, then  $H$  and its complement  $H' = \{k : f_k(k) \uparrow\} = \{k : k \notin S_k\}$  are recursively enumerable. Hence, the Halting Problem can be expressed by a question, whether there exist a partial recursive function (a Turing machine program), say  $h$ , such that  $H' = \text{domain}(h)$ .

**TURING'S HALTING THEOREM.** *The set  $H$  is recursively enumerable but not recursive.*



PROOF. Suppose that the set  $H$  is recursive. Let  $X$  and  $Y$  be the set of Gödel numbers of all recursively enumerable sets (alternatively: of all partial recursive function). Define the array  $\tau$  by

$$\tau(k, n) = \begin{cases} 1, & \text{if } n \in S_k \text{ i.e. } f_k(n) \downarrow, \\ 0, & \text{if } k \notin S_k \text{ i.e. } f_k(n) \uparrow, \end{cases}$$

Let  $\beta$  be the identity function. Then  $\delta$  is given by

$$\delta(k, k) = \begin{cases} 1, & \text{if } \tau(k, k) = 1, \text{ i.e. } k \in S_k \text{ i.e. } f_k(k) \downarrow, \\ 0, & \text{if } \tau(k, k) = 0, \text{ i.e. } k \notin S_k \text{ i.e. } f_k(k) \uparrow, \end{cases}$$

and  $\alpha$  is given by

$$\alpha(k, k) = \begin{cases} 1, & \text{if } \tau(k, k) = 0, \text{ i.e. } k \notin S_k \text{ i.e. } f_k(k) \uparrow, \\ 0, & \text{if } \tau(k, k) = 1, \text{ i.e. } k \in S_k \text{ i.e. } f_k(k) \downarrow. \end{cases}$$

Observe that if  $H$  is recursive, then  $H$  and its complement  $H'$  are recursively enumerable. Hence both  $\delta$  and  $\alpha$  occur as a row of  $\tau$ . By Diagonalization Lemma,  $\alpha$  does not occur as a row of  $\tau$ . Contradiction. Therefore,  $H'$  cannot be recursively enumerable, so  $H$  cannot be recursive.

Suppose that  $H'$  is recursively enumerable. Then  $h = g(H')$  is a member  $X$ , and

$$\tau(h, k) = \begin{cases} 1, & \text{if } \tau(k, k) = 0, \text{ i.e. } k \notin S_k, \\ 0, & \text{if } \tau(k, k) = 1, \text{ i.e. } k \in S_k. \end{cases}$$

When we put  $k = h$  in the equation above, we obtain

$$\tau(h, h) = 1 \text{ iff } \tau(h, h) = 0.$$

Since the equivalence can be expressed by

$$h \notin H' \text{ iff } h \in H',$$

it is a certain version of the Russell's Antinomy (because  $h = g(H')$ ).  $\square$

### 3.2.3. Tarski's undefinability theorem for truth

TARSKI'S THEOREM. *The set  $T$  of sentences of  $L_\Sigma$  that are true in the intended interpretation  $M$  of  $L_\Sigma$  (or the set  $g(T)$ ) is not definable in  $L_\Sigma$ .*

PROOF. Suppose otherwise. Let  $\Sigma$  be a consistent extension of the Peano Arithmetic, and suppose that the set of (Gödel numbers of) sentences of  $L_\Sigma$  that are true in the intended interpretation of  $L_\Sigma$  is definable in  $L_\Sigma$ . That is: there is a formula of  $L_\Sigma$ , say  $True(x)$ , that defines the set  $g(T)$  of Gödel numbers of all true sentences (we note that if  $True(x)$  defines  $g(T)$ , then  $k \in g(T)$  iff  $True(\bar{k}) \in T$ ). Now, by  $S_k$  we denote the set defined by  $A_k(x)$ .

