2017 ACTA UNIVERSITATIS CAROLINAE PAG. 33–43
PHILOSOPHICA ET HISTORICA 2 / MISCELLANEA LOGICA

DIAGONAL ARGUMENTS

JAROSLAV PEREGRIN
Institute of Philosophy, Czech Academy of Science and Faculty of Philosophy, University of Hradec Králové
E-mail: jaroslav.peregrin@uhk.cz

ABSTRACT
It is a trivial fact that if we have a square table filled with numbers, we can always form a column which is not yet contained in the table. Despite its apparent triviality, this fact can lead us the most of the path-breaking results of logic in the second half of the nineteenth and the first half of the twentieth century. We explain how this fact can be used to show that there are more sequences of natural numbers than there are natural numbers, that there are more real numbers than natural numbers and that every set has more subsets than elements (all results due to Cantor); we indicate how this fact can be seen as underlying the celebrated Russell's paradox; and we show how it can be employed to expose the most fundamental result of mathematical logic of the twentieth century, Gödel's incompleteness theorem. Finally, we show how this fact yields the unsolvability of the halting problem for Turing machines.
Keywords: diagonalization, cardinality, Russell's paradox, incompleteness of arithmetic, halting problem

1. General formulation

Imagine a square table populated by natural numbers.

5	0	3
1	1	3
3	7	3

Is it possible to add a column that the table does not yet contain? There are, of course, many such columns that could be added. Now suppose that the table is populated only by zeros and ones.

0	0	1
1	1	0
1	0	0

https://doi.org/10.14712/24647055.2017.14

Is it still possible to add a new column? Of course, it is – for example 0, 0, 0; or 1, 1, 1. Now suppose that the table is very large. Can we still do the same? It seems that the answer must still be positive, though now it might be not so easy to find a new column. Here is an easy method: make the first number of the new column different from that in the first row of the first column of the original table, make the second number in the new column different from that in the second row of the second column, etc. Hence, the number in the *n*th row of the new column is different from that in the *n*th row of the *n*th column and the new column is thus different from each column of the original table.

0	0	1	**1**
1	1	0	**0**
1	0	0	**1**

We can call this method diagonalization, and we can call the column produced by diagonalization the *antidiagonal* of the table (we will abreviate it to AD). Note that we can speak about *the* antidiagonal only thanks to the fact that the table we talk about cannot contain more than two numbers (0 and 1 in our case). If the values in the table were allowed to be drawn from a set consisting of more than two elements, there would be many antidiagonals. Note also that the presupposition of the application of diagonalization is that the table is square, i.e. that the number of rows of the original table equals the number of its columns.

Simple as this method may seem to be, it lays the foundation of many path-breaking results of logic in the second half of the nineteenth and the first half of twentieth century.[1] Let us assume that each row in a table we are considering has a label and let us use the sign **D** for the set of all the labels. Each column of the table can then be considered as a function mapping **D** (in our introductory examples **D** could be {1,2,3}, for the tables have three rows) on a set **R** of those values that can occur within the table (in our first example, **R** may be {0,1,3,5,7} (or any set containing it), in the second one it would be {0,1}).

	f_1	f_2	f_3	f_4	f_5	...
d_1	$f_1(d_1)$	$f_2(d_1)$	$f_3(d_1)$	$f_4(d_1)$	$f_5(d_1)$	...
d_2	$f_1(d_2)$	$f_2(d_2)$	$f_3(d_2)$	$f_4(d_2)$	$f_5(d_2)$	...
d_3	$f_1(d_3)$	$f_2(d_3)$	$f_3(d_3)$	$f_4(d_3)$	$f_5(d_3)$	...
d_4	$f_1(d_4)$	$f_2(d_4)$	$f_3(d_4)$	$f_4(d_4)$	$f_5(d_4)$	...
d_5	$f_1(d_5)$	$f_2(d_5)$	$f_3(d_5)$	$f_4(d_5)$	$f_5(d_5)$	...
...	...	...	...	...	...	...

1 In this paper we concentrate at the most perspicuous presentation of the diagonal argument. For more detailed, deeper and more technical accounts see Smullyan [1994], Boolos et al. [2002], or Gaifman [2006].

The table thus presents a set **F** of functions from **D** to **R**, such that the number of elements of **F** coincides with that of **D** (the table is square); and the diagonal method shows that there is a function from **D** to **R** which does not belong to **F:**

Theorem. Let **F** be a set of functions with the domain **D** and range **R**. Let **R** consist of at least two elements. Then, if the cardinality of **F** is the same as that of **D**, there exists a function from **D** to **R** which is not an element of **F.**

Proof: Let i be a one-one mapping of **D** on **F**. Let **f** be such that $\mathbf{f}(x) \neq f_x(x)$, where $f_x = i(x)$, for every x from **D**. Then **f** is – obviously – not an element of **F**.

This formulation allows us to extend our considerations to infinite "tables" – even to "tables" with a non-denumerable number of rows and columns. But by saying this we make our exposition basically a-historical, for diagonalization was first used to prove the very existence of non-denumerable cardinalities.

2. Cardinality issues

A straightforward application of diagonalization shows that however we order infinite sequences of natural numbers into a succession, the succession will not contain all the sequences.

	1	2	3	4	5	...	AD
1	5	0	3	8	4	...	**1**
2	1	1	3	3	6	...	**2**
3	3	7	3	7	7	...	**1**
4	4	4	4	1	1	...	**2**
5	9	6	7	3	2	...	**1**
...	...	...	...	...	...	...	...

This is usually interpreted in such a way that there are more sequences of natural numbers than there are natural numbers; hence, that there is an infinity greater than the infinity of natural numbers.[2] Note that this result keeps holding even if we only consider sequences of some restricted subset of natural numbers, such as $\{0,1,2,3,4,5,6,7,8,9\}$ or indeed $\{0,1\}$.

Now consider real numbers between zero and one, i.e. numbers of the shape $0{,}x_1x_2x_3$ ..., where x_1, x_2, x_3 ... is an infinite sequence of one-digit numerals. Each such number can therefore be identified with an infinite sequence of natural numbers;[3] it follows that there are more real numbers than natural numbers. The acceptance of this view by Cantor [1874] marked a break in the foundations of mathematics.

2 Though this is nowadays an almost universally accepted interpretation, it is perhaps not quite inevitable – we might for example insist that the reason that we are not able to order all the sequences in a single succession is not a matter of their number, but rather of some peculiarities of the ordering procedure.

3 In fact, with some trivial exceptions: a real number of the shape $0{,}x_1x_2x_3...x_n999...$ (with no other digit than 9 thereafter) is considered to be the same as $0{,}x_1x_2x_3...x_n'000...$ (with zeros thereafter), where $x_n'=x_n+1$. But it is easy to show that these exceptions are not relevant.

Still more generally, take the labels of rows to be elements of an arbitrary set S and the columns of the table as the characteristic functions of its subsets: i.e. every column represents that subset of S which consists of those elements of S to which it assigns the value 1. If $S = \{e_1, e_2, e_3, ...\}$ and we denote its subsets as $s_1, s_2, s_3, ...$, we have the following table:

S	s_1	s_2	s_3	s_4	s_5	...	AD
e_1	0	1	1	0	1	...	**1**
e_2	0	0	1	0	0	...	**1**
e_3	0	0	1	0	1	...	**0**
e_4	0	0	0	1	0	...	**0**
e_5	0	0	0	1	1	...	**0**
...	...	...	...	...	...	...	...

Here diagonalization shows us that there are more subsets of any set than the elements of the set. These results are connected with the birth of set theory and especially, again, with the research of Cantor [1890].

3. Paradoxes

Suppose we have all the properties that there are (*being black*, *being a fish*, *being a color*, *being a property*, ...). Denote them as $p_1, p_2, p_3, ...$. Make them the labels of both rows and columns of a table and fill the cell in the intersection of the ith row with the jth column with 1 iff the ith property has the jth property (hence if, e.g., *being black is a color*, or if *being a property is a fish*) and with 0 otherwise; and construct the antidiagonal.

	p_1	p_2	p_3	p_4	p_5	...	p^*
p_1	0	1	1	0	1	...	**1**
p_2	0	0	1	0	0	...	**1**
p_3	0	0	1	0	1	...	**0**
p_4	0	0	0	1	0	...	**0**
p_5	0	0	0	1	1	...	**0**
...	...	...	...	...	...	...	...

Now the antidiagonal contains 1 in its ith row iff p_i does not have itself; hence, it corresponds to a property p^* that maps every property on 1 iff the property does not have itself. In this sense, p^* would seem to be the property of *not having itself*. And as this indeed appears to be a property and as we assumed that the table contained all the properties that there are, p^* must be identical with p_j for some j. Then, however, the jth element of the diagonal is the jth element of both the column p_j and the column p^*. In other words, it is $p_j(p_j)$ and at the same time it is $\neg p_j(p_j)$. Thus, $p_j(p_j)$ is true iff it is false. This is the celebrated paradox of Russell [1908].

What to make of it? One interpretation of this fact is that *not having itself* only *seems* to be a property, but in fact it is not. For how could it be a property if it cannot be a member of any set of properties? Another interpretation is that there is such a property and that hence there are real inconsistencies plaguing natural language[4] and that the role of logic is to establish artificial languages gerrymandered in such a way that no inconsistencies are let into them. Hence, let us now turn to the case when $p_1, p_2, p_3, \ldots$ are all properties *that are expressible in terms of a fixed language.*

4. Incompleteness of arithmetic

Let now $p_1, p_2, p_3, \ldots$ be not properties, but formulas of the language of Peano arithmetic (PA), each of which contains exactly one free variable. We will call such formulas *pseudopredicates*; they can clearly be considered as expressing (numerical) properties. In particular, every such formula is true of some numbers and false for others and expresses the property which a number has iff the formula is true of it. Thus, for example, the formula $x>5$ expresses the property of being bigger than five, whereas the formula $\exists y\ (x = 2.y)$ expresses the property of being even. If we denote this last formula as p, we shall denote by $p(1)$,$p(2)$, $p(365)$, etc., the result of the replacement of its single free variable by 1, 2, 365, etc., respectively, i.e. the respective formulas $\exists y\ (1 = 2.y)$, $\exists y\ (2 = 2.y)$, $\exists y\ (365 = 2.y)$, etc.

At the same time, if we fix a Gödel numbering,[5] every such pseudopredicate p will have a number $\ulcorner p \urcorner$. Then if, for instance, p is $\exists y\ (x = 2.y)$ and the Gödel number $\ulcorner p \urcorner$ of this formula is 365, we can form the formula $\exists y\ (365 = 2.y)$ (which is, by the way, obviously false), which results from replacing the only free variable of p by its own Gödel number; i.e. it is the formula $p(\ulcorner p \urcorner)$. Then, if we denote the truth value of a (closed) formula f as $|f|$ and the opposite value as $\overline{|f|}$, we can form the table such that the number in the intersection of the ith row and the jth column indicates whether the number $\ulcorner p_i \urcorner$ has the property p_j, i.e. it is the truth value of the formula $p_j(\ulcorner p_i \urcorner)$ (which is the formula that results from the replacement of all occurrences of the single free variable of p_j by the numeral $\ulcorner p_i \urcorner$):

4 In an extreme form this may lead to the theory of dialetheism (see Priest [1998]), according to which there really are propositions that are both true and false.

5 As Gödel showed, the expressions of the language of arithmetic can be "enumerated", i.e. mapped on numerals in such a way that the mapping is one-to-one and that we are able to compute the number of any formula and find the formula with any given number.

	p_1	p_2	p_3	…	AD
p_1	$\lvert p_1(\ulcorner p_1 \urcorner)\rvert$	$\lvert p_2(\ulcorner p_1 \urcorner)\rvert$	$\lvert p_3(\ulcorner p_1 \urcorner)\rvert$	…	$\overline{\lvert \boldsymbol{p}_1(\ulcorner \boldsymbol{p}_1 \urcorner)\rvert}$
p_2	$\lvert p_1(\ulcorner p_2 \urcorner)\rvert$	$\lvert p_2(\ulcorner p_2 \urcorner)\rvert$	$\lvert p_3(\ulcorner p_2 \urcorner)\rvert$	…	$\overline{\lvert \boldsymbol{p}_2(\ulcorner \boldsymbol{p}_2 \urcorner)\rvert}$
p_3	$\lvert p_1(\ulcorner p_3 \urcorner)\rvert$	$\lvert p_2(\ulcorner p_3 \urcorner)\rvert$	$\lvert p_3(\ulcorner p_3 \urcorner)\rvert$	…	$\overline{\lvert \boldsymbol{p}_3(\ulcorner \boldsymbol{p}_3 \urcorner)\rvert}$
p_4	$\lvert p_1(\ulcorner p_4 \urcorner)\rvert$	$\lvert p_2(\ulcorner p_4 \urcorner)\rvert$	$\lvert p_3(\ulcorner p_4 \urcorner)\rvert$	…	$\overline{\lvert \boldsymbol{p}_4(\ulcorner \boldsymbol{p}_4 \urcorner)\rvert}$
p_5	$\lvert p_1(\ulcorner p_5 \urcorner)\rvert$	$\lvert p_2(\ulcorner p_5 \urcorner)\rvert$	$\lvert p_3(\ulcorner p_5 \urcorner)\rvert$	…	$\overline{\lvert \boldsymbol{p}_5(\ulcorner \boldsymbol{p}_5 \urcorner)\rvert}$
…	…	…	…	…	**…**

The antidiagonal now maps $\ulcorner p_i \urcorner$ on the truth value opposite to that of $p_i(\ulcorner p_i \urcorner)$, and it is immediately clear that no formula of arithmetic can yield this mapping. And, unlike the case of *not having itself* considered in the context of all conceivable properties, this conclusion is not problematic – on the contrary, it is good for arithmetic to be put together so that it avoids the paradox.

Now Gödel showed, among other things, that the function mapping the number of a formula p on the number of $p(\ulcorner p \urcorner)$ can be expressed by a term with a free variable of the language of PA – we can introduce the function symbol Dg that expresses it. Suppose that we have a pseudopredicate Tr such that is true precisely of numbers of true formulas. In this case, we could form the formula $\neg Tr(Dg(x))$, which would produce precisely the antidiagonal column – indeed, it would be true of a formula p just in case $p(\ulcorner p \urcorner)$ would not be true. It follows that the language of PA cannot contain the pseudopredicate Tr. (This result is sometimes referred to as *Tarski's theorem*.)

On the other hand Gödel showed that there is a pseudopredicate which is true (and provably so) precisely of numbers of formulas *provable* within PA, and that hence we can introduce the predicate symbol Pr with this property. Hence we do have the formula $\neg Pr(Dg(x))$ which appears to be an analogue of the previous one with provability in place of truth. So consider a variation of the previous table in which the number in the cell at the intersection of the ith row and the jth column now indicates whether the number $\ulcorner p_i \urcorner$ has the property p_j *provably*, i.e. it is 1 iff the formula $p_j(\ulcorner p_i \urcorner)$ is provable (i.e. *provably* true) and is 0 iff it is *refutable* (i.e. *provably* false, its negation being provable). If $\lVert f \rVert$ is 1 for a provable f and is 0 for a refutable f (and $\overline{\lVert f \rVert}$ is the opposite value), we have

	p_1	p_2	p_3	…	AD
p_1	$\lVert p_1(\ulcorner p_1 \urcorner)\rVert$	$\lVert p_2(\ulcorner p_1 \urcorner)\rVert$	$\lVert p_3(\ulcorner p_1 \urcorner)\rVert$	…	$\overline{\lVert \boldsymbol{p}_1(\ulcorner \boldsymbol{p}_1 \urcorner)\rVert}$
p_2	$\lVert p_1(\ulcorner p_2 \urcorner)\rVert$	$\lVert p_2(\ulcorner p_2 \urcorner)\rVert$	$\lVert p_3(\ulcorner p_2 \urcorner)\rVert$	…	$\overline{\lVert \boldsymbol{p}_2(\ulcorner \boldsymbol{p}_2 \urcorner)\rVert}$
p_3	$\lVert p_1(\ulcorner p_3 \urcorner)\rVert$	$\lVert p_2(\ulcorner p_3 \urcorner)\rVert$	$\lVert p_3(\ulcorner p_3 \urcorner)\rVert$	…	$\overline{\lVert \boldsymbol{p}_3(\ulcorner \boldsymbol{p}_3 \urcorner)\rVert}$
p_4	$\lVert p_1(\ulcorner p_4 \urcorner)\rVert$	$\lVert p_2(\ulcorner p_4 \urcorner)\rVert$	$\lVert p_3(\ulcorner p_4 \urcorner)\rVert$	…	$\overline{\lVert \boldsymbol{p}_4(\ulcorner \boldsymbol{p}_4 \urcorner)\rVert}$
p_5	$\lVert p_1(\ulcorner p_5 \urcorner)\rVert$	$\lVert p_2(\ulcorner p_5 \urcorner)\rVert$	$\lVert p_3(\ulcorner p_5 \urcorner)\rVert$	…	$\overline{\lVert \boldsymbol{p}_5(\ulcorner \boldsymbol{p}_5 \urcorner)\rVert}$
…	…	…	…	…	**…**

Now does the pseudopredicate $\neg Pr(Dg(x))$ produce this antidiagonal? We know this cannot be the case – if it were, then we would have a contradiction, for the antidiagonal is different from all the columns in the table, and yet as $\neg Pr(Dg(x))$ is a pseudopredicate of the language of PA, it would have to be one of the columns. Why this is not the case?

It is the case that if $||p(\ulcorner p \urcorner)|| = 1$, i.e. if p is provably true of itself, $\neg Pr(Dg(\ulcorner p \urcorner))$ is provably false and hence $||\neg Pr(Dg(\ulcorner p \urcorner))|| = \overline{||p(\ulcorner p \urcorner)||} = 0$. (This follows from the fact that $p(\ulcorner p \urcorner)$ is provable iff $Pr(\ulcorner p(\ulcorner p \urcorner) \urcorner)$ is provable, and $Pr(\ulcorner p(\ulcorner p \urcorner) \urcorner)$ is equivalent to $Pr(Dg(\ulcorner p \urcorner))$ and hence to $\neg\neg Pr(Dg(\ulcorner p \urcorner))$). Conversely, if $||p(\ulcorner p \urcorner)|| = 0$, then $\overline{||p(\ulcorner p \urcorner)||} = 1$. Hence the new column contains 1 iff the diagonal contains 0. Thus, the new column would be the antidiagonal – and the contradiction would be inevitable – if it were the case that any formula were provably true iff it were not provably false. (For in this case all cells of the diagonal which would not contain 1's would contain 0's and the corresponding cells of the new column would contain 1's.) But while no formula is at the same time provable and refutable (at least as long as PA is consistent), it need not be the case that every formula is either provable, or refutable. And we see that it *cannot* be the case, in pain of contradiction. Hence if PA is consistent, then it is not complete, in pain of contradiction. This is the celebrated *incompleteness* discovered and proven by Gödel [1931].

5. Fixed points

Let us investigate an alternative way of reaching incompleteness via diagonalization, a way that is closer to the way Gödel himself proceeded. Consider a property q of numbers, i.e. a mapping of numbers on truth values. Let us form a column, p^*, by associating every pseudopredicate p with q applied to $Dg(\ulcorner p \urcorner)$:

	p^*
p_1	$q(\ulcorner p_1(\ulcorner p_1 \urcorner) \urcorner)$
p_2	$q(\ulcorner p_2(\ulcorner p_2 \urcorner) \urcorner)$
p_3	$q(\ulcorner p_3(\ulcorner p_3 \urcorner) \urcorner)$
p_4	$q(\ulcorner p_4(\ulcorner p_4 \urcorner) \urcorner)$
p_5	$q(\ulcorner p_5(\ulcorner p_5 \urcorner) \urcorner)$
...	...

Whether this column coincides with one of the columns of the table or not depends on the specific nature of q, in particular on whether q is expressible in the language of PA (in view of the obvious fact that $q(\ulcorner p_i(\ulcorner p_i \urcorner) \urcorner)$ is expressible in arithmetic just in case q is). If, for instance, q is *is not true*, then the column becomes the antidiagonal. On the other hand, if q is expressible in arithmetic, then there must be a p_j which expresses $q(\ulcorner p_i(\ulcorner p_i \urcorner) \urcorner)$.

In this case, consider the cell in the intersection of the jth row and the jth column. According to the definition of the table, it will contain the truth value of $p_j(\ulcorner p_j \urcorner)$. On the other hand, in view of the fact that this column coincides with that for p^*, it will also contain the truth value of $q(\ulcorner p_j(\ulcorner p_j \urcorner)\urcorner)$. As a result, the values of $q(\ulcorner p_j(\ulcorner p_j \urcorner)\urcorner)$ and $p_j(\ulcorner p_j \urcorner)$ are bound to coincide; schematically $q(\ulcorner p_j(\ulcorner p_j \urcorner)\urcorner) \leftrightarrow p_j(\ulcorner p_j \urcorner)$. Hence, we have shown what is usually called the *fix point lemma*: for every property q expressible in arithmetic there will be a sentence s_q of arithmetic such that $q(\ulcorner s_q \urcorner) \leftrightarrow s_q$.

In this way, we arrive at the inexpressibility of the truth property in arithmetic by an alternative route. If the property were expressible, then its negation would be also and it would have a fixed point $s_{\neg Tr}$ such that $\neg Tr(\ulcorner s_{\neg Tr} \urcorner) \leftrightarrow s_{\neg Tr}$ and hence that $Tr(\ulcorner s_{\neg Tr} \urcorner) \leftrightarrow \neg s_{\neg Tr}$. But as Tr is a truth predicate only if $Tr(\ulcorner s \urcorner) \leftrightarrow s$ for every statement s[6], it is also the case that $Tr(\ulcorner s_{\neg Tr} \urcorner) \leftrightarrow s_{\neg Tr}$. Putting the two equivalences together, we have $s_{\neg Tr} \leftrightarrow \neg s_{\neg Tr}$; and hence we have a contradiction.

Now imagine that we take q to be the property of non-provability, i.e. a property which a number has iff it is a number of a formula not provable in PA. We already know that this property is expressible in arithmetic, so it *does* have a fixed point. Hence, there is a $s_{\neg Pr}$ so that $\neg Pr(\ulcorner s_{\neg Pr} \urcorner) \leftrightarrow s_{\neg Pr}$. Now suppose that $s_{\neg Pr}$ is provable; if so, then so is $\neg Pr(\ulcorner s_{\neg Pr} \urcorner)$. But this could only be if $s_{\neg Pr}$ were not provable, hence the assumption of the provability of $s_{\neg Pr}$ leads to the contradictory conclusion of its non-provability; hence, $s_{\neg Pr}$ cannot be provable. Suppose, then, that $s_{\neg Pr}$ is refutable, hence that $\neg s_{\neg Pr}$ is provable. Then $Pr(\ulcorner s_{\neg Pr} \urcorner)$ is provable, and, as a result, $s_{\neg Pr}$ is provable. Hence $s_{\neg Pr}$ cannot be refutable either – in pain of inconsistency.

6. Turing machines

The problem of the decidability of an axiomatic system is the problem of whether we can always decide if a given formula of the system is a theorem. Note that if the system is such that every formula that is not provable is refutable, then the decision procedure is always at hand: we use the axioms and rules to continue generating the theorems and, sooner or later, we must reach either the formula, or its negation. (True, it might be a procedure that is not very practical since reaching the result may take a lot of time, but it works.) If, on the other hand, this is not the case (and in case of languages of pure logic it cannot be the case, for their theorems are only logical truths, and certainly not every negation of a sentence that is not logically true is logically true), the existence of a decision procedure is not guaranteed.

Alan Turing [1937], when he dealt with this problem, saw the necessity of exactly specifying what is a "procedure" or an "algorithm". His answer to this question was the abstract machines which later came to bear his name: *Turing machines*. For simplicity's sake, let us assume that the machines deal only with natural numbers, i.e. that if any such machine is fed with a natural number it starts computing and, if it halts, it yields another natural number. Thus, any such machine "realizes" a function from natural numbers to natural

6 The fact that this is precisely what characterized the property of truth was argued for by Tarski [1932].

numbers. We will not talk about the inner structure of the machines here, but we note that any such machine is uniquely describable by language and hence can be identified with a certain (sometimes perhaps very long) expression. Thus, all the machines can be enumerated (M_1, M_2, ...) and we can also always find the nth machine according to the enumeration.

Now consider the table with rows labeled with natural numbers and columns labeled with Turing machines. The number in the intersection of the ith row and the jth column is the value yielded by M_j for the input i (as the machine may not stop, the cell may be also empty).[7] Construct an antidiagonal as indicated in the table (where we take $M_i[i]$+1 to be 0 iff M_i does not stop for the input i):

	M_1	M_2	M_3	M_4	M_5	...	?
1	$M_1[1]$	$M_2[1]$	$M_3[1]$	$M_4[1]$	$M_5[1]$	...	$M_1[1]$+1
2	$M_1[2]$	$M_2[2]$	$M_3[2]$	$M_4[2]$	$M_5[2]$	...	$M_2[2]$+1
3	$M_1[3/]$	$M_2[3]$	$M_3[3]$	$M_4[3]$	$M_5[3]$	...	$M_3[3]$+1
4	$M_1[4]$	$M_2[4]$	$M_3[4]$	$M_4[4]$	$M_5[4]$	...	$M_4[4]$+1
5	$M_1[5]$	$M_2[5]$	$M_3[5]$	$M_4[5]$	$M_5[5]$	...	$M_5[5]$+1
...	...	...	...	...	...	...	

The antidiagonal cannot be computable by a Turing machine (since it is different from every column of the original table, which correspond to every Turing machine). But is it really not computable? Imagine the following computation: given a number j we find the machine M_j (we have already noted that we can do this), we let it run on the input j, add 1 and ... *voilà*! There is, of course, a snag. We must wait until M_j stops and yields its result; but what if it never stops? We would be waiting forever (for we would never know whether it merely has not stopped *yet*, or whether it would *never* stop). Hence, what we would need is an algorithm which would be able to tell us, for any given machine M and any input i, whether M ever stops for i.

Hence, there cannot be a Turing machine solving this *halting problem* – and insofar as we are convinced that everything that is solvable is solvable by a Turing machine, the problem is generally unsolvable. And, as it can be shown that the halting problem would be solvable if the predicate calculus were decidable (the stopping of every Turing machine

---

7 In fact, the result that at least some machines cannot stop for every argument can be established by means of a consideration similar to that by which we established the incompleteness of arithmetic. Imagine a *universal* Turing machine U, a machine that is able to simulate any Turing machine in the sense that if it gets, as its input, the description of some Turing machine m plus some data d (we will write $U\ (m \oplus d)$ (where '$\oplus$' symbolizes concatenation by means of some kind of separator) it stops just in case m stops for the input d and in that case it yields the same value: $U(m \oplus d) = m(d)$. It is easy to turn U into a machine U' such that $U'(m \oplus d) \neq m(d)$ whenever m stops for d. Further, it is easy to turn U' into U'' such that $U''(d) = U'(d \oplus d)$. Now $U''(U'') = U'(U'' \oplus U'') \neq U''(U'')$. This shows that U'' can never stop for the data U''.

turns out to be equivalent to a certain formula being true – see Boolos et al. [2002]), predicate calculus is undecidable.

7. Conclusion

Diagonalization is, in essence, a trivial method of constructing, for a square table, a column that is not yet contained in the table. However, it has far-reaching consequences; in fact, consequences that reach as far as the most path-breaking problems and results of modern logic. It allows us to extend the trivial observation that there are more subsets than elements of a set from the finite case to infinite ones, thus establishing the need for a hierarchy of infinities, instantiated by the infinite of natural numbers, that of real numbers, etc. Also, it allows us to see that *not having itself* is a property of properties that is strangely anomalous in that once it is expressed in a language, it makes the language inconsistent. Within the framework of the exactly delimited language of arithmetic, this yields us, first, the consequence that the concept of truth cannot be expressed by any pseudopredicate of the language; and, second, as there is a predicate expressing the concept of provability, the consequence that the language must be incomplete. Applied to the realm of Turing machines, it further yields us the result that the halting problem for these machines must be unsolvable. In this way, the *prima facie* simple observation of the possibility to diagonalize any square table leads us to a battery of very nontrivial results constituting, as it were, the central nervous system of modern logic.

Acknowledgments

Work on this paper has been supported by Research Grant No. 17-15645S of the Czech Science Foundation. I am grateful to Pavel Arazim for valuable critical comments.

References

G. S. Boolos, J. R. Burgess, and R. C. Jeffrey. *Computability and Logic (4th edition)*. Cambridge University Press, 2002.

G. Cantor. Über eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Journal für die reine und angewandte Mathematik*, 77: 258–262, 1874.

G. Cantor. Über eine elementare Frage der Mannigfaltigkeitslehre. *Jahresbericht der Deutschen Mathematiker-Vereingumg*, 1: 75–78, 1890.

H. Gaifman. Naming and diagonalization, from Cantor to Gödel to Kleene. *Oxford Journal of the IGPL*, 15: 709–728, 2006.

K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38: 173–198, 1931.

G. Priest. What is so bad about contradictions? *The Journal of Philosophy*, 95: 410–426, 1998.

B. Russell. Mathematical logic as based on the theory of types. *American Journal of Mathematics*, XXX: 222–262, 1908.

R. M. Smullyan. *Diagonalization and self-reference*. Clarendon Press, 1994.

A. Tarski. Der Wahrheitsbegriff in den Sprachen der deduktiven Disziplinen. *Anzeiger der Österreichischen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse*, 69: 23–25, 1932.

A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. *Proceedings of The London Mathematical Society*, 42: 230–265, 1937.