

Foreword

For a book to be interesting to read, it must have a plot—even when its subject is mathematics. Furthermore, a mathematics book for nonmathematicians is exactly the sort of work that should be written so that even the professional mathematician will find the developments of the plot interesting to follow. And such books should also be very well written, even of literary worth. The book you hold in your hands has all these virtues.

It consists of ten mathematical “novellas”, united by recurring “characters”. For example, the main character of the discussion of induction, of which the first novella is devoted to providing a firm foundation, makes an appearance in practically every succeeding one in order to play its assigned part. On the other hand, the set of natural numbers represents the first example of the so-called “number systems” developed successively in novellas 3, 8, and 9. Such internal connections are a standard feature of the book.

It is difficult to define the genre of this book. It is not a problem book, nor a textbook, nor a “book for reading about mathematics”. It is most of all reminiscent of a good lecture course, from which a thoughtful student comes away with more than was actually spoken about in the lectures. I like how each chapter/novella begins with elementary problems serving as a basis for the succeeding exposition, and how the author always knows when he should stop in expounding theoretical material. I like the arrangement and character of the multitude of exercises, which, on the one hand, make the book easier to read, and, on the other, provide the reflective reader with the opportunity of testing his or her understanding of the subject matter.

The “elementary mathematics” that this book deals with should be understandable by students of high schools specializing in physics and mathematics. Above all, however, I would recommend that all teachers in such schools saturate themselves in its ideas, since one should not confuse the profound inculcation of mathematics with the teaching of “higher” mathematics at technical institutes.

I would wish all students in mathematical departments of institutes of teacher-education to study this book, since that person cannot make a good mathematics teacher for whom “school” mathematics ends with the solution of various kinds of equations and inequalities by means of routine or cunning transformations, and “higher” mathematics has been reduced to a collection of more and more complicated concepts and constructions only distantly related to what he or she will have to expound to their classes.

It seems to me that this “elementary mathematics” book will be of interest and use also to university teachers—especially those working with lower-year students—as illustrating how to introduce new concepts to the students by demonstrating their

inevitability, providing them with exercises in their use, and without overwhelming them with technical details.

I hope that this book is destined for a long life—which, of course, it certainly deserves.

*A. S Merkurjev,
Doctor of physico-mathematical sciences,
Professor at the University of California at Los Angeles.*

Introduction:

What This Book Is About and For Whom It Is Intended

“Now drop that! When I say I’ll learn a man the river, I mean it. And you can depend on it, I’ll learn him or kill him.”*

**“Teach” is not in the river vocabulary.
From *Life on the Mississippi* by Mark Twain.*

In 1995 the St. Petersburg University publishing house issued the author’s book *Selected Chapters from Elementary Mathematics* [11].¹ Unfortunately, the Russian printing was very small, and the author long nourished the project of preparing a second edition. However, notwithstanding all the virtues of *Selected Chapters*, it did have crucial defects. The fact is that it was based on notes of a lecture course for teachers-to-be already possessing a degree. For this reason the author felt free to use, for example, the concept of the determinant in the definition of the resultant of two polynomials, and to refer to Fubini’s theorem in calculating the volume of a 4-dimensional sphere. Although most of *Selected Chapters* should certainly have been accessible to more senior students, all the same the transitions from elementary problems to questions of “higher mathematics” would very likely have defeated them—and the basic structural idea of the book thus be hidden from them. Hence, in thinking over the plan of a new version, the author decided above all to write it so as to be accessible to teachers and upper-year students in high schools providing enhanced instruction in mathematics. Although the titles of six of the chapters, as well as some of the material they contain, remain as before, the present new “elementary mathematics” text is a completely different book.

Don’t be misled by the word “elementary”; after all, the book is about *mathematics*. It is highly regrettable that what is studied at school under that name is no more like mathematics than a plucked flower gathering dust for many years in a herbarium or pressed between the pages of a book is like that same flower in the meadow, besprinkled with dewdrops sparkling in the rays of the rising sun. As the author understands it, the term “elementary” signifies that an understanding of the material in question requires neither a developed ability to reason abstractly nor skill in using the refined techniques of mathematical analysis. In those places in the book where it is natural to introduce concepts outside the high school syllabus, the requisite definitions are given, with examples. And in order to bring out the meanings of such concepts clearly, appropriate (but not too many) theorems are proved concerning them.

¹An English translation entitled *Easy as π ?* appeared in 1999, and an Italian translation in 2000.

Thus if you are acquainted with mathematics, at least to the extent of a standard high school course, and you like it enough to want to know more, and if, in addition, you are prepared to do some serious work, then you may boldly begin to study elementary mathematics as it is represented in this book.

Content, basic ideas, and structure

The book consists of ten chapters, most of which bear titles that are completely traditional for textbooks intended for use in elective mathematics courses. In each chapter elementary problems are considered, accompanied by theoretical material directly related to them. There are over 300 problems in the book, most of which are intended to be solved independently by the reader. (Their solutions are given at the back of the book.)

It must be mentioned that the author's pedagogical style was developed under the influence of the celebrated books by George Pólya and Felix Klein.² The basic aim of this book is to develop the reader's ability to think mathematically. Although, as already mentioned, the author does not presuppose that the reader possess any skills in the manipulation of abstract concepts and in their use in logical reasoning, such concepts *are* introduced in various chapters of the book. A typical example is the introduction in Chapter 1 ("Induction") of Peano's axioms in order to define the set of natural numbers, and, in particular, to provide a foundation for the method of mathematical induction. In the other direction, it often happens that a little theory is useful in bringing out the significance of certain of the problems. A very simple example of this is afforded by the connection between Problem 3.5 and Fermat's Little Theorem (Theorem 3.8). It is useful to mention also that what that theorem asserts becomes clearer once it has been formulated in terms of the notion of congruence of numbers with respect to a given modulus. Finally, it is always of interest to contemplate the unexpected applications of a theory that has been developed for other reasons. Chapter 3 contains a description of the so-called "public-key codes", which at the present time represent the most popular and secure tool in cryptography. The proof in this context of the feasibility of encoding and decoding is, as it happens, based on classical theorems of number theory: Fermat's Little Theorem and Euler's theorem (3.9) generalizing it.

Practically every page of the book contains a variety of exercises (mostly of a theoretical character); they number over 200. Their primary role is to develop the reader's mathematical culture. In many cases the solution of an exercise constitutes a step in the proof of some statement or other. However, in solving it you are not only filling a gap in that proof, but also acquiring a better understanding of the concepts involved. Furthermore, in the author's opinion, the exercises foster understanding of the proofs, which in some instances are rather lengthy. For the same reason the chains of argument are often broken up into several steps. In any case, it is not at all necessary that all details be thoroughly understood at a first reading,³ although it is important to understand at least the statements of the theorems. The reader wishing to understand the proofs fully may find it easier to

²George Pólya, *How to solve it*. Second Edition, Princeton Univ. Press, 1957, and *Mathematics and plausible reasoning*, Vols. I, II, Princeton Univ. Press, 1954. Felix Klein, *Elementarmathematik vom höheren Standpunkte aus* (Elementary mathematics from an advanced viewpoint). Leipzig, 1908.

³Although there are those personalities who like to work painstakingly through a book, understanding everything as they go. *Trans.*

do so at a second reading. It is also recommended that the reader compare his or her solutions with the author's, reproduced at the end of each chapter.

What has been said above concerning the exercises applies also to the book as a whole. For instance, in each chapter the reader will inevitably encounter sections difficult to understand. He or she should omit these at a first reading, or rather read the statements of the theorems carefully without immediately attempting to work through their proofs. As can be seen from the table of contents, the chapters are only slightly interrelated, so can be read in practically any order. There are, of course, *some* logical links between certain chapters. For example, in order to read the basic part of Chapter 4 (“Geometric transformations”), one must already have become acquainted with the concept of a “group”, introduced in the final section of the preceding chapter. Although every mathematics student knows what set the symbol \mathbb{Z} stands for,⁴ the reader should not rush to omit §3.9 (“The set of whole numbers”), where by means of a very simple and natural example the concept of a factor set is introduced, since this concept is needed in the sequel—for example, in the definition of a ring of residues.

Among the theorems appearing in the book, there are those that may be called mathematical *chefs d'oeuvre*, such as the following:

- Euler's theorem on the number of partitions of a natural number as a sum of natural numbers (§2.6);
- Euler's theorem (Theorem 3.13) on the frequency of occurrence of prime numbers in the sequence of natural numbers, and Gauss's theorem on the cyclicity of the group of units of the ring of residues relative to a prime modulus;
- Euler's formula (Theorem 6.2) for planar graphs, and Hilbert's proof of the Jordan curve theorem (Theorem 6.18) in the case of a polygon;
- Ramsey's theorem (Theorem 7.5), Minkowski's lemma (Theorem 7.12), and proofs of the theorems of Euler (Theorem 7.14) and Lagrange (Theorem 7.15) on the representation of a natural number as a sum of two and four squares of natural numbers;
- Gauss's proof of the constructibility of a regular 17-gon by means of straightedge and compasses (Theorem 8.9); d'Alembert's proof of “the fundamental theorem of algebra” (Theorem 8.15) to the effect that every nonconstant polynomial has a (complex) root; Euler's formula for a number raised to a complex exponent (§8.7).

The following constructions and statements might be called “mathematical gems”:

- The construction of public-key codes (§3.8);
- The classification of ornaments (Theorem 4.19);
- Muirhead's theorem (Theorem 5.4);
- Pick's formula for the area of a polygon with its vertices at points of the integer lattice (Theorem 6.17);
- The construction of an efficient sorting using binary heaps (§6.8);
- Dilworth's lemma (Theorem 7.3), the theorems of Erdős and Szekeres (Theorem 7.2) and Sperner (Theorem 7.4);

⁴Just in case, here is the definition: \mathbb{Z} stands for the set $\{\dots, -3, -2, -1, 0, 1, 2, 3, 4, \dots\}$ of all positive and negative integers, together with zero. *Trans.*

- Connes’ proof of Morley’s theorem (§8.5 and Theorem 8.11) on the points of intersection of the trisectors of the angles of an arbitrary triangle.

However, it must be stressed once more that most of these theoretical facts, constructions, and theorems are of an essentially elementary character. Ramsey’s theorem, for example, is an extension of the well-known proposition that among any set of six people there must necessarily be either three mutually acquainted or else three mutually unacquainted people. The idea behind the construction of a regular 17-gon can be motivated by the analogous construction of a regular pentagon. For solving the combinatorial problems considered in Chapter 2, the concept of the derived function turns out to be useful. In the same chapter the ring concept makes an appearance (in the guise of a ring of formal power series), to appear again in the next chapter (“The whole numbers”). The expediency of introducing the group concept is clearly seen in Chapter 4, where groups of geometric transformations are considered—and so on.

Chapter 10 (“Mathematics and the computer”) occupies a special place in the book. The author has always been interested in the question as to how modern symbolic computing packages such as *Maple* and *Mathematica* might be used to good effect in the teaching of mathematics. For instance, it would seem perfectly senseless to use the operator `Plot[]` just for the sake of looking at a few dozen graphs, or the operator `Solve[]` to solve a few dozen equations. On the other hand, it would be inappropriate to “tie the hands” of students by concealing from them the possibilities represented by such computer packages. The results of the author’s ruminations on this question are reflected in this chapter. The basic idea there is expressed by the proposition that the solving of a problem begins only *after* its solution has appeared on the computer screen.

In conclusion, a few words about the system of numeration of problems, statements, and exercises used in the book. The symbols used as labels consist of two numerals separated by a dot. The first numeral indicates the chapter. The second indicates the order of the item within that chapter, with theoretical statements (theorems and lemmas), problems, and exercises numbered independently.

Sources and recommended literature

A list of works used by the author can be found at the back of the book. A source is referred to in the text only if it was used in an essential way. For example, the exposition of discrete fast Fourier transforms (§8 of Chapter 8) is based on material from §32 of the book [5], and the proofs of the theorems of Euler and Lagrange on the representation of a number as a sum of two and four squares are taken from the book [20].

As for the problems, only a small number were thought up by the author.⁵ The main sources of problems were the books [3], [7], [24], and [30], and Problem 10.2 was taken from the book [21]. However, as far as most of the problems in the book are concerned, it would be impossible to say who invented them or where they were first published. To those who have led mathematical problem-solving groups, they have been “well known for a long time”.

Of the books listed as sources, the author can recommend the following ones for further, or parallel, study: [1], [3], [7], [13], [14], [21], [22], [23], [24], [26], [28], and [29].

⁵In the sense that neither the author nor his colleagues had encountered them earlier.

Acknowledgments

Much in our lives is determined by our circle of acquaintance. The author is grateful to the following among his colleagues in the mathematics-mechanics faculty of St. Petersburg State University for our many discussions about mathematics and mathematics teaching, and for their help, support and constructive criticism: S. M. Ananyevskii, B. M. Bekker, V. G. Bykov, V. M. Gol'khovoĭ, and N. A. Vavilov. The author was greatly helped by G. I. Vol'fson, a graduate of SPbSU and a teacher at Gymnasium No. 261, and by Lev Spodyneĭko, a graduate of the Academic Gymnasium of SPbSU, both of whom read the book carefully in manuscript form. The author also thanks A. L. Gromov, associate professor in our faculty, and V. R. Meshkov for preparing the drawings. Special thanks are due to A. S. Merkurjev, our former colleague and now a professor at the University of California, for writing the preface to the book and arranging for the author to have the use of the computing package *Mathematica*.

The author would also like to thank Robert Burns, the translator. Our discussions (by email)—not to mention his corrections of various mistakes, inaccuracies, and misprints—afforded me the greatest pleasure.

Last but not least, the author cannot but mention the role played in his life by the Physics and Mathematics School (now Gymnasium) No. 30. What *real* mathematics is was made clear to me there, and all in the short span of the two years that I and many of my colleagues had the great good fortune to study there. However, it is appropriate to speak of this in greater detail, especially in view of the fact that when the author was writing this book there arrived the bitter news that Iosif Yakovlevich Verebeĭchik had quit this life—Iosif Yakovlevich, who for many of our generation will always remain in memory as *our teacher*.

In memory of a teacher

Iosif Yakovlevich Verebeĭchik was a teacher-legend. It is difficult to believe that he worked in “No. 30” for only a little over ten years, yet managed to educate almost a whole generation of the mathematicians of our city. In the mathematics-mechanics faculty of SPbSU alone, of his former students there are two professors in the department of higher algebra and number theory, two professors in the department of higher geometry, two professors in the department of theoretical cybernetics, and professors in the departments of mathematical analysis, probability and statistics, and general mathematics and information technology—and how many more are working in other tertiary institutions and in other countries!

This all started in Grade 9. We—former top students at ordinary high schools—were happy if we received a grade of “three minus minus”.⁶ The book we always kept by us was “Modenov” [20], we studied with great satisfaction axiomatic geometry “à la Hilbert” [7], and learned the theory of area from V. A. Rokhlin’s article in the *Encyclopaedia of Elementary Mathematics*.⁷ And when Iosif Yakovlevich was unable to extricate himself from difficulties in defining the real numbers in terms of their infinite decimal expansions, the author went and studied Dedekind sections as expounded by Fikhtengol’ts. We solved problems, and participated in

⁶In the Soviet and Russian educational system the grading scheme has always been out of 5, with 3 representing a bare pass, and less than 3 failure. *Trans.*

⁷Vol. 5. *Geometry*. M.: Nauka, 1966.



Iosif Yakovlevich Verebeichik (23.03.1921–12.03.2007)

“mathematical battles” invented by Iosif Yakovlevich. And even if only a few of my classmates became professional mathematicians, we shared forever those qualities that Verebeichik had imbued us with: responsibility and the ability to work hard, precision of reasoning, openness to new ideas, and faith in our own ability—whoever managed to survive need fear no more!

Of course, at No. 30 there were other—I won’t be afraid to use the word—outstanding teachers, but at that time I. Ya. Verebeichik was the leader. “Everything returns to its origins”, and indeed our school has been restored to its former site at the corner of Seventh Line and Middle Avenue. “The chain of time has not been broken”, and our (still *our!*) school lives and educates, and one of the most important links in that chain is engraved with the name of Iosif Yakovlevich Verebeichik.

St. Petersburg, 2008

CHAPTER 1

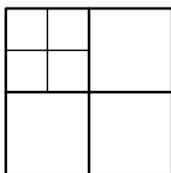
Induction

1.1. Reasoning “by induction”

As the word “induction” is usually understood, it means a type of reasoning for inferring the general from the particular. Generally speaking, this is also what it signifies from a mathematical point of view.

PROBLEM 1.1. Prove that a square can be subdivided into six, eight, or nine smaller squares. Into what other numbers of squares can a square be subdivided?

Answer: Into any number except 2, 3 and 5. It is indeed not very difficult to subdivide a square into six, eight, or nine smaller squares, and subdividing it into seven is very easy: just draw a big plus sign in the original square, and then do the same to one of the four quarters thus obtained (as in the diagram).



In fact this is the basic idea of the solution: by inserting a plus sign, the number of smaller squares increases by 3. Therefore, since we can dissect the square into six subsquares, we can dissect it into nine, and then twelve, fifteen subsquares, and so on. Having subdivided the square into four subsquares, we can in the same way then obtain subdivisions into seven, ten, thirteen, etc., squares. Finally, we can subdivide our square into eight, eleven, fourteen, etc., subsquares. Thus by adding 3 enough times to the numbers 6, 7 and 8, we can obtain all natural numbers exceeding 5. Hence it only remains to show that it is impossible to subdivide a square into five smaller squares (it being obvious that it cannot be subdivided into two or three squares).

In the initial stages (especially for students in the lower grades) it is appropriate to limit the reasoning to this informal level.

The following problem yields to similar reasoning.

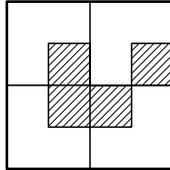
PROBLEM 1.2. Prove that you can pay any sum of money larger than seven pennies even if you have in your pocket only threepenny and fivepenny coins (although, of course, in unlimited quantities...).

For, since $8 = 3 + 5$, $9 = 3 + 3 + 3$, and $10 = 5 + 5$, we can pay eight, nine, or ten pennies exactly. Hence by adding the appropriate number of threepenny coins, we can come up with any sum over seven pennies.

One more problem of this type:

PROBLEM 1.3. Prove that if a square $2^n \times 2^n$ “chessboard” has any one of its cells removed, then the remainder of the board can be tiled with “corner” tiles of the form .

If a cell is removed from a 2×2 board, then what remains is a single corner tile. Next consider a 4×4 board with one cell removed. Subdivide the board into four 2×2 boards and place a corner tile so that it does not overlap with the square that contained the cell that was removed (as in the diagram). Then what remains can be tiled with four corner tiles!



The argument now continues in analogous fashion. If we subdivide the board into four equal squares and place a corner tile so that it does not overlap with that one of the four squares that contained the cell that was removed, but overlaps each of the other three, then we have left to tile four squares of half the original size, each of which contains exactly one cell that is not to be tiled over (since it is either removed or already tiled over). But we already know how to tile such squares!

As you can see, in each of the problems considered, the reasoning went from a particular case to the general case, from an examination of particular examples to a general idea allowing *unlimited* transition from one case to the next.

The next example is interesting in that at first glance it would appear that the inductive method of reasoning does not apply.

PROBLEM 1.4. Prove that if $a \geq b \geq c \geq d \geq e$, then

$$a^2 - b^2 + c^2 - d^2 + e^2 \geq (a - b + c - d + e)^2.$$

It can be difficult to solve a problem that is actually a special case of a general fact yet to be discovered. However, in the present case we can first try to solve a simpler problem. Thus suppose $a \geq b \geq c$. Then

$$\begin{aligned} a^2 - b^2 + c^2 - (a - b + c)^2 &= (a - b)(a + b) - (a - b)(a - b + 2c) \\ &= (a - b)(2b - 2c) = 2(a - b)(b - c) \geq 0, \end{aligned}$$

whence $a^2 - b^2 + c^2 \geq (a - b + c)^2$.

Hence

$$\begin{aligned} a^2 - b^2 + c^2 - d^2 + e^2 &\geq (a - b + c)^2 - d^2 + e^2 \\ &\geq (a - b + c - d + e)^2, \end{aligned}$$

since clearly $a - b + c \geq d$.

EXERCISE 1.1. Find a statement generalizing Problem 1.4.

To solve the final problem of this section, a little more in the way of mathematical technicalities is required; however, the solution brings out the “method of mathematical induction” especially clearly.

PROBLEM 1.5. Find the least number c such that the inequality $n^3 \leq c \cdot 2^n$ holds for all n .

Clearly, the number c is simply the largest term of the sequence $(x_n) = (\frac{n^3}{2^n})$. To begin with, let's calculate a little. The first few terms of this sequence are $\frac{1}{2}, 2, \frac{27}{8}, 4, \frac{125}{32}, \frac{216}{64} = \frac{27}{8}$. It would appear that, beginning with the fourth term, the sequence is decreasing; that is, each term from $x_4 = 4$ on is greater than its successor. If we can prove this, then it will follow that x_4 is the largest term, so that $c = 4$. Multiplying the (hoped-for) inequality $x_n = \frac{n^3}{2^n} > \frac{(n+1)^3}{2^{n+1}} = x_{n+1}$ by 2^{n+1} , expanding $(n+1)^3$, and gathering like powers of n , it becomes $n^3 > 3n^2 + 3n + 1$. It then only remains to observe that since $n \geq 4$, we have

$$n^3 \geq 3n^2 + n^2 \geq 3n^2 + 4n \geq 3n^2 + 3n + 4 > 3n^2 + 3n + 1.$$

Let's take another look at this solution in order to see what exactly was behind the claim that the term x_4 is the largest among *all* terms of the *infinite* sequence (x_n) . First we established (by direct calculation) that x_4 is greater than x_1, x_2 and x_3 . Then we proved that $x_n > x_{n+1}$ for all $n \geq 4$. Hence $x_4 > x_5 > x_6 > \dots$, and so on, to infinity, whence it follows that $x_4 > x_n$ for all $n \geq 5$.

1.2. The method of mathematical induction

From a formal viewpoint, the essence of the “method of mathematical induction” consists in the avoidance of the words “and so on, to infinity”, or, in other words, in avoiding the use of a sequence of dots, or “ellipsis”. The recipe for doing this is as follows. Suppose we wish to prove *infinitely many* mathematical statements, or, more precisely, suppose we have to prove that some statement $\mathcal{P}(n)$ is true for every natural number n : $\mathcal{P}(n) \forall n \in \mathbb{N}$. We are dealing here in essence with an infinite sequence $\mathcal{P}(1), \mathcal{P}(2), \dots$ of mathematical statements. It would appear that the best approach would be to prove $\mathcal{P}(n)$ once and for all, that is, for *any* natural number n . However, as the examples in the preceding section show, it may happen that to prove, for example, $\mathcal{P}(5)$, it is useful to know already that $\mathcal{P}(4)$ is true. Suppose that the very first proposition of the sequence, namely $\mathcal{P}(1)$, has been established as true; this is the *basis of the induction* or *first step of the induction*. Next, for any k , we prove that *on the assumption* that $\mathcal{P}(k)$ is true, it follows that the statement $\mathcal{P}(k+1)$ is true in turn—this is the *inductive step*. Then we are entitled to assert that all of the statements $\mathcal{P}(n)$ are valid; that is, that $\mathcal{P}(n)$ is true for all n .

The solution of Problem 1.6 below has all the standard features of an application of the method of mathematical induction. However, let's first consider how the method applies to our solution of Problem 1.5 above. We showed there that $x_n > x_{n+1}$ for all natural numbers $n \geq 4$. We wish to infer from this that $x_4 > x_n$ for all $n \geq 5$. The initial step of the induction is the inequality $x_4 > x_5$. We now carry out the inductive step. Thus we *assume* that for some (arbitrary) natural number n it is true that $x_4 > x_n$. Then, since we have already proved that $x_n > x_{n+1}$, it follows that $x_4 > x_{n+1}$. Hence by the method of mathematical induction we infer that $x_4 > x_n$ for all $n \geq 5$.

The solution of the next problem is representative in its use of this method.

PROBLEM 1.6. Prove the identity $\sum_{k=1}^n k^3 = \frac{n^2(n+1)^2}{4}$.

Direct substitution shows that the identity holds for $n = 1$, so we have the basis of the induction. Now we carry out the inductive step; that is, assuming that $\sum_{k=1}^n k^3 = \frac{n^2(n+1)^2}{4}$, we wish to deduce that $\sum_{k=1}^{n+1} k^3 = \frac{(n+1)^2(n+2)^2}{4}$.

We have

$$\begin{aligned}\sum_{k=1}^{n+1} k^3 &= \sum_{k=1}^n k^3 + (n+1)^3 = \frac{n^2(n+1)^2}{4} + (n+1)^3 \\ &= \frac{(n+1)^2}{4}(n^2 + 4n + 4) = \frac{(n+1)^2(n+2)^2}{4}.\end{aligned}$$

This example shows especially clearly that the method of mathematical induction affords a convenient approach to proving a statement somehow arrived at in advance.

Observe that we could have proved the identity “without induction” by taking the simple identity $k^3 = \frac{k^2(k+1)^2}{4} - \frac{(k-1)^2k^2}{4}$, and adding its instances from $k = 1$ to n :

$$\begin{aligned}1^3 &= 1^2 \cdot \frac{2^2}{4} - 0, \\ 2^3 &= 2^2 \cdot \frac{3^2}{4} - 1^2 \cdot \frac{2^2}{4}, \\ &\dots \\ n^3 &= \frac{n^2(n+1)^2}{4} - \frac{(n-1)^2n^2}{4}.\end{aligned}$$

However, first, the inductive proof is simpler, and second, it involves only two equations: the one we are assuming and the one we wish to establish as true. And, finally, the aim of the method of mathematical induction is to do away with ellipses in proofs.

PROBLEM 1.7 (The tower of Hanoi). We have three vertical pegs, on one of which ten discs-with-holes of differing diameters are threaded, so that each disc rests on a larger, with the largest at the bottom. What is the least number of moves of one disc at a time, needed to transfer all ten discs to one of the other two pegs, using the third as an auxiliary, without ever placing a larger disc on a smaller?

Strictly speaking, the solution of this problem does not require the method of mathematical induction since it specifies a specific number n of discs: $n = 10$. By direct experimentation in the cases $n = 1, 2, 3$, one finds that they require at least one, three, and seven moves, respectively. Now in order to transfer a “tower” of *four* discs to another peg, it is clearly necessary to first transfer the top three discs onto one of the other two pegs, then place the bottom, largest, one on the unoccupied peg, and, finally, arrange the three smaller pegs on top of the largest. These three steps altogether require $7 + 1 + 7 = 15$ moves of one disc at a time. Let’s now attempt to solve the problem in the general case of n discs. Clearly, the reasoning in the case of four discs generalizes to yield the following recurrence relation for the least number p_n of moves needed to transfer a tower consisting of n pegs:

$$p_{n+1} = 2p_n + 1.$$

For, in order to transfer the bottom, largest, of $n + 1$ discs (which is the case we now wish to consider), we first need to move the n discs lying on top of it to one of the other two pegs, and this requires at least p_n moves. Then we move the largest disc to the free peg, and need another p_n moves of the smaller discs to arrange them on top of it in the appropriate order.

The following table gives the values of p_n for $n = 1, 2, \dots, 7$:

n	1	2	3	4	5	6	7
p_n	1	3	7	15	31	63	127

It is not hard to guess that in fact $p_n = 2^n - 1$. There now remains just one further step: apply the method of mathematical induction to prove the formula we have found—in other words, solve the following problem.

PROBLEM 1.8. Given that $x_1 = 1$ and $x_{n+1} = 2x_n + 1$, prove that $x_n = 2^n - 1$.

Since $p_1 = 1 = 2^1 - 1$, we have the basis, or first step, of the induction. Now for the inductive step: if $p_n = 2^n - 1$, then

$$p_{n+1} = 2p_n + 1 = 2 \cdot (2^n - 1) + 1 = 2^{n+1} - 2 + 1 = 2^{n+1} - 1.$$

Although the Hanoi tower puzzle is one of the best problems for illustrating the inductive method of reasoning, the method of mathematical induction *per se* does not play a very substantial part in its solution. A careful examination of the above solution reveals that the leading role is played by a “recursive algorithm”, so-called, in the following sense. If we denote by $\mathcal{T}_n(i, j)$ the shortest sequence of moves one disc at a time needed to transfer the n discs from the i th peg of the three to the j th, then.

$$\mathcal{T}_1(i, j) = (\text{move the topmost disc from the } i\text{th to the } j\text{th peg}),$$

$$\text{and } \mathcal{T}_{n+1}(i, j) = (\mathcal{T}_n(i, k), \mathcal{T}_1(i, j), \mathcal{T}_n(k, j)).$$

PROBLEM 1.9. Into how many parts do

- a) n points subdivide a straight line;
- b) n straight lines subdivide a plane, if no two of the lines are parallel and no three are concurrent, that is, no three intersect at a single point?

It is totally obvious that n points divide a line into $n + 1$ segments. So what is there to prove here, and what relevance does induction have? The point here is that the first part of the problem provides a hint for solving the second.

Denote by c_n the number of regions of the plane into which n straight lines (no two of which are parallel and no three concurrent) divide it. If we now draw in another straight line, then it will intersect the n already present in exactly one point each. These n points of intersection divide our line up into $n + 1$ segments, each of which clearly divides some region into two parts, thus producing a further $n + 1$ regions. Hence $c_{n+1} = c_n + n + 1$. Then, since $c_1 = 2$, we have

$$c_n = c_{n-1} + n = 1 + (1 + 2 + 3 + \cdots + n) = 1 + \frac{n(n+1)}{2}.$$

This problem might have been formulated differently: What is the largest number of regions into which n straight lines divide the plane? Confronted with this version, the students would have to discover for themselves the geometrical conditions on the lines in the problem as stated (which are the conditions for the lines to be, as they say, “in general position”). It is interesting that even without those conditions, that is, for *any* n lines, there is still a simple formula for the number of regions into which they divide the plane.

EXERCISE 1.2. Suppose we have n straight lines intersecting in s points in the plane. Let k_1, k_2, \dots, k_s be the numbers of lines intersecting in these points, in some order. Prove that the number of regions into which the lines divide the plane is $n + 1 + \sum_{i=1}^s (k_i - 1)$.

There are problems that do not yield to an inductive argument directly, yet can easily be proved when recast in a stronger form.

PROBLEM 1.10. Prove the inequality

$$\frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{2n} < \frac{3}{4}.$$

Denoting the left-hand side of this inequality by a_n , we have

$$a_{n+1} = a_n + \frac{1}{2n+1} + \frac{1}{2n+2} - \frac{1}{n+1} = a_n + \frac{1}{2n+1} - \frac{1}{2n+2} > a_n,$$

so it is not clear how we might infer from $a_n < \frac{3}{4}$ that $a_{n+1} < \frac{3}{4}$. However, it is not difficult to see that if $a_n \leq \frac{3}{4} - \frac{1}{4n}$, then $a_{n+1} \leq \frac{3}{4} - \frac{1}{4n+4}$ (prove it!), whence $a_n \leq \frac{3}{4} - \frac{1}{4n} < \frac{3}{4}$ for all natural n .

1.3. The principle of mathematical induction

That a proof using “the method of mathematical induction” really *is* a proof was taken as obvious above. After all, if $\mathcal{P}(1)$ is true, then so is $\mathcal{P}(2)$, and therefore in turn $\mathcal{P}(3)$, and so on!

Any doubters may be satisfied with the following reasoning. Suppose we have a proposition depending on n that is known to hold for $n = 1$, and for which its truth for $n = k$ implies its truth for $n = k + 1$, and yet is *not* true for all natural numbers n . Consider the smallest number n_0 for which the given proposition is false. Clearly $n_0 > 1$, so that $n_0 - 1$ is a natural number, and for it the statement is true. However, this then implies that it is true for n_0 !

To an even greater skeptic the following question might then occur: Why should every nonempty subset of \mathbb{N} have a smallest element?¹ After all, it is well known that this is not the case for the set \mathbb{Q} !² Let \mathcal{E} be a nonempty subset of \mathbb{N} , and let $a \in \mathcal{E} \subset \mathbb{N}$. If a is not smallest in \mathcal{E} , then there must be another element $a^{(1)} \in \mathcal{E}$ such that $a^{(1)} < a$. Then if $a^{(1)}$ is not smallest in \mathcal{E} , there must exist a smaller element $a^{(2)} \in \mathcal{E}$, and so on, until we reach 1, which is less than every other natural number. But how can we avoid using the words “and so on”?

The truth of the method of mathematical induction ultimately rests on axioms for the system of natural numbers—the axioms of arithmetic. In speaking of the “principle of mathematical induction”, we have in mind the correctness of the method of mathematical induction.

How can we describe *infinite* mathematical objects, such as, for example, the various sets of numbers: natural numbers, integers, rationals, etc.? As we shall see later—in Chapters 3, 8, and 9—once we know what \mathbb{N} is, we will be able to construct, by means of certain relatively simple set-theoretical constructions, the set \mathbb{Z} of whole numbers and the set \mathbb{Q} of rational numbers, and then, once we have the set \mathbb{R} of real numbers, easily construct the set \mathbb{C} of complex numbers. But we have to start somewhere, since each construction is based on a previously defined set. The remainder of this chapter is, in fact, devoted to an examination of what exactly the familiar set \mathbb{N} of natural numbers might be. Although later on in this

¹Standard notation for the set of natural numbers: $\mathbb{N} := \{1, 2, 3, \dots\}$. *Trans.*

²Standard notation for the set of all rational numbers, that is, numbers expressible in the form $\frac{m}{n}$, where n is a natural number and m an integer. The set of all positive rational numbers, for instance, has no smallest element, since it does not contain zero. *Trans.*

chapter we shall show how to derive it from the axioms of arithmetic, for the time being we shall assume as fact that *every nonempty subset of the set \mathbb{N} of natural numbers has a smallest element*.

There are situations where knowing that a given statement is true for n does not immediately yield its truth for $n + 1$.³ We have already met with examples of this: see Problems 1.1 and 1.2.

PROBLEM 1.11. Prove that every plane polygon (not necessarily convex) can be subdivided into pairwise nonoverlapping triangles by means of diagonals not intersecting in the interior of the polygon.

EXERCISE 1.3. Prove that every plane polygon has a diagonal lying in its interior.

We argue by induction (on the number of sides of the polygon) using Exercise 1.3. If the polygon is a triangle, there is nothing to prove. Suppose the assertion is true for all polygons with fewer than n sides. Consider any n -gon. By Exercise 1.3, it has a diagonal dividing it into two polygons with fewer sides. Since (by the inductive hypothesis) each of these can be subdivided into triangles by means of pairwise nonintersecting diagonals, it follows that the same is true of the given n -gon.

In this solution we have used the so-called “extended version” of the method of mathematical induction. Here is the justification of this, that is, its derivation from our as yet unproved assertion that every nonempty subset of natural numbers has a least element.

THEOREM 1.1. *Let $(\mathcal{P}(n))_{n \in \mathbb{N}}$ be a sequence of mathematical statements. Suppose that the statement $\mathcal{P}(l)$ is true, and that for every natural number n the truth of the statements $\mathcal{P}(k)$ for $l \leq k \leq n$ implies the truth of the statement $\mathcal{P}(n + 1)$. Then the statement $\mathcal{P}(n)$ is true for all natural $n \geq l$.*

Consider the number

$$n_0 = \min\{n \in \mathbb{N} \mid n \geq l \text{ and } \mathcal{P}(n) \text{ is false}\}.$$

Since $\mathcal{P}(n_0)$ is a false statement, we must have $n_0 > l$. By definition of n_0 the statement $\mathcal{P}(k)$ must be true for $k = l, l + 1, \dots, n_0 - 1$, whence, by the second assumption in the theorem, $\mathcal{P}(n_0)$ is also true. This contradiction completes the proof (since it shows that the above set must be empty). \square

1.4. Peano's axioms

So, what exactly is “the set \mathbb{N} of natural numbers”?

Consider an arbitrary triple of the form $\{\mathcal{N}, 1, s\}$, consisting of:

- some set \mathcal{N} ,
- a distinguished element 1 of it,
- a mapping $s : \mathcal{N} \rightarrow \mathcal{N}$.

³That is, where the inductive step is nonobvious. *Trans.*

We assume that this triple satisfies the following conditions, called *Peano's*⁴ *axioms*:⁵

- (Pe1) if $x, y \in \mathcal{N}$ and $x \neq y$, then $s(x) \neq s(y)$;
- (Pe2) the distinguished element 1 is not the image of any element of the set \mathcal{N} under the mapping s ;
- (Pe3) if M is a subset of \mathcal{N} , such that $1 \in M$ and $s(x) \in M$ for all $x \in M$, then $M = \mathcal{N}$.

Axiom (Pe1) means simply that the mapping (or function) s is one-to-one, or “injective”.⁶ Recall that the *image set* $f(A)$ of a subset $A \subset X$ under a mapping (or function) $f : X \rightarrow Y$, is defined as the set consisting of the images $f(a)$ of all the elements $a \in A$:

$$f(A) = \{f(a) \mid a \in A\}.$$

In this notation, axiom (Pe1) may be written in the form: $1 \notin s(\mathcal{N})$. We shall say that a subset M of the set \mathcal{N} is *inductive* if $1 \in M$ and $s(M) \subset M$. In this terminology, axiom (Pe3) states that \mathcal{N} has no nontrivial inductive subsets, that is, different from the whole set \mathcal{N} . These axioms were introduced by the Italian mathematician Giuseppe Peano in the 1890s. The idea behind the function s is that $s(x)$ should be the next natural number after x , that is, its successor $x + 1$. However, we have yet to define addition on the set \mathcal{N} .

We now deduce some basic consequences of Peano's axioms.

LEMMA 1.2. *We have $s(\mathcal{N}) = \mathcal{N} \setminus 1$, and $s(a) \neq a$ for all $a \in \mathcal{N}$.*

We merely have to prove that each of the subsets

$$M_1 = \{1\} \cup s(\mathcal{N}) \text{ and } M_2 = \{a \in \mathcal{N} \mid s(a) \neq a\}$$

is inductive. The set M_1 contains the distinguished element 1 by definition. Further, if $a \in M_1$, then $s(a) \in M_1$ simply because M_1 contains the images of *all* elements of the set \mathcal{N} . Hence by axiom (Pe3), $M_1 = \mathcal{N}$, and moreover, since $1 \notin s(\mathcal{N})$ in view of axiom (Pe2), it follows that $s(\mathcal{N}) = \mathcal{N} \setminus 1$. The proof that M_2 is inductive is left for the reader to carry out independently. (First use axiom (Pe2), and then (Pe1)). \square

A (binary) operation on a set may be given precise meaning as a mapping F associating with each ordered pair (a, b) of elements of the set, an element $c = F(a, b)$ of the set. The standard examples—and the ones that will be of interest to us—are the usual “arithmetic operations”, that is, addition and multiplication of numbers. Two of their most basic properties are *commutativity*, that is, the property that $a + b = b + a$ and $ab = ba$ for all a, b , and *associativity*: $a + (b + c) = (a + b) + c$ and $a(bc) = (ab)c$. But, of course, we have still not defined these operations.

⁴Giuseppe Peano (1858–1932), Italian mathematician. One of the founders of mathematical logic.

⁵These “axioms of arithmetic” presuppose basic set theory, so someone wishing to descend further into the foundations of mathematics might ask for axioms for Cantor's set theory. The almost standard such axioms are those of Zermelo and Fraenkel, in terms of which the set of natural numbers and their arithmetic—and much else besides—can be defined. Descending further, one comes to the programs of G. Frege and Russell-Whitehead of the early 1900s for inferring mathematics from basic principles of logic, considerably refined by later logicians, notably W. V. O. Quine. *Trans.*

⁶That is, images of distinct elements are distinct.

Note that commutativity of an operation F is equivalent to its symmetry, $F(a, b) = F(b, a)$, while associativity is equivalent to the property

$$F(a, F(b, c)) = F(F(a, b), c).$$

We shall now consider an operation F on the set \mathcal{N} satisfying the following conditions:

- (Add1) $F(a, 1) = s(a)$ for all $a \in \mathcal{N}$;
 (Add2) $F(a, s(b)) = s(F(a, b))$ for all $a, b \in \mathcal{N}$.

THEOREM 1.3. *Any operation $F : \mathcal{N} \times \mathcal{N} \rightarrow \mathcal{N}$, satisfying conditions (Add1), (Add2), is commutative and associative.*

Here is the proof of associativity. Fix on any two elements $a, b \in \mathcal{N}$, and set

$$M = \{c \in \mathcal{N} \mid F(a, F(b, c)) = F(F(a, b), c)\}.$$

Since $F(a, F(b, 1)) = F(a, s(b))$ (by condition (Add1)) $= s(F(a, b))$ (by condition (Add2)) $= F(F(a, b), 1)$ (by (Add1) again), we have $1 \in M$. Consider any element $c \in M$. Then $F(a, F(b, s(c))) = F(a, s(F(b, c))) = s(F(a, F(b, c))) = s(F(F(a, b), c))$ (it is here that we have used the assumption that $c \in M$) $= F(F(a, b), s(c))$, whence $s(c) \in M$. Hence M is an inductive subset, so that in fact $M = \mathcal{N}$.

EXERCISE 1.4. Prove the commutativity of such an operation F . \square

THEOREM 1.4. *There is at most one mapping satisfying conditions (Add1), (Add2).*

Fix on any element $a \in \mathcal{N}$, and consider the set

$$M = \{b \in \mathcal{N} \mid F(a, b) = F'(a, b)\},$$

where F and F' are any two mappings satisfying conditions (Add1), (Add2). Since $F(a, 1) = s(a) = F'(a, 1)$, we have $1 \in M$. Consider any element $b \in M$. Then $F(a, s(b)) = s(F(a, b)) = s(F'(a, b)) = F'(a, s(b))$, whence $s(b) \in M$. Thus M is an inductive subset of \mathcal{N} , so that again $M = \mathcal{N}$, implying that $F = F'$. \square

1.5. Addition, order, and multiplication

As already indicated above, Peano's axioms for the set of natural numbers can be used to introduce the arithmetic operations on that set, as well as the familiar order relation. This was first pointed out by the German mathematician Hermann Grassmann.⁷

THEOREM 1.5. *There exists an operation of addition on the set \mathcal{N} , that is, a mapping F satisfying conditions (Add1), (Add2). Furthermore, there is only one such operation, and it defines a commutative and associative operation on \mathcal{N} .*

Consider the set M consisting of all elements $a \in \mathcal{N}$, for which there exists a mapping $f_a : \mathcal{N} \rightarrow \mathcal{N}$, such that

$$f_a(1) = s(a), \text{ and } f_a(s(b)) = s(f_a(b)) \text{ for all } b \in \mathcal{N}.$$

Define $f_1(b) = s(b)$. Then $f_1(1) = s(1)$ and $f_1(s(b)) = s(s(b)) = s(f_1(b))$, whence $1 \in M$. For each $a \in M$, we define $f_{s(a)}$ by the formula $f_{s(a)}(b) = s(f_a(b))$. We then

⁷Hermann Günther Grassmann (1809–1877), German polymath.

have $f_{s(a)}(1) = s(f_a(1)) = s(s(a))$, and $f_{s(a)}(s(b)) = s(f_a(s(b))) = s(s(f_a(b))) = s(f_{s(a)}(b))$ for all $b \in \mathcal{N}$. We have thus proved that the set M is inductive, so that $M = \mathcal{N}$, and the formula $F(a, b) = f_a(b)$ defines a mapping F satisfying conditions (Add1), (Add2). By Theorem 1.4 this function is unique, and by Theorem 1.3 the operation it defines is commutative and associative. \square

LEMMA 1.6. *We have $a + b \neq a$ for all $a, b \in \mathcal{N}$.*

Observe first that $1 + b = b + 1 = s(b) \neq 1$ in view of Lemma 1.2. Proceeding further, if $a \in \mathcal{N}$ is such that $a + b \neq a$ for all $b \in \mathcal{N}$, then, since the mapping s is injective, we have $s(a) \neq s(a + b) = s(b + a) = b + s(a) = s(a) + b$. Hence the set $M = \{a \in \mathcal{N} \mid a + b \neq a \ \forall b \in \mathcal{N}\}$ is inductive, as required. \square

We now define the *order relation* on \mathcal{N} : We shall say that a is *greater than* b , and write $a > b$, if there is an element $k \in \mathcal{N}$ such that $a = b + k$.

THEOREM 1.7. *For any two elements $a, b \in \mathcal{N}$, exactly one of the following possibilities is realized: 1) $a > b$; 2) $a = b$; 3) $a < b$.*

That no two of the listed possibilities can be realized simultaneously follows from Lemma 1.6. In order to prove that at least one of them holds, fix on an arbitrary element $b \in \mathcal{N}$ and consider the set

$$M = \{a \mid a > b \text{ or } a = b \text{ or } a < b\}.$$

EXERCISE 1.5. Prove that the set M just defined is inductive. \square

COROLLARY. *The elements $a, a + 1$ are adjacent in the above ordering of the set \mathcal{N} ; that is, there is no element $b \in \mathcal{N}$ such that $a < b < a + 1$.*

EXERCISE 1.6. Prove this corollary.

THEOREM 1.8. *Every nonempty subset of \mathcal{N} has a least element.*

Let E be any nonempty subset of the set \mathcal{N} . Consider the set $M = \{a \in \mathcal{N} \mid a \leq b \ \forall b \in E\}$, consisting of all elements of \mathcal{N} less than or equal to everything in E . Obviously, $1 \in M$. If for every $u \in M$ we had $u + 1 \in M$, then M would be inductive, and so coincide with \mathcal{N} , and then E would be empty. Hence there is an element $u \in \mathcal{N}$ such that $u \in M$ but $u + 1 \notin M$. If we can show that u also belongs to E , it will follow that u is the least element of E . Now since $u + 1 \notin M$, there must exist an element $b_0 \in E$ such that $b_0 < u + 1$. If u did not belong to E , then since $u \in M$, we would have $u < b$ for all $b \in E$, whence, in particular, $u < b_0 < u + 1$, contradicting the corollary to Theorem 1.7. \square

Multiplication in the set of natural numbers is introduced in a similar fashion. The following theorem holds.

THEOREM 1.9. *There exists a unique mapping $G : \mathcal{N} \times \mathcal{N} \rightarrow \mathcal{N}$, such that:*

(Mult1) $G(a, 1) = a$;

(Mult2) $G(a, s(b)) = G(a, b) + a$.

The mapping G defines an operation on \mathcal{N} which is commutative, associative, and distributive with respect to addition, that is, for all $a, b, c \in \mathcal{N}$ the equality $G(a, b + c) = G(a, b) + G(a, c)$ holds.

LEMMA 1.10. *If $G : \mathcal{N} \times \mathcal{N} \rightarrow \mathcal{N}$ is a mapping satisfying conditions (Mult1), (Mult2), then*

- 1) $G(1, b) = b$;
- 2) $G(s(a), b) = G(a, b) + b$.

The proof proceeds along the usual lines. Let

$$M = \{b \mid G(s(a), b) = G(a, b) + b\}.$$

Since $G(s(a), 1) = s(a) = a + 1 = G(a, 1) + 1$, we have $1 \in M$. Let b be any element from M . Since

$$\begin{aligned} G(s(a), s(b)) &= G(s(a), b) + s(a) = (G(a, b) + b) + (a + 1) \\ &= (G(a, b) + a) + (b + 1) = G(a, s(b)) + s(b), \end{aligned}$$

we have $s(b) \in M$, whence $M = \mathcal{N}$.

EXERCISE 1.7. Prove Theorem 1.9. \square

EXERCISE 1.8. Prove that $2 \times 2 = 4$. (Here $2 = s(1)$, $3 = s(2)$, $4 = s(3)$).

EXERCISE 1.9. Prove that if $a < b$ and $c \in \mathcal{N}$, then $ac < bc$.

Our final observation concerning Peano's axioms is that they constitute a full axiomatization of the natural numbers in the sense that, as we shall now show, there is essentially just one set satisfying them. In such cases, it is said that all models of the axiom system are "isomorphic". Thus in contrast with the axioms for a group or those for a topological space, Peano's axioms define the unique object that we call the *set of natural numbers* and denote by \mathbb{N} .

THEOREM 1.11. *If $\{\mathcal{N}, 1, s\}$ and $\{\mathcal{N}', 1', s'\}$ are two triples satisfying Peano's axioms, then there exists a one-to-one correspondence $\Phi : \mathcal{N} \rightarrow \mathcal{N}'$, such that $\Phi(1) = 1'$ and for all $x \in \mathcal{N}$ the equation $\Phi(s(x)) = s'(\Phi(x))$ holds.*

Suppose we have a mapping $\Phi : \mathcal{N} \rightarrow \mathcal{N}'$, such that $\Phi(1) = 1'$ and $\Phi(s(x)) = s'(\Phi(x))$ for all $x \in \mathcal{N}$. We shall show that then Φ must be a bijection.⁸ Write $M = \Phi(\mathcal{N})$. By the first property of Φ , we have $1' = \Phi(1) \in M$. If $y = \Phi(x) \in M$, then $s'(y) = s'(\Phi(x)) = \Phi(s(x)) \in M$. Thus we have established that M is inductive, whence $M = \mathcal{N}'$. Hence $\Phi(\mathcal{N}) = \mathcal{N}'$, that is, Φ is a surjection.⁹

If $x \neq 1$, then $x = s(x_1)$ for some $x_1 \in \mathcal{N}$, whence $\Phi(x) = \Phi(s(x_1)) = s'(\Phi(x_1))$, from which it follows that $\Phi(x) \neq 1'$. Hence the element $1' \in \mathcal{N}'$ has only one preimage, namely 1, under the mapping Φ .

Consider the set

$$M = \{y \in \mathcal{N} \mid x \neq y \implies \Phi(x) \neq \Phi(y)\}.$$

We have just finished proving that $1 \in M$. Let $y \in M$ be such that $x \neq s(y)$. If $x = 1$, then $\Phi(x) = 1' \neq \Phi(s(y)) = s'(\Phi(y))$. If $x \neq 1$, then $x = s(x_1)$. Since we are assuming that $x \neq s(y)$, it follows that $x_1 \neq y$, whence $\Phi(x_1) \neq \Phi(y)$, and, in turn, $s'(\Phi(x_1)) \neq s'(\Phi(y))$, which, by the second condition on Φ , is equivalent to $\Phi(x) \neq \Phi(s(y))$. We have thus shown that if $y \in M$, then $s(y) \in M$; that is, (since also $1 \in M$) that M is an inductive set. Hence $M = \mathcal{N}$, and whenever $x, y \in \mathcal{N}$ are such that $x \neq y$, then $\Phi(x) \neq \Phi(y)$, completing the proof that Φ is a bijection.

It remains to prove that such a mapping Φ actually exists. This comes down, in essence, to the correctness of a recursive definition. In order not to frighten off

⁸That is, both one-to-one and onto. *Trans.*

⁹That is, a mapping onto \mathcal{N}' . *Trans.*

the reader completely, we shall refrain from giving a formal proof; in any case, the situation is intuitively obvious, in the sense that once $\Phi(x)$ is defined, then the formula $\Phi(s(x)) = s'(\Phi(x))$ defines Φ on the successor $s(x)$. \square

EXERCISE 1.10. Prove that:

- 1) $\Phi(x + y) = \Phi(x) + \Phi(y)$ and $\Phi(xy) = \Phi(x)\Phi(y)$ for all $x, y \in \mathcal{N}$;
- 2) if $x < y$, then $\Phi(x) < \Phi(y)$.

1.6. The number of elements of a set

We often use the phrase “the number of elements of a set” or “the size of a set”. How exactly does one associate a number, that is, a natural number, or element of \mathbb{N} , with a given set?

Consider the following example—taken from life, as it were.

What’s the simplest method of determining whether there are more girls than boys or the other way around, at the dance-hall? Of course, one might simply count the boys, and then the girls, and compare the resulting numbers. But there is a more natural method. If “ladies’ choice” is announced, then each girl will rush to find a partner, and the answer to our question will be decided by whether there remain boys or girls left standing along the wall.

Thus we shall say that a set A has k elements, if there exists a one-to-one correspondence (or bijection) between it and the set $I_k = \{n \in \mathbb{N} \mid n \leq k\}$, consisting of all natural numbers not exceeding the natural number k (an “initial segment” of \mathbb{N}). We need to show that the number k of elements of A is unique; that is, that, a set cannot consist simultaneously of, say, three and four elements. Thus suppose there is a one-to-one correspondence between A and I_k , and also between A and I_l , where $k \neq l$; suppose, for definiteness, that $k < l$. Then I_k is a proper subset of I_l , but this does not immediately imply that there cannot be a one-to-one correspondence between them—after all the mapping s determines such a correspondence between \mathbb{N} and $\mathbb{N} \setminus 1$. You might object that in the case in question the sets are finite. But then, just what *is* a finite set?...

THEOREM 1.12. *If $k \neq l$, then there is no one-to-one correspondence between I_k and I_l .*

EXERCISE 1.11. Prove this theorem using the method of mathematical induction. \square

EXERCISE 1.12. Construct a one-to-one correspondence between the sets $[0, 1)$ and $[0, 1]$.¹⁰

We prove one more useful fact.

LEMMA 1.13. *For all $k, l \in \mathbb{N}$ there is a one-to-one correspondence between the sets I_k and $I_{k+l} \setminus I_l$.*

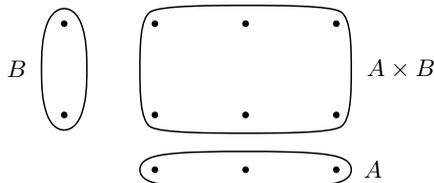
We define a mapping $\psi : I_k \rightarrow \mathbb{N}$ by setting $\psi(i) = i + l$. From the inequalities $1 \leq i \leq k$, it follows that $l + 1 \leq \psi(i) \leq k + l$, which means that for all $i \in I_k$, we have $\psi(i) \in I_{k+l} \setminus I_l$. It is not difficult to see that ψ is an injection, so it remains to prove only that $\psi(I_k) = I_{k+l} \setminus I_l$. Now if $j \in I_{k+l} \setminus I_l$, then $j > l$, so that $j = l + i$, where, clearly, $i \leq k$. \square

¹⁰These are intervals of real numbers. Here $[0, 1)$ consists of all real numbers x satisfying $0 \leq x < 1$, and $[0, 1]$ of all real numbers x such that $0 \leq x \leq 1$. *Trans.*

From now on we shall denote the number of elements in a (finite) set A by $|A|$.

COROLLARY. *If A and B are non-intersecting finite sets, then $|A \cup B| = |A| + |B|$. In particular, if B is a singleton set, that is, consists of a single element, then $|A \cup B| = s(|A|)$.*

In standard notation, $A \times B$ denotes the *Cartesian product* of the sets A and B , defined as the set of all ordered pairs (a, b) , where $a \in A$ and $b \in B$.



EXERCISE 1.13. Define $H(|A|, |B|) := |A \times B|$. Prove that this formula defines a binary operation on the set \mathbb{N} , satisfying conditions (Mult1), (Mult2), whence $|A \times B| = |A| \cdot |B|$.

We conclude with a joke-theorem taken from the book *Physicists Continue Joking*, and justified by the proverb “Many a true word spoken in jest”.

THEOREM 1.14 (joke). *Alexander the Great never existed.*

We first need to prove the following lemma.

LEMMA 1.15 (false). *All objects have the same color.*

The proof is by induction. Let n stand for the number of objects in a collection. The initial case $n = 1$ of the induction is trivially true. Suppose the lemma true for $n = k$, that is, that any k objects have the same color, and consider any set of $k + 1$ objects. If we remove one of them, we are left with k , which by the inductive hypothesis all have the same color. Replacing that object and removing some other, we conclude, again by the inductive hypothesis, that the first must have had the same color as the rest. \square

We are now ready to prove the theorem. Suppose the contrary, that is, that Alexander the Macedonian lived. Historians assert—and historians never lie—that Alexander rode on a black stallion called Bucephalus. Now it is known that there exist white objects.¹¹ Hence, in view of the lemma just proved, all objects must be white, and therefore it is absolutely impossible for Alexander the Great to have ridden a black horse. This contradiction completes the proof. \square

Supplementary problems

PROBLEM 1.12. Prove Bernoulli’s¹² inequality:

$$(1 + h)^n \geq 1 + hn \text{ for all } n \in \mathbb{N} \text{ and } h > -1.$$

PROBLEM 1.13. Using the method of mathematical induction, prove that $(x - 1)(x^{n-1} + x^{n-2} + \dots + 1) = x^n - 1$.

PROBLEM 1.14. Prove the inequality $\frac{1}{\sqrt{1}} + \frac{1}{\sqrt{2}} + \dots + \frac{1}{\sqrt{n}} > \sqrt{n}$ for $n > 1$.

¹¹See Mark Twain’s short story “The stolen white elephant”, published in 1882.

¹²Jacob (or Jacques) Bernoulli (1654–1705), Swiss mathematician. One of the many eminent mathematicians in the Bernoulli family.

PROBLEM 1.15. Find all natural n for which the inequality $5n^2 > 2^n$ holds.

PROBLEM 1.16. Prove that the inequality

$$\frac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots 2n} \leq \frac{1}{\sqrt{2n+1}}$$

holds for every natural number n .

PROBLEM 1.17. Prove that for every natural number n the number $4^n + 15n - 1$ is divisible by 9.

PROBLEM 1.18. Prove the inequality

$$1 < \frac{1}{n+1} + \frac{1}{n+2} + \cdots + \frac{1}{3n+1} < 2.$$

PROBLEM 1.19. Prove that the number of ways of tiling a rectangular $2 \times n$ strip by means of rectangular 2×1 tiles is equal to the n th Fibonacci number F_n .¹³

PROBLEM 1.20. Given that the sequence (x_1, x_2, \dots) is such that $x_1 = 0$, $x_2 = 4$, and $x_{n+1} = 4(x_n - x_{n-1})$, prove that $x_n = (n-1)2^n$. Find the general term of such a sequence if the first two terms are given instead as $x_1 = 1$ and $x_2 = 3$.

PROBLEM 1.21. What is the largest number of regions into which n planes subdivide space?

PROBLEM 1.22. Find a formula in closed form for the sum

$$1 \cdot 2 \cdot 3 + 2 \cdot 3 \cdot 4 + \cdots + n(n+1)(n+2).$$

PROBLEM 1.23. Prove that if the (real) numbers x_1, x_2, \dots, x_n are all positive, not all equal, and satisfy $x_1 x_2 \cdots x_n = 1$, then

$$x_1 + x_2 + \cdots + x_n > n.$$

PROBLEM 1.24. Prove Chebyshev's¹⁴ inequality:

If $a_1 \leq a_2 \leq \cdots \leq a_n$ and $b_1 \leq b_2 \leq \cdots \leq b_n$, then

$$\sum_{i=1}^n a_i \sum_{i=1}^n b_i \leq n \sum_{i=1}^n a_i b_i.$$

PROBLEM 1.25. Prove that the sum of the interior angles of an arbitrary plane (not necessarily convex) n -gon is equal to $\pi(n-2)$.

PROBLEM 1.26. The Chebyshev polynomials of the first kind are defined by the formula $T_n(x) = \cos(n \arccos x)$, $n \in \mathbb{N}$. Prove that $2^{1-n} T_n(x)$ is a polynomial of degree n with leading coefficient 1.

PROBLEM 1.27. Find an explicit formula for calculating $\cos \frac{\pi}{1024}$, not involving any trigonometric functions.

PROBLEM 1.28. Let (q_1, q_2, \dots) be a sequence of natural numbers such that $q_1 = 1$ and $q_{n+1} \leq 1 + q_1 + \cdots + q_n$. Prove that every natural number can be expressed as a sum of different terms (or just one term) of this sequence.

¹³We remind the reader that the *Fibonacci numbers* are the terms of the sequence (F_n) , where $F_0 = F_1 = 1$, and succeeding terms are produced using the recurrence relation $F_{n+1} = F_n + F_{n-1}$. Leonardo of Pisa (Fibonacci) (ca. 1170–1250). His *Book of calculation* introduced the Hindu decimal place-value system of numeration to Europe.

¹⁴Pafnutii L'vovich Chebyshev (1821–1894), Russian mathematician. Made important contributions to number theory and probability theory.

Comments of a pedagogical nature

One of the aims of the first few sections of this chapter is to demonstrate the difference in meaning of the phrases the “inductive method of reasoning”, the “method of mathematical induction”, and the “principle of mathematical induction”, all three of which are often used unreflectingly. It is not difficult to teach high school students to carry out an inductive proof of an identity along the lines of the solution to Problem 1.6. However, as far as their mathematical education is concerned, it is far more important (and vastly more difficult) to imbue them with the ability to reason inductively, a skill employed generally in searching for solutions (see the solution of Problem 1.7), the essence of which comes down to an ability to glimpse the basic idea of a general proof by examining particular cases. Note that Problem 1.6 can be made more difficult by reformulating it as follows: Prove that the sum of the cubes of the first n natural numbers is the square of a natural number. Clearly, in this form it should not be set the more junior students—for whom, however, there is a good substitute: Prove that the sum of the first n odd numbers is the square of a natural number. Incidentally, the identity at the center of the latter problem has a very pretty geometrical interpretation. (And how might Problem 1.6 be interpreted geometrically?). In the spirit of the “pigeon-hole principle” (see Chapter 7), one may suggest the “pine-tree principle” to the more junior students: In order to climb to the top of a pine tree, one has somehow to get oneself up onto the lowest branch, and then think of a way of moving from each branch up to the next one.

The author has limited himself in this chapter to a moderate selection of problems, serving mainly as illustrations, since in almost every subsequent chapter there are statements whose proofs use the method of mathematical induction. It is perhaps appropriate to stress once again that in many cases the principle of mathematical induction (as the most basic of the axioms defining the set \mathbb{N}) does not appear explicitly—for example, when ellipses (...) are used, or the phrase “and so on”. In such cases, the teacher need not (or even *should* not) remind the students of the principle of induction—but he or she should certainly have it in mind.

A few words on Peano’s axioms. Clearly, high school students must have the axiomatic method demonstrated to them using familiar examples; usually, the foundations of geometry serve as the context for a first introduction to axiomatics. However, the axiom system of Euclidean geometry is rather complicated, so that it is practically impossible within the limits of the high school syllabus to cover completely (even if sketchily) the construction of elementary geometry from the axioms—although it might be worthwhile defining the concept of a ray in terms of Hilbert’s axioms [7]. It might be thought that Weyl’s axiom system represents an exception, but then that is more linear algebra than geometry. . . . Peano’s axiom system, serving to define an object so very familiar to all, so compact and simple in structure, and so closely connected to the “method of mathematical induction”, which forms part of the high school syllabus, affords the students the possibility of looking afresh at the set of natural numbers and the two binary operations on it. It is strange, therefore, that it has not hitherto been included in the syllabi of schools specializing in mathematics and the sciences. The author of this book was, it may be said, lucky. He was introduced to Peano’s axioms in his senior high school mathematics classes, and, although 40 years have elapsed since then, still remembers the impression they made on him.

Solutions of the exercises

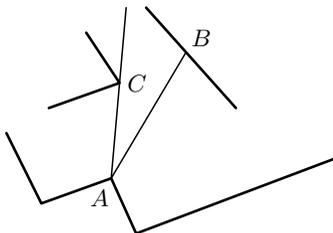
1.1. The desired general formulation is: If $a_1 \geq a_2 \geq \dots \geq a_{2n+1}$, then

$$a_1^2 - a_2^2 + \dots + a_{2n+1}^2 \geq (a_1 - a_2 + \dots + a_{2n+1})^2.$$

The proof is by induction on n . Generalize the inductive step given in the solution to Problem 1.4.

1.2. Strictly speaking, if, as here, the formula is given, then it is hardly worthwhile proving it—especially in this case, where the proof scarcely differs from the solution of Problem 1.9. On adding an arbitrary line, the number of regions increases by $l + 1$, where l is the number of points in which the new line meets those already there.

1.3. If all interior angles of the polygon are less than π , then the polygon is convex, and there is nothing to prove since then *all* its diagonals lie in its interior. Hence we may assume that there is at least one interior angle exceeding π , at a vertex A , say. Consider any ray emanating from the point A , and directed into the interior of the polygon. Let B be the first point in which the ray meets some edge of the polygon. We now rotate the ray about the point A in either direction, keeping it in the interior of the polygon. Since the angular interval of this rotation is greater than π , at some moment, clearly, the (variable) point B either moves continuously to a vertex different from A , or else “jumps” onto a different side of the polygon, which can occur only if the ray meets a vertex C as shown in the diagram. In this case, the segment AC is an interior diagonal.



The author is not completely satisfied with this proof. Can the reader think of a better one?

1.4. Consider the set

$$M = \{a \in \mathcal{N} \mid F(a, 1) = F(1, a)\}.$$

Clearly, $1 \in M$. Assume inductively that $a \in M$. In the following argument, we use the associativity of the operation F in addition to the conditions (Add1) and (Add2). Thus we have

$$\begin{aligned} F(s(a), 1) &= s(F(a, 1)) = s(F(1, a)) \\ &= F(F(1, a), 1) = F(1, F(a, 1)) = F(1, s(a)), \end{aligned}$$

whence $s(a) \in M$. The set M is therefore inductive, so that $M = \mathcal{N}$. We have thus proved that $F(a, 1) = F(1, a)$ for all $a \in \mathcal{N}$. We now fix on an arbitrary element $a \in \mathcal{N}$, and consider the set

$$M = \{b \in \mathcal{N} \mid F(a, b) = F(b, a)\}.$$

We have just shown that $1 \in M$. Suppose inductively that $b \in M$. Then

$$\begin{aligned} F(a, s(b)) &= F(a, F(b, 1)) = F(F(a, b), 1) = F(F(b, a), 1) \\ &= F(1, F(b, a)) = F(F(1, b), a) = F(F(b, 1), a) = F(s(b), a), \end{aligned}$$

whence $s(b) \in M$. Thus once again we have that M is inductive, and consequently $M = \mathcal{N}$. We conclude that $F(a, b) = F(b, a)$ for all $b \in \mathcal{N}$ and “any particular” $a \in M$, in other words, that the operation F is commutative.

1.5. We already know (see Lemma 1.2) that $s(\mathcal{N}) = \mathcal{N} \setminus 1$, so that if $b \neq 1$, there must be an element $c \in \mathcal{N}$, such that $b = s(c) = c + 1 = 1 + c$, implying that $b > 1$. Hence $b \geq 1$ for all $b \in \mathcal{N}$, whence $1 \in M$. Assume inductively that $a \in M$. If $a = b$, then $s(a) = a + 1 > b$, whence $s(a) \in M$. If $a > b$, then $a = b + k$, so $s(a) = a + 1 = (b + k) + 1 = b + (k + 1)$, whence $s(a) > b$. Finally, if $a < b$, then $b = a + k$. If $k = 1$, then $s(a) = a + 1 = b$. If $k \neq 1$, then, as already proven, $k > 1$, whence $k = 1 + l$, and $b = a + (1 + l) = (a + 1) + l = s(a) + l$, that is, $b > s(a)$. This completes the proof of the inductive step: if $a \in M$, then $s(a) \in M$.

1.6. If $b > a$, then $b = a + k$. If $k = 1$, then $b = a + 1$, while if $k \neq 1$, then $k = 1 + l$, and $b = a + (1 + l) = (a + 1) + l > a + 1$. Hence if $b > a$, then $b \geq a + 1$.

1.7. The existence and uniqueness of the mapping G is established along the same lines as the existence and uniqueness of the operation of addition. We shall prove that if the mapping G satisfