Metamathematics of Elementary Mathematics

Lectures 1 and 2

Alexandre V. Borovik

Introduction

Toutes les grandes personnes ont d'abord été des enfants
(Mais peu d'entre elles s'en souviennent.)
Antoine de Saint-Exupéry, *Le Petit Prince.*

These lectures is an attempt to gain a better understanding of the specific nature of mathematical practice by looking at mathematics from a new and somewhat unusual point of view. I propose to systematically record and analyse *mathematically* logical difficulties experienced—and occasionally overcome—by children in their early learning of mathematics. This explains the title of the course: according to a widely accepted understanding of this term, *metamathematics* is mathematics applied to study of mathematics. In my case I apply mathematics for the sake of better understanding of the practice of elementary school mathematics.

Quite naturally, the real life material in my research is limited to analysis of recollections of my fellow mathematicians about their childhood. Unfortunately, only mathematicians posses an adequate language which allows them to describe in some depths their personal experiences of learning mathematics. So far my approach is justified only by warm welcome it found in some of my mathematician friends, and I am most grateful to them for their support.

I am a mathematician and restrict myself to clearly describing hidden structures of elementary mathematics which may intrigue and—like shadows in the night—sometimes scare an inquisitive child. I hope that my notes could be useful to specialists in mathematical education and in psychology of education. But I refrain from making any recommendations on mathematics teaching. For me, the primary aim of my project is to understand the nature of hardcore mainstream "research" mathematics.

I hope that my proposal provides an answer to frequently asked question: "Is mathematics special?" Of course it is! The emphasis on child's experiences makes my programme akin to linguistic and cognitive science. However, when a linguist studies formation of speech in a child, he studies language, not the structure of linguistic as a scientific discipline. When I propose to study formation of mathematical concepts in a child, I wish to get insights into interplay of mathematical structures in *mathematics*. Mathematics has an astonishing power of reflection,

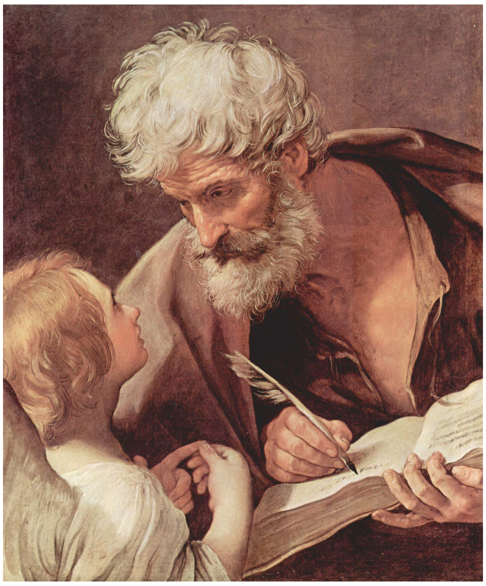

Figure 1. *L'Evangelista Matteo e l'Angelo.* Guido Reni, 1630–1640. Pinacoteca Vaticana. Source: *Wikipedia Commons*. Public domain. Guido Reni was one of the first artists in history of visual arts who paid attention to psychology of children. Notice how the little angel counts on his fingers the points he is sent to communicate to St. Matthew.

and a self-referential study of mathematics by mathematical means plays an increasingly important role within mathematical culture. I simply suggest to make a step further (or aside, or back in life) and take a look back in time, in one's child years.

The lectures are also an attempt to trigger the chain of memories in my listeners: every, even most minute, recollection of difficulties and paradoxes of their early mathematical experiences are most welcome. Please write to me at

`borovik@manchester.ac.uk`.

Contents

1. "Named" numbers

1.1. Dividing apples between people. I take the liberty to tell a story from my own life[1]; I believe it is relevant for the principal theme of the paper.

When, as a child, I was told by my teacher that I had to be careful with *"named"* numbers and not to *add* apples and people, I remember asking her why in that case we can *divide* apples by people:

$$10\,\text{apples} \;:\; 5\,\text{people} = 2\,\text{apples}. \tag{1}$$

Even worse: when we distribute 10 apples giving 2 apples to a person, we have

$$10\,\text{apples} \;:\; 2\,\text{apples} = 5\,\text{people} \tag{2}$$

Where do "people" on the right hand side of the equation come from? Why do "people" appear and not, say, "kids"? There were no "people" on the left hand side of the operation! How do numbers on the left hand side know the name of the number on the right hand side?

There were much deeper reasons for my discomfort. I had no bad feelings about dividing 10 apples among 5 people, but I somehow felt that the problem of deciding how many people would get apples if each was given 2 apples from the total of 10, was completely different. (My childhood experience is confirmed by experimental studies, see Bryant and Squire [**6**].)

In the first problem you have a fixed data set: 10 apples and 5 people, and you can easily visualize giving apples to the people, in rounds, one apple to a person at a time, until no apples were left. But an attempt to visualize the second problem in a similar way, as an orderly distribution of apples to a queue of people, two apples to each person, necessitated dealing with a potentially unlimited number of recipients. In horror I saw an endless line of poor wretches, each stretching out his hand, begging for his two apples. This was visualization gone astray. I was not in control of the queue! But reciting numbers, like chants, while counting *pairs* of apples, had a soothing, comforting influence on me and restored my shattered confidence in arithmetic.

I did not get a satisfactory answer from my teacher and only much later did I realize that the correct naming of the numbers should be

$$10\,\text{apples} \;:\; 5\,\text{people} = 2\,\frac{\text{apples}}{\text{people}}, \qquad 10\,\text{apples} \;:\; 2\,\frac{\text{apples}}{\text{people}} = 5\,\text{people}. \tag{3}$$

It is a commonplace wisdom that the development of mathematical skills in a student goes alongside the gradual expansion of the realm of numbers with which

[1] Call me AVB; I am male, have a PhD in Mathematics, teach mathematics at a research-led university.

FIGURE 2. The First Law of Arithmetic: you do not add fruit and people. Giuseppe Arcimboldo, *Autumn*. 1573. Musée du Louvre, Paris. Source: *Wikipedia Commons*. Public domain.

he or she works, from natural numbers to integers, then to rational, real, complex numbers:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

What is missing from this natural hierarchy is that already at the level of elementary school arithmetic children are working in a much more sophisticated structure, a graded ring

$$\mathbb{Q}[x_1, x_1^{-1}, \dots, x_n, x_n^{-1}].$$

of Laurent polynomials in n variables over $\mathbb{Q}$, where symbols $x_1, \dots, x_n$ stand for the names of objects involved in the calculation: apples, persons, etc. This explains why educational psychologists confidently claim that the operations (1) and (2) have little in common [**6**]—indeed, operation (2) involves an operand "apple/people" of a much more complex nature than basic "apples" and "people" in operation (1): "apple/people" could appear only as a result of some previous division.

Usually, only Laurent monomials are interpreted as having physical (or real life) meaning. But the addition of heterogeneous quantities still makes sense and is done componentwise: if you have a lunch bag with (2 apples + 1 orange), and another bag, with (1 apple + 1 orange), together they make

$$(2 \text{ apples } + 1 \text{ orange}) + (1 \text{ apple } + 1 \text{ orange}) = (3 \text{ apples } + 2 \text{ oranges}).$$

Notice that this gives a very intuitive and straightforward approach to vectors.[2]

Of course, there is no need to teach Laurent polynomials to kids; but it would not harm to teach them to teachers. I have an ally in François Viéte who in 1591 wrote in his *Introduction to the Analytic Art* [**23**] that

> *If one magnitude is divided by another, [the quotient] is heterogeneous to the former ... Much of the fogginess and obscurity of the old analysts is due to their not paying attention to these [rules].*

1.2. Dimensions. Physicists love to work in the Laurent polynomial ring

$$\mathbb{R}[\text{length}^{\pm 1}, \text{time}^{\pm 1}, \text{mass}^{\pm 1}]$$

because they love to measure all physical quantities in combinations (called "dimensions") of the three basic units: for length, time and mass. But then even this ring becomes too small since physicists have to use fractional powers of basic units. For example, velocity has dimensions length/time, while electric charge can be meaningfully treated as having dimensions

$$\frac{\text{mass}^{1/2}\text{length}^{3/2}}{\text{time}}.$$

Indeed if we choose our units in such a way that the permittivity ϵ_0 of free space is dimensionless, then from Coulomb's law

$$F = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r^2}$$

applied to two equal charges $q_1 = q_2 = q$, we see that q^2/r^2 has the dimensions of force.

It pays to be attentive to the dimensions of quantities involved in a physical formula: the balance of names of units (*dimensions*) on the left and right hand sides may suggest the shape of the formula. Such *dimensional analysis* quickly leads to immensely deep results, like, for example, Kolmogorov's celebrated "5/3 Law" for the energy spectrum of turbulence.

Exercises.

Exercise 1.1. To scare the reader into acceptance of the intrinsic difficulty of division, I refer to paper *Division by three* [**10**] by Peter Doyle and John Conway. I quote their abstract:

> We prove without appeal to the Axiom of Choice that for any sets A and B, if there is a one-to-one correspondence between $3 \times A$ and $3 \times B$ then there is a one-to-one correspondence between A and B. The first such proof, due to Lindenbaum, was announced by Lindenbaum and Tarski in 1926, and subsequently 'lost'; Tarski published an alternative proof in 1949.

[2]By the way, this "lunch bag" approach to vectors allows a natural introduction of duality and tensors: the total cost of a purchase of amounts g_1, g_2, g_3 of some goods at prices p^1, p^2, p^3 is a "scalar product"-type expression $\sum g_i p^i$. We see that the quantities g_i and p_i could be of completely different nature. The standard treatment of scalar (dot) product in undergraduate linear algebra usually conceals the fact that dot product is a manifestation of duality of vector spaces, creating immense difficulties in the subsequent study of tensor algebra.

Here, of course, 3 is a set of 3 elements, say, $\{0, 1, 2\}$. An exercise for the reader: prove this in a naive set theory *with* the Axiom of Choice.

Exercise 1.2. Theoretical physicists occasionally love to use a system of measurements based on fundamental units:

- speed of light $c = 299,792,458$ meters per second,
- gravitational constant $G = (6.67428 \pm 0.00067) \times 10^{-11}\ \mathrm{m^3 kg^{-1} s^{-2}}$ and
- Planck' constant $h = 6.62606896 \times 10^{-34}\ \mathrm{m^2 kg \cdot s^{-1}}$.

Express the more common physical units: meter, kilogram, second in terms of c, G, h.

2. The triumph of "named numbers": Kolmogorov's "5/3" Law

To demonstrate the power of "named numbers", I reproduce what remains the most striking and beautiful example of dimensional analysis in mathematics. The deduction of Kolmogorov' seminal "5/3" law for the energy distribution in the turbulent fluid [**14**] is so simple that it can be done in a few lines. I was lucky to study at a good secondary school where my physics teacher (Anatoly Mikhailovich Trubachov, to whom I express my eternal gratitude) derived the "5/3" law in one of his improvised lectures. In my exposition, I borrow some details from Arnold [**2**] and Ball [**3**] (where I have also picked the idea of using a woodcut by Katsushika Hokusai, Figure 3, as an illustration).

Figure 3. Multiple scales in the motion of a fluid, from a woodcut by Katsushika Hokusai *The Great Wave off Kanagawa* (from the series *Thirty-six Views of Mount Fuji*, 1823–29). This image is much beloved by chaos scientists. Source: *Wikipedia Commons*. Public domain.

2.1. Turbulent flows: basic setup. The turbulent flow of a liquid consists of vortices; the flow in every vortex is made of smaller vortices, all the way down the scale to the point when the viscosity of the fluid turns the kinetic energy of motion into heat (Figure 3). If there is no influx of energy (like the wind whipping up a storm in Hokusai's woodcut), the energy of the motion will eventually dissipate and the water will stand still. So, assume that we have a balanced energy flow, the storm is already at full strength and stays that way. The motion of a liquid is made of waves of different lengths; Kolmogorov asked the question, what is the share of energy carried by waves of a particular length?

Here it is a somewhat simplified description of his analysis. We start by making a list of the quantities involved and their dimensions. First, we have the *energy flow* (let me recall, in our setup it is the same as the dissipation of energy). The dimension of energy is

$$\frac{\text{mass} \cdot \text{length}^2}{\text{time}^2}$$

(remember the formula $K = mv^2/2$ for the kinetic energy of a moving material point). It will be convenient to make all calculations *per unit of mass.* Then the energy flow ϵ has dimension

$$\frac{\text{energy}}{\text{mass} \cdot \text{time}} = \frac{\text{length}^2}{\text{time}^3}$$

For counting waves, it is convenient to use the *wave number*, that is, the number of waves fitting into the unit of length. Therefore the wave number k has dimension

$$\frac{1}{\text{length}}.$$

Finally, the *energy spectrum* $E(k)$ is the quantity such that, given the interval $\Delta k = k_1 - k_2$ between the two wave numbers, the energy (per unit of mass) carried by waves in this interval should be approximately equal to $E(k_1)\Delta k$. Hence the dimension of E is

$$\frac{\text{energy}}{\text{mass} \cdot \text{wavenumber}} = \frac{\text{length}^3}{\text{time}^2}.$$

2.2. Subtler analysis. To make the next crucial calculations, Kolmogorov made the major assumption that amounted to saying that[3]

> The way bigger vortices are made from smaller ones is the same throughout the range of wave numbers, from the biggest vortices (say, like a cyclone covering the whole continent) to a smaller one (like a whirl of dust on a street corner).

Then we can assume that the energy spectrum E, the energy flow ϵ and the wave number k are linked by an equation which does not involve anything else. Since the three quantities involved have completely different dimensions, we can combine them only by means of an equation of the form

$$E(k) \approx C\epsilon^x \cdot k^y.$$

Here C is a constant; since the equation should remain the same for small scale and for global scale events, the shape of the equation should not depend on the choice of units of measurements, hence C should be dimensionless.

[3] This formulation is a bit cruder than most experts would accept; I borrow it from Arnold [**2**].

Let us now check how the equation looks in terms of dimensions:

$$\frac{\text{length}^3}{\text{time}^2} = \left(\frac{\text{length}^2}{\text{time}^3}\right)^x \cdot \left(\frac{1}{\text{length}}\right)^y.$$

After equating lengths with lengths and times with times, we have

$$\begin{aligned} \text{length}^3 &= \text{length}^{2x} \cdot \text{length}^{-y} \\ \text{time}^2 &= \text{time}^{3x}, \end{aligned}$$

which leads to a system of two simultaneous linear equations in x and y,

$$\begin{aligned} 3 &= 2x - y \\ 2 &= 3x \end{aligned}$$

This can be solved with ease and gives us

$$x = \frac{2}{3} \quad \text{and} \quad y = -\frac{5}{3}.$$

Therefore we come to *Kolmogorov's "5/3" Law*:

$$E(k) \approx C\epsilon^{2/3}k^{-5/3}.$$

We have done it!

It is claimed in the specialist literature that the dimensionless constant C can be determined from experiments and happens to be pretty close to 1.

2.3. Discussion. The status of this celebrated result is quite remarkable. In the words of an expert on turbulence, Alexander Chorin [**7**],

> Nothing illustrates better the way in which turbulence is suspended between ignorance and light than the Kolmogorov theory of turbulence, which is both the cornerstone of what we know and a mystery that has not been fathomed.
>
> The same spectrum [...] appears in the sun, in the oceans, and in manmade machinery. The 5/3 law is well verified experimentally and, by suggesting that not all scales must be computed anew in each problem, opens the door to practical modelling.

Arnold [**2**] reminds us that the main premises of Kolmogorov's argument remain unproven—after more than 60 years! Even worse, Chorin points to the rather disturbing fact that

> Kolmogorov's spectrum often appears in problems where his assumptions clearly fail. [...] The 5/3 law can now be derived in many ways, often under assumptions that are antithetical to Kolmogorov's. Turbulence theory finds itself in the odd situation of having to build on its main result while still struggling to understand it.

2.4. History of dimensional analysis. It would be interesting to have an account of the history of dimensional analysis. It can be traced back at least to *Froude's Law of Steamship Comparisons* used to great effect in D'Arcy Thompson's book *On Growth and Form* [**21**, p. 24] for the analysis of speeds of animals:

> the maximal speed of similarly designed steamships is proportional to the square root of their length.

William Froude (1810–1879) was the first to formulate reliable laws for the resistance that water offers to ships and for predicting their stability. And here is the deduction for his law adapted from [**21**, p. 24].

We use the property of that, at small speeds, the drag F (force of resistance) offered by water to the ship is proportional to the area of cross section of the ship (and hence to the square L^2 of its typical linear size L) and to the velocity V of the ship. We shall write it symbolically as

$$F \sim L^2 V.$$

To sustain constant speed, the ship has to produce power FV. Now we have to recall that we are talking about *steamships*, powered by an engine inside, this engine being fed by coal, carried by the ship, etc. Therefore it is reasonable to assume that the power produced by the ship is proportional to its volume, and therefore

$$FV \sim L^3.$$

Combining the two proportions together, we have

$$L^2 V^2 \sim L^3$$

and therefore

$$V \sim L^{1/2}.$$

D'Arcy Thompson applied the same derivation to fish because of the validity of the principal assumption: the energy is produced internally.

Exercises.

Exercise 2.1. How would Froude's Law look for solar powered ships?

Exercise 2.2. Why does a mouse have (relatively) slimmer body build than an elephant?

Exercise 2.3. Prove the following corollary of Froude's Law:

> The *relative* speed of a fish (that is, speed measured in numbers of its lengths covered by the fish per unit of time) is *inverse proportional* to the square root of its length.

This explains a well-known phenomenon: little fish in a stream appear to be *very* quick.

Exercise 2.4. Estimate, who is relatively faster: an ant or a racehorse?

Exercise 2.5. Analyse the following quote from Greenhill [**12**, p. 233]:

> Thus treating a man's leg as swinging like a pendulum through a fixed definite angle, the length of a pace is proportional to his height, but the number of paces a minute is inversely as the square root of the height; so that his pace of walking and getting over the ground will vary as the square root of his height, as in Froude's Law again.

Using dimensional analysis, prove:

- the number of swings of a pendulum a minute is inversely as the square root of the length of the pendulum;

and use this fact to derive Greenhill's conclusion:

Figure 4. Walking on stilts increases speed (for Exercise 2.5).

- his pace of walking and getting over the ground will vary as the square root of his height (or of length of legs, see Figure 4).

Investigate further:

- how do the frequency of a pendulum's swings and the pace of a man's walking depend on the acceleration of gravity? (Read more on that in [**22**].)

Exercise 2.6. Building on ideas from Exercise 2.2, develop a method for estimation of a maximal possible height of a tree of given species.

The problem is of serious practical value. To explain it to an Englishman, you have to mention just one word: *Leylandii*. For a foreigner, *Wikipedia* provides an explanation:

> The Leyland Cypress, *X Cupressocyparis leylandii*, is often referred to as just Leylandii. It is a fast-growing evergreen tree much used in horticulture, primarily for hedges and screens.
>
> The Leyland Cypress is a hybrid between the Monterey Cypress, *Cupressus macrocarpa*, and the Nootka Cypress, *Cupressus nootkatensis*. The hybrid has arisen on nearly 20 separate occasions, always by open pollination. [. . .]
>
> Leyland Cypresses are commonly planted in gardens to provide a quick boundary or shelter hedge. However, their rapid growth (up to a metre per year), heavy shade and great potential height (often over 20 m tall in garden conditions, they can reach at least 35 m) make them a problem. In Britain they have been the source of a number of high profile disputes between neighbours, even leading to violence (and in one recent case, murder), because of their capacity to cut out light.

The problem is that no-one knows the maximal height of some of the latest hybrids—all known specimens continue to grow. . .

FIGURE 5. *X Cupressocyparis leylandii*, 35 meters and still growing.

3. Adding one by one

My colleague EHK[4] told me about a difficulty she experienced in her first encounter with arithmetic, aged 6. She could easily solve "put a number in the box" problems of the type

$$7 + \square = 12,$$

by counting how many 1's she had to add to 7 in order to get 12 but struggled with

$$\square + 6 = 11,$$

because she did not know where to start. Worse, she felt that she could not communicate her difficulty to adults. Her teacher forgot to explain to her that addition was commutative.

Another one of my colleagues, AB[5], told me how afraid she was of subtraction. She could easily visualise subtraction of 4 from 100, say, as a stack of 100 objects;

[4]For the record: EHK is female, has a PhD in Mathematics, teaches mathematics at a highly selective secondary school.

[5]AB is female, has a PhD in Mathematics, teaches mathematics in a research-led university.

after removing 4 objects from the top (by reverse counting: 100, 99, 98, 97), 96 are left. But what happens if we remove 4 objects from the bottom of the stack?

A brief look at axioms introduced by Dedekind (but commonly called Peano axioms) provides some insight in EHK's and AB's difficulties.

3.1. Dedekind-Peano axioms. Recall that the Dedekind-Peano axioms describe the properties of natural numbers $\mathbb{N}$ in terms of a "successor" function $S(n)$. (There is no canonical notation for the successor function, in various books it is denoted $s(n)$, $\sigma(n)$, n', or even n^{++}, as in popular computer languages C and C^{++}.)

Axiom 1: 1 is a natural number.

Axiom 2: For every natural number n, $S(n)$ is a natural number.

Axioms 1 and 2 define a unary representation of the natural numbers: the number 2 is $S(1)$, and, in general, any natural number n is

$$S^{n-1}(1) = S(S(\cdots S(1)\cdots)) \quad (n-1 \text{ times}).$$

As we shall soon see, the next two axioms deserve to be treated separately; they define the properties of this representation.

Axiom 3: For every natural number n other than 1, $S(n) \neq 1$. That is, there is no natural number whose successor is 1.

Axiom 4: For all natural numbers m and n, if $S(m) = S(n)$, then $m = n$. That is, S is an injection.

The final axiom (Axiom of Induction) has a very different nature and is best understood as a method of reasoning about all natural numbers.

Axiom 5: If K is a set such that:

- 1 is in K, and
- for every natural number n, if n is in K, then $S(n)$ is in K,

then K contains every natural number.

Thus, Dedekind-Peano arithmetic is a formalisation of that very counting by one that little EHK did, and addition is defined in precisely the same way as EHK learned to do it: by a recursion

$$\begin{aligned} m+1 &= S(m) \\ m+S(n) &= S(m+n). \end{aligned}$$

Commutativity of addition is a non-trivial (although still accessible to a beginner) theorem. To force you to feel some sympathy to poor little EHK and to poor little AB, I will prove it to you in the next section.

3.2. A brief digression: is 1 a number? Having postponed more serious work, we can spend a few minutes discussing Axiom 1: 1 *is a natural number.*

Even this axiom is not self-evident as it appears to be. In many languages, including English, the word 'number' can denote some collection or ensemble of objects with tacit understanding that it contains at least a few, and in any case more than one, objects. For example, a phrase

"A number of people feel that 1 is not a number"

makes sense and means that more than one person thinks that 1 is not a number. Such usage reflects an earlier stage of development of the system of numerals when 1 was not a number; numbers were made of ones, of basic units; but one is not made of ones.

FIGURE 6. Guido Reni. A fragment of ??? Musée du Louvre. Source: *Wikipedia Commons*. Public domain.

What is very important for history of mathematics, it appears that, for similar reasons, 1 was not a number for ancient Greek mathematicians, as evidenced in Euclid's *Elements* by careful separation of the use of words 'number' and 'unit'.

And, as a digression within digression, I want to mention the issue of collective nouns—I shall discuss them again in later lectures, so the present deviation is not waste of time. The English language has a peculiar tendency to form or find a special word to denote groups of particular animals. For example, Englishmen say

a herd of cows,
a flock of sheep,
a pack of dogs,
a school of fish.

To illustrate how far things go, it will suffice to mention that ducks, while on water, form a *paddling*, while in flight they are a *flush.* Some nouns are absolutely obscure; for example, I found in Wikipedia a *sedge of bitterns*, but I do not even know what bitterns are.

Invention of collective nouns for groups of people from various professional groups is a popular genre of English humor; to my taste,

a number of mathematicians

appears to be one of the more obvious solutions.

4. Properties of addition

There are several alternative forms of notation for the successor function: $S(n)$, $s(n)$, $\sigma(n)$, n' and even n^{++}, the latter used in programming languages C and

C++. I shall use notation n'; as the reader will soon see, it is very convenient—and natural—to write a symbol for the successor function *after* the number that has to be incremented.

In this new notation, the recursive rule for addition looks like

$$n + 1 = n' \tag{4}$$
$$n + m' = (n + m)'. \tag{5}$$

I will also use Axiom 5 in a more conventional form, obviously equivalent to the original one.

Axiom 5: Assume that a certain statement about numbers If K is a set such that:

- the statement is true for 1 (*Basis of Induction*)
- if the statement is true for a natural number n (*Inductive Assumption*) then it is true for the next number n' (*Inductive Step*).

Then the statement is true for all natural numbers.

I will prove two canonical properties of addition.

4.1. Associativity of addition.

THEOREM 4.1. *Assume that + is a binary operation which satisfies conditions* (4) *and* (5)*. Then + is associative, that is,*

$$(a + b) + c = a + (b + c)$$

for all a, b, c.

PROOF. The proof will use induction on c.
Basis of Induction.

$$\begin{aligned}(a + b) + 1 &\overset{\text{by (4)}}{=} (a + b)' \\ &\overset{\text{by (5)}}{=} a + b' \\ &\overset{\text{by (4)}}{=} a + (b + 1).\end{aligned}$$

Inductive Assumption:

$$(a + b) + c = a + (b + c).$$

Inductive Step.

$$\begin{aligned}(a + b) + c' &\overset{\text{by (5)}}{=} ((a + b) + c)' \\ &\overset{\substack{\text{by inductive}\\ \text{assumption}}}{=} (a + (b + c))' \\ &\overset{\text{by (5)}}{=} a + (b + c)' \\ &\overset{\text{by (5)}}{=} a + (b + c').\end{aligned}$$

□

4.2. Commutativity of addition. We shall start with a very special, but crucially important case.

THEOREM 4.2. *Assume that + is a binary operation which satisfies conditions* (4) *and* (5). *Then*

$$1 + a = a + 1$$

for all a.

PROOF. We shall prove the theorem by induction on a.

Basis of Induction.

$$1 + 1 = 1 + 1.$$

There is nothing to prove here.

Inductive Assumption:

$$1 + a = a + 1.$$

Inductive Step.

$$\begin{aligned} 1 + a' &\overset{\text{by (5)}}{=} (1 + a)' \\ &\overset{\substack{\text{by inductive}\\ \text{assumption}}}{=} (a + 1)' \\ &\overset{\text{by (4)}}{=} (a')' \\ &\overset{\text{by (4)}}{=} a' + 1. \end{aligned}$$

□

THEOREM 4.3. *Assume that + is a binary operation which satisfies conditions* (4) *and* (5). *Then + is commutative, that is,*

$$a + b = b + a$$

for all a and b.

PROOF. We shall prove the theorem by induction on b.

Basis of Induction: Theorem 4.2.

Inductive Assumption:

$$a + b = b + a.$$

Inductive Step.

$$\begin{aligned} a + b' &\overset{\text{by (5)}}{=} (a + b)' \\ &\overset{\substack{\text{by inductive}\\ \text{assumption}}}{=} (b + a)' \\ &\overset{\text{by (4)}}{=} (b + a) + 1 \\ &\overset{\text{by Theorem 4.1}}{=} b + (a + 1) \\ &\overset{\text{by Theorem 4.2}}{=} b + (1 + a) \\ &\overset{\text{by Theorem 4.1}}{=} (b + 1) + a \\ &\overset{\text{by (4)}}{=} b' + a. \end{aligned}$$

□

5. Dark clouds

Notice that I was careful to formulate Theorems 4.1–4.3 in the most cautious way, by emphasising their conditional nature:

> *if* $+$ *is a binary operation which satisfies conditions* (4) *and* (5) *then ...*

The reason for my restraint is that writing does not mean to define a function. If you look at the proofs of Theorems 4.1–4.3, you notice that they do not refer to Axioms 3 and 4 and are based entirely on the Induction Axiom, Axiom 5.

David Pierce [**20**] makes an incisive comment:

> Indeed, if one thinks that the recursive definitions of addition and multiplication—
>
> $$\begin{aligned} n+0 &= n, \\ n+(k+1) &= (n+k)+1; \\ n \cdot 0 &= 0, \\ n\cdot(k+1) &= n\cdot k + n \end{aligned}$$
>
> —are *obviously* justified by induction alone, then one may think the same for exponentation, with
>
> $$\begin{aligned} n^0 &= 1 \\ n^{k+1} &= n^k \cdot n. \end{aligned}$$
>
> However, while addition and multiplication are well-defined on $\mathbb{Z}/n\mathbb{Z}$ (which admits induction), exponentiation is not; rather, we have
>
> $$\begin{aligned} (x,y) &\mapsto x^y \\ (\mathbb{Z}/n\mathbb{Z})^* \times \mathbb{Z}/\phi(n)\mathbb{Z} &\to \mathbb{Z}/n\mathbb{Z}, \end{aligned}$$

where $(\mathbb{Z}/n\mathbb{Z})^*$, as usual, denotes the group of invertible elements of the residue ring $\mathbb{Z}/n\mathbb{Z}$.

Indeed, the recursive definition of exponentiation fails in $\mathbb{Z}/3\mathbb{Z}$:

n	n^2	n^3	$n^3 \times n$	n^4
2	1	2	1	2

but holds in $\mathbb{Z}/6\mathbb{Z}$:

n	n^2	n^3	n^4	n^5	n^6	n^7
1	1	1	1	1	1	1
2	4	2	4	2	4	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	1	5	1	5	1	5
6	6	6	6	6	6	6

The former is an exception rather than rule, as clarified by David Pierce's theorem.

THEOREM 5.1. David Pierce [**20**]) *The identities*

$$a^1 = a, \qquad a^{b+1} = a^b \times a \tag{6}$$

hold on $\mathbb{Z}/n\mathbb{Z}$ *if and only if* $n \in \{0,1,2,6\}$.

I share David Pierce's indignation at the sate of affairs [**20**]:

> Yet the confusion continues to be made, even in textbooks intended for students of mathematics and computer science who ought to be able to understand the distinction. Textbooks also perpetuate related confusions, such as suggestions that induction and 'strong' induction (or else the 'well-ordering principle') are logically equivalent, and that either one is sufficient to axiomatize the natural numbers. [...]
>
> This is one example to suggest that getting things straight may make a pedagogical difference.

But I have to admit that I shared the widespread ignorance until David Pierce brought my attention to the issue—despite the fact that, in a calculus course that I took in the first year of my university studies, the lecturer (Gleb Pavlovich Akilov) explicitly proved the existence of a function of natural argument defined by a recursive scheme [**1**].

To save our theory from collapse, in the next section we shall prove the existence of addition.

Exercises.

Exercise 5.1. Prove Theorem 5.1 for prime values of n. You may wish to use Fermat's Theorem:

> If p is a prime integer and $0 < a < p <$ then
> $$a^p \equiv a \bmod p.$$

Exercise 5.2. Then try to prove Theorem 5.1 in full generality.

6. Landau's proof of the existence of addition

I decide to borrow *verbatim* a proof of the existence of addition from Edmund Landau's famous book *Grundlagen der Analysis* [**17**]. Also, I picked up from his book notation

$$S(n) = n'$$

which I had already used in my proofs. Although this is not emphasised by Landau, the proof of consistency of addition is not using Axioms 3 and 4. Are these axioms of any use at all? We shall return to this question later.

Theorem 6.1. [**17**, Theorem 4] *To every pair of numbers x, y, we may assign in exactly one way a natural number, called $x + y$, such that*

(1) $x + 1 = x'$ *for every* x,
(2) $x + y' = (x + y)'$ *for every* x *and every* y.

Proof. (A) First we will show that for each fixed x there is at most one possibility of defining $x + y$ for all y in such a way that $x + 1 = x'$ and $x + y' = (x + y)'$ for every y.

Let a_y and b_y be defined for all y and be such that

$$a_1 = x', \quad b_1 = x', \quad a_{y'} = (a_y)', \quad b_{y'} = (b_y)' \text{ for every } y.$$

Let $\mathcal{M}$ be the set of all y for which

$$a_y = b_y.$$

(I) $a_1 = x' = b_1$; hence 1 belongs to $\mathcal{M}$.

(II) If y belongs to $\mathcal{M}$, then $a_y = b_y$, hence by Axiom 2,

$$(a_y)' = (b_y)',$$

therefore

$$a_{y'} = (a_y)' = (b_y)' = b_{y'},$$

so that y' belongs to $\mathcal{M}$.

Hence $\mathcal{M}$ is the set of all natural numbers; i.e. for every y we have $a_y = b_y$.

(B) Now we will show that for each x it is actually possible to define $x+y$ for all y in such a way that

$$x + 1 = x' \text{ and } x + y' = (x + y)' \text{ for every } y.$$

Let $\mathcal{M}$ be the set of all x for which this is possible (in exactly one way, by (A)).
(I) For $x = 1$, the number $x + y = y'$ is as required, since
$x + 1 = 1' = x'$,
$x + y' = (y')' = (x + y)'$.
Hence 1 belongs to $\mathcal{M}$.
(II) Let x belong to $\mathcal{M}$, so that there exists an $x+y$ for all y. Then the number $x' + y = (x + y)'$ is the required number for x', since

$$x' + 1 = (x + 1)' = (x')'$$

and

$$x' + y' = (x + y')' = ((x + y)')' = (x' + y)'.$$

Hence x' belongs to $\mathcal{M}$. Therefore $\mathcal{M}$ contains all x. □

Landau's book is characterised by a specific austere beauty of entirely formal axiomatic development, dry, cut to the bone, streamlined. Not surprisingly, it is claimed that logical austerity and precision were Landau's characteristic personal traits.[6]

Grundlagen der Analysis opens with two prefaces, one intended for the student and the other for the teacher; we already quoted *Preface for the Teacher*, it is a remarkable pedagogical document. The preface for the student is very short and begins thus:

> 1. Please don't read the preface for the teacher.
> 2. I will ask of you only the ability to read English and to think logically-no high school mathematics, and certainly no higher mathematics. [...]
> 3. Please forget everything you have learned in school; for you haven't learned it.
>   Please keep in mind at all times the corresponding portions of your school curriculum; for you haven't actually forgotten them.
> 4. The multiplication table will not occur in this book, not even the theorem,
>
> $$2 \times 2 = 4,$$

[6] Asked for a testimony to the effect that Emmy Noether was a great woman mathematician, Landau famously said: "I can testify that she is a great mathematician, but that she is a woman, I cannot swear."

> but I would recommend, as an exercise for Chap. I, section 4, that you define
>
> $$\begin{aligned} 2 &= 1+1, \\ 4 &= (((1+1)+1)+1), \end{aligned}$$
>
> and then prove the theorem.

Exercises.

EXERCISE 6.1. Follow Edmund Landau's advise and prove from the axioms of Peano arithmetic that

$$2 \times 2 = 4.$$

7. Numbers in computer science

I quote a professional computer scientist who left the following comment on my blog:

> *I would caution everyone ... not to confuse "mathematical thinking" with "The thinking done by computer scientists and programmers".*
>
> *Unfortunately, most people who are not computer scientists believe these two modes of thinking to be the same.*
>
> *The purposes, nature, frequency and levels of abstraction commonly used in programming are very different from those in mathematics.*

I would like to illustrate the validity of these words on a very simple example.

Let us look at a simple calculation with MATLAB, an industry standard software package for mathematical computations.

```
>> t= 2

t =      2

>> 1/t

ans =     0.5000
```

What is shown in this screen dump is a basic calculation which uses floating point arithmetic for computations with rounding.

Next, let us make the same calculation with a different kind of integers:

```
>> s=sym('2')

s =     2

>> 1/s

ans =   1/2
```

Here we use "symbolic integers", designed for use as coefficients in symbolic expressions.

Next, let us try to combine the two kinds of integers in a single calculation:


```
>>  1/(s+t)

ans =   1/4
```

We observe that the sum $s + t$ of a floating point number t and a symbolic integer s is a symbolic integer.

Examples involving analytic functions are even more striking:

```
>> sqrt(t)

ans =       1.4142

>> sqrt(s)

ans =       2^(1/2)

>> sqrt(t)*sqrt(s)

ans =       2
```

We see that MATLAB can handle two absolutely different representations of integers, remembering, however, the intimate relation between them.

The example above is written in C++. When represented in C++, even the simplest mathematical objects and structures appear in the form of (a potentially infinite variety of) *classes* linked by mechanisms of *inheritance* and *polymorphism*. This is a manifestation of one of the paradigms of the computer science: if mathematicians instinctively seek to build their discipline around a small number of "canonical" (actually) infinite structures, computer scientists frequently prefer to work with a host of similarly looking structures. We shall look in the next lectures how they keep control of their bestiary. For a time being, we have to notice that Dedekind-Peano axioms allow for a variety of realisations.

Acknowledgements

I am grateful to my correspondents EHK and AB for sharing with me their childhood memories, to parents of DW for allowing me to write about the boy, to David Pierce for many comments and for a permission to use his paper [**20**], to Sergei Utyuzhnikov for explaining to me some points related to turbulence, to Mikael Johansson who brought my attention to the cohomological nature of carries, and to Hovik Khudaverdyan and Gábor Megyesi for an example of zero divisors in the ring of 10-adic integers.

This text would not appear without kind invitations to give a talk at "Is Mathematics Special" conference in Vienna in May 2008 and to give a lecture course "Elementary mathematics from the point of view of "higher" mathematics" at Nesin Mathematics Village in Şirince, Turkey, in July 2008.

My work on these lectures was partially supported by a grant from the John Templeton Foundation, a charitable institution which describes itself as a *"philanthropic catalyst for discovery in areas engaging in lifes biggest questions."* However, the opinions expressed in the book are those of the author and do not necessarily reflect the views of the John Templeton Foundation.

Finally, my thanks go to the blogging community—I have picked in the blogosphere some ideas and quite a number of references—and especially to numerous (and mostly anonymous) commentators on my blog, especially to my old friend who prefers to be known only as Owl.

References

[1] G. P. Akilov and V. N. Dyatlov, **Elements of Functional Analysis**, Novosibirsk State University,1979. (In Russian.)

[2] V. I. Arnold, **What is Mathematics?** Moscow, MTSNMO, 2004.

[3] R. Ball, The Kolmogorov cascade, an entry in **Encyclopedia of Nonlinear Science** (A. Scott, ed.). Routledge, 2004. (The text is availabe at http://wwwrsphysse.anu.edu.au/∼rxb105/cascade.pdf)

[4] D. Birmajer and J. B. Gill, Arithmetic in the ring of formal power series with integer coefficients, The American Mathematical Monthly 115 no. 6 (2008) 541–549.

[5] A. V. Borovik, **Mathematics under the Microscope**, submitted; available for free download from http://www.maths.manchester.ac.uk/∼avb/micromathematics/downloads

[6] P. Bryant and S. Squire, The influence of sharing on children's initial concept of division, J. Experimental Child Psychology 81 no. 1 (January 2002) 1–43.

[7] A. J. Chorin, Book Review: Kolmogorov spectra of turbulence I: Wave turbulence, by V. E. Zakharov, V. S. Lvov, and G. Falkovich. Bull. Amer. Math. Soc. 29 no. 2 (1993) 304-306.

[8] R. Cowan and S. Biddle, Children's understanding of one-to-one correspondence in the context of sharing, Educational Psychology, 9 no, 2 (1989)133–140.

[9] R. Dedekind, **Essays on the theory of numbers. I: Continuity and irrational numbers. II: The nature and meaning of numbers**, Dover, 1963.

[10] P. Doyle and J. Conway, Division by three, http://arxiv.org/abs/math.LO/0605779.

[11] A. Gardiner, **Mathematical puzzling**, UK Mathematics Foundation, Birmingham, 1996. ISBN0-7044-17545.

[12] G. Greenhill, Mechanical Similitude, The Mathematical Gazette 8 no. 122 (Mar., 1916) 229–233.

[13] L. Henkin, On mathematical induction, The American Mathematical Monthly, 67 (1960), no. 4, 323–338.

[14] A. N. Kolmogorov, Local structure of turbulence in an incompressible fluid for very large Reynolds numbers, Doklady Acad Sci. USSR 31 (1941) 301–305.

[15] V. A. Krutetskii, **The Psychology of Mathematical Abilities in Schoolchildren**. (Translated from Russian by J. Teller, edited by J. Kilpatrick and I. Wirszup.) The University of Chicago Press, 1976. (ISBN 0-226-45485-1)

[16] G. Lakoff and R. Núnez, **Where Mathematics Comes From: How the Embodied Mind Brings Mathematics Into Being**. Basic Books, New York, 2000.

[17] E. Landau, **Foundations of Analysis**, Chelsea, 1966.

[18] R. Núnez, Do *real* numbers really move? Language, thought and gesture: the embodied cognitive foundations of mathematics. In: **18 Unconventional Essays on the Nature of Mathematics** (R. Hersh, ed.). Springer, 2005, pp. 160–181.

[19] G. Peano, The principles of arithmetic, presented by a new method, in **From Frege to Gödel**, (Jean van Heijenoort, editor), Harvard, 1967.

[20] D. Pierce, Induction and recursion. A talk at Logic Colloquium 2008—ASL European Summer Meeting, Bern, Switzerland, 3–8 July 2008.

[21] D'A. W. Thompson, **On Growth and Form**. Cambridge University Press, 1961.

[22] C . Vaughan and M . O'Malley, Froude and the contribution of naval architecture to our understanding of bipedal locomotion, Gait and Posture 21 no. 3 350–362.

[23] F. Viéte, **The Analytic Art**, translated by T. Richard Witmer. Kent, Ohio: The Kent State University Press, 1983.