Let  $X$  be the set of 1-place defining ( $T$ -representing) formulas of  $L_\Sigma$ , and let  $Y$  be the set of natural numbers. Define the array  $\tau$  by

$$\tau(A_k, n) = \begin{cases} 1, & \text{if } A_k(x/\bar{n}) \in T, \text{ i.e. } n \in S_k, \\ 0, & \text{if } A_k(x/\bar{n}) \notin T, \text{ i.e. } n \notin S_k. \end{cases}$$

Let  $\beta$  be a Gödel numbering with the domain  $X$ <sup>15</sup>. Then  $\delta$  is given by

$$\delta(A_k, k) = \begin{cases} 1, & \text{if } \tau(A_k, k) = 1, \text{ i.e. } A_k(x/\bar{k}) \in T, \text{ i.e. } k \in S_k, \\ 0, & \text{if } \tau(A_k, k) = 0, \text{ i.e. } A_k(x/\bar{k}) \notin T, \text{ i.e. } k \notin S_k, \end{cases}$$

and  $\alpha$  is given by

$$\alpha(A_k, k) = \begin{cases} 1, & \text{if } \tau(A_k, k) = 0, \text{ i.e. } A_k(x/\bar{k}) \notin T, \text{ i.e. } k \notin S_k, \\ 0, & \text{if } \tau(A_k, k) = 1, \text{ i.e. } A_k(x/\bar{k}) \in T, \text{ i.e. } k \in S_k. \end{cases}$$

Under the assumption then the set  $T$  (or  $g(T)$ ) is definable in  $L_\Sigma$ , the set  $\{k : \alpha(A_k, k) = 1\} = \{k : A_k(x/\bar{k}) \notin T\}$  of Gödel numbers of all ordinary formulas of  $L_\Sigma$  is also definable in  $L_\Sigma$ , and then  $\alpha$  occurs as a row of  $\tau$ , i.e. there exists a formula of  $L_\Sigma$ , that is true (in standard interpretation of  $L_\Sigma$ ) of exactly the Gödel numbers of those formulas that are not true (in standard interpretation of  $L_\Sigma$ ) of the own Gödel numbers. By Diagonalization Lemma,  $\alpha$  does not occur as a row of  $\tau$ , i.e. there is no a formula of  $L_\Sigma$ , that defines the set of Gödel numbers of all ordinary formulas of  $L_\Sigma$ . Contradiction. Therefore, the set cannot be definable, and hence  $g(T)$  cannot be definable.  $\square$

Observe that following theorem holds:

**THE GRELLING'S (HETEROLOGICAL) PARADOX. THE SECOND VERSION.** *Let  $\Sigma$  be extension of PA. If there exists a formula of  $L_\Sigma$  that defines in  $L_\Sigma$  the set of Gödel numbers of all ordinary formulas of  $L_\Sigma$ , then  $\Sigma$  is inconsistent<sup>16</sup>.*

<sup>15</sup>Alternatively, both  $X$  and  $Y$  are a set of Gödel numbers of 1-place defining formulas, and  $\beta$  is the identity function.

<sup>16</sup>In more general form, the theorem says that if there exists a formula  $True(x)$  of language  $L_\Sigma$  that defines in  $L_\Sigma$  the set of (Gödel numbers of) all true sentences of  $L_\Sigma$ , then  $\Sigma$  is inconsistent.



PROOF. The proof of this theorem is the appropriate interpretation of Diagonalization Lemma and is quite similar to that of first version of the theorem. We claim that there is a formula of  $L_\Sigma$ , say  $Ordinary(x)$ , defines the set of Gödel numbers of ordinary formulas of  $L_\Sigma$ . Thus  $Ordinary(x) \in X$  (= the set of all 1-place defining formulas of  $L_\Sigma$ ), and  $\alpha$  occurs as a row of  $\tau$ . But then

$$\tau(Ordinary(x), k) = \begin{cases} 1, & \text{if } \tau(A_k, k) = 0, \text{ i.e. } A_k(x/\bar{k}) \notin T, \\ 0, & \text{if } \tau(A_k, k) = 1, \text{ i.e. } A_k(x/\bar{k}) \in T. \end{cases}$$

From the equation above, by taking  $g(Ordinary(x))$  for  $k$ , we have:

$$\tau(Ordinary(x), g(Ordinary(x))) = 1 \text{ iff } \tau(Ordinary(x), g(Ordinary(x))) = 0,$$

which is a contradiction. In other words,

$$Ordinary(\ulcorner Ordinary \urcorner) \notin T \text{ iff } Ordinary(\ulcorner Ordinary \urcorner) \in T.$$

The last equivalence is somewhat other version of the heterological paradox.  $\square$

### 3.2.4. Gödel's first incompleteness theorem

GÖDEL'S FIRST INCOMPLETENESS THEOREM.  $\Sigma$  is incomplete. More specifically: there is a sentence  $A$  of the language  $L_\Sigma$  such that  $A$  is not provable in  $\Sigma$  and  $A$  is not refutable in  $\Sigma$ .

Here we give four proofs of the theorem which employ results above.

**a) The proof using Church's result.** The proof is based on Church's undecidability theorem and the following lemma:

LEMMA A. Let  $\Sigma$  be a formal theory that meets conditions (i) and (ii). Then, if  $\Sigma$  is undecidable (which means that the set of theorems of  $\Sigma$  is not recursive), then  $\Sigma$  is incomplete.

**b) The proof using Turing's result.** The proof is based on Turing's halting theorem and the following lemma:

LEMMA B. If the set  $H$  is weakly representable in  $\Sigma$  ( $H$  is recursive enumerable but is not recursive), then  $\Sigma$  is undecidable, well also  $\Sigma$  is incomplete.

REMARK. Emil Post gave the modern and abstract versions of Gödel's theorem, which looks like Japanese haiku: there exists a recursively enumerable set of natural numbers, that is not recursive (for example, the set of all theorems of  $\Sigma$  or the set  $H$ ).  $\dashv$



**c) The proof using provability and refutability.** Gödel's original method of proving incompleteness use provability and refutability. To do this, however, Gödel had to make a certain assumption known as  $\omega$ -consistency, which we will not discuss here<sup>17</sup>. We say only that under the assumption that  $\Sigma$  is  $\omega$ -consistency, the sets  $\{k : A_k(\bar{k}) \in R\} \subseteq g(R)$  and  $\{k : A_k(\bar{k}) \in P\} \subseteq g(P)$  are weakly representable in  $\Sigma$ .

LEMMA C. THE FIRST VERSION. *Let  $\Sigma$  be conservative extension of PA. If the set  $\{k : A_k(\bar{k}) \in R\}$  (of Gödel numbers of all  $R$ -ordinary formulas) is weakly representable in  $\Sigma$ , then  $\Sigma$  is either inconsistent or incomplete. Differently: if  $\Sigma$  is consistent and the set  $\{k : A_k(\bar{k}) \in R\}$  is weakly representable in  $\Sigma$ , then  $\Sigma$  is incomplete.*

PROOF. (Smullyan 1994, p. 57). Suppose that the set  $\{k : A_k(\bar{k}) \in R\}$  (of all  $R$ -ordinary formulas) is weakly representable in  $\Sigma$ , say by  $Ordinary(x)$ . This means that, if  $A_k(x)$  is any weakly representing formula of  $L_\Sigma$  (with Gödel numbers  $k$ ), then

$$Ordinary(\bar{k}) \in P \text{ iff } A_k(\bar{k}) \in R,$$

that is,

$$Ordinary(\bar{k}) \in P \text{ iff } \neg A_k(\bar{k}) \in P,$$

Now to show that  $\Sigma$  is either inconsistent or incomplete, we need only the formula  $Ordinary(x)$  for  $A_k(x)$ . Then we get the following equivalence

$$Ordinary(\ulcorner Ordinary \urcorner) \in P \text{ iff } \neg Ordinary(\ulcorner Ordinary \urcorner) \in P.$$

There are two possible ways of understanding the equivalence. First, the sentence  $Ordinary(\ulcorner Ordinary \urcorner)$  is both provable and refutable:

$$Ordinary(\ulcorner Ordinary \urcorner) \in P \text{ and } \neg Ordinary(\ulcorner Ordinary \urcorner) \in P.$$

Second, the sentence  $Ordinary(\ulcorner Ordinary \urcorner)$  is neither provable nor refutable:

$$Ordinary(\ulcorner Ordinary \urcorner) \notin P \text{ and } \neg Ordinary(\ulcorner Ordinary \urcorner) \notin P.$$

---

<sup>17</sup>Let  $\Sigma$  be a theory in the language of Peano Arithmetic. It is called  $\omega$ -inconsistent if there is a formula  $A(x)$  (with just  $x$  free) such that the sentence  $\exists x A(x)$  is provable in  $\Sigma$ , yet all the infinitely many sentences  $A(\bar{0}), \dots, A(\bar{n}), \dots$  are refutable in  $\Sigma$ . The theory  $\Sigma$  is  $\omega$ -consistent if it is not  $\omega$ -inconsistent. Notice that the notion of  $\omega$ -consistency refers only to proofs. It allows us to replace semantical arguments by ones that only involve the notion of provability, which has constructive content. Any  $\omega$ -consistent theory is consistent in the ordinary sense, since everything is provable in an inconsistent theory.



If  $\Sigma$  is consistent, then the first alternative cannot hold, hence  $\Sigma$  is incomplete.  $\square$

LEMMA C. THE SECOND VERSION. *Let  $\Sigma$  be conservative extension of PA. If the set  $\{k : A_k(\bar{k}) \in P\}$  (of all formulas that are not  $P$ -ordinary) is contrarepresentable in  $\Sigma$ , then  $\Sigma$  is either inconsistent or incomplete. Differently: if  $\Sigma$  is consistent and the set is contrarepresentable in  $\Sigma$ , then  $\Sigma$  is incomplete.*

PROOF. (Smullyan 1994, p. 57) The proof of this lemma is quite similar to that of first version. Suppose that the set (of all formulas that are not  $P$ -ordinary) is contrarepresentable in  $\Sigma$  by  $Ordinary(x)$ . This means that, if  $A_k(x)$  is any weakly representing formula of  $L_\Sigma$  (with Gödel numbers  $k$ ), then

$$Ordinary(\bar{k}) \in R \text{ iff } A_k(\bar{k}) \in P,$$

that is,

$$\neg Ordinary(\bar{k}) \in P \text{ iff } A_k(\bar{k}) \in P,$$

Take  $A_k(x)$  to be  $Ordinary(x)$  itself. Then we have

$$\neg Ordinary(\ulcorner Ordinary \urcorner) \in P \text{ iff } Ordinary(\ulcorner Ordinary \urcorner) \in P. \quad \square$$

**d) The proof using Tarski's result.** Now we give a version of the incompleteness argument using Tarski's result — replacing the assumption that  $\Sigma$  is  $\omega$ -consistent by the (strongly) assumption then  $\Sigma$  is correct. The system is called *correct* if every provable sentence is true and no refutable (or disprovable) sentence is true. Then  $g(P) \subseteq g(T)$ . If  $\Sigma$  is correct, then the set  $g(P)$  of Gödel numbers of provable sentences is definable in  $L_\Sigma$ . Thus  $g(P)$  is definable in  $L_\Sigma$  but  $g(T)$  is not (Tarski's theorem). Hence these two sets are different, i.e.  $g(P) \neq g(T)$ . Thus, for some Gödel number  $k, k \in g(T)$  and  $k \notin g(P)$ . This means that there is a sentence that is true in standard interpretation of  $L_\Sigma$  but not provable in  $\Sigma$ . Hence  $\Sigma$  is incomplete. More specifically:

LEMMA D. *If  $\Sigma$  is correct and the set of Gödel numbers of  $P$ -ordinary formulas is  $T$ -representable (definable) in  $L_\Sigma$ , then  $\Sigma$  is incomplete.*

PROOF. (Smullyan 1994, p. 53). Let  $\Sigma$  be correct, and suppose that the set of  $P$ -ordinary formulas is definable in  $L_\Sigma$  by  $Ordinary(x)$ . This means that, if  $A_k(x)$  is any weakly representing formula of  $L_\Sigma$ , then

$$Ordinary(\bar{k}) \in T \text{ iff } A_k(\bar{k}) \notin P.$$

Now to show that  $\Sigma$  is incomplete, we need only the formula  $Ordinary(x)$  for  $A_k(x)$ . Then we get the following equivalence

$$Ordinary(\ulcorner Ordinary \urcorner) \in T \text{ iff } Ordinary(\ulcorner Ordinary \urcorner) \notin P.$$

There are two possible ways of understanding the equivalence. First, the sentence  $Ordinary(\ulcorner Ordinary \urcorner)$  is true but is not provable:

$$Ordinary(\ulcorner Ordinary \urcorner) \in T \text{ and } Ordinary(\ulcorner Ordinary \urcorner) \notin P.$$

Second, the sentence  $Ordinary(\ulcorner Ordinary \urcorner)$  is provable but is not true:

$$Ordinary(\ulcorner Ordinary \urcorner) \notin T \text{ and } Ordinary(\ulcorner Ordinary \urcorner) \in P.$$

If  $\Sigma$  is correct, then the second alternative cannot hold. Since  $T$  is disjoint from the set  $R$  of refutable sentences, we have also:

$$Ordinary(\ulcorner Ordinary \urcorner) \in T \text{ and } Ordinary(\ulcorner Ordinary \urcorner) \notin R.$$

That is, the sentence  $Ordinary(\ulcorner Ordinary \urcorner)$  is true but is not refutable. Thus

$$Ordinary(\ulcorner Ordinary \urcorner) \notin P \text{ and } Ordinary(\ulcorner Ordinary \urcorner) \notin R.$$

Hence  $Ordinary(\ulcorner Ordinary \urcorner)$  is undecidable sentence in  $\Sigma$ , and  $\Sigma$  is incomplete.  $\square$

I think that the results presented here are sufficient to show that analogical thinking plays an important role also in mathematical creativity.

## References

- [1] Cantor, G., "Über eine elementare Frege der Mannigfaltigkeitslehre", *Jahresbericht der Mathematiker-Vereinigung* 1 (1890–91), 75–78, reprinted in: E. Zermelo (ed.), *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, Berlin, J. Springer 1932, pp. 278–281.
- [2] Church, A., "An unsolvable problem of elementary number theory", *American Journal of Mathematics* 58 (1936), 345–363, reprinted in: M. Davis (ed.), *The Undecidable*, Raven Press, New York 1965, pp. 89–107.
- [3] Gödel, K., "Über formal unentscheidbare Sätze der 'Principia mathematica' und verwandter Systeme I", *Monatshefte für Mathematik und Physik* 38 (1931), 173–198; translated in: J. van Heijenoort (ed.), *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, Harvard University Press, Cambridge, MA, 1967, pp. 596–616.



- [4] Gumański, L., “Remarks on Cantor’s Diagonal Method and Some Related Topics”, *Ruch Filozoficzny* LVII, 3/4 (2000), 437–448.
- [5] Hallett, M., *Cantorian Set Theory and Limitation of Size*, Clarendon Press, Oxford, 1984.
- [6] Krajewski, S., *Twierdzenie Gödla i jego interpretacje filozoficzne. Od mechanicyzmu do postmodernizmu*, Wydawnictwo IFiS PAN, Warszawa, 2003.
- [7] Martin, R. L., “On a Puzzling Classical Validity”, *Philosophical Review* LXXXVI, 4, (1977), 454–473.
- [8] Murawski, R., *Filozofia matematyki. Zarys dziejów*, PWN, Warszawa, 1995.
- [9] Murawski, R., *Recursive Functions and Metamathematics*, Synthese Library vol. 286, Kluwer Academic Publishers, Dordrecht – Boston – London, 1999.
- [10] Russell, B., *The Principles of Mathematics*, Cambridge University Press, Cambridge, 1903.
- [11] Simmons, K., “The diagonal argument and the Liar”, *Journal of Philosophical Logic* 19 (1990), 277–303.
- [12] Smullyan, R., *Diagonalization and Self-Reference*, Clarendon Press, Oxford, 1994.
- [13] Tarski, A., *Pojęcie prawdy w językach nauk dedukcyjnych*, Warszawa, 1933; translated in: *Logic, semantics, metamathematics. Papers from 1923 to 1938*, Clarendon Press, Oxford 1956, pp. 152–277.
- [14] Turing, A., “On computable numbers, with an application to Entscheidungsproblem”, *Proceedings of the London Mathematical Society* Ser. 2, Vol. 42 (1937), 230–265; reprinted in: M. Davis (ed.), *The Undecidable*, Raven Press, New York 1965, pp. 116–154.
- [15] Woleński, J., *Metamatematyka a epistemologia*, PWN, Warszawa, 1993.

ZBIGNIEW TWORAK  
Instytut Filozofii UAM  
Zakład Logiki i Metodologii Nauk  
ul. Szamarzewskiego 89c  
60-569 Poznań  
Poland  
tworak@amu.edu.pl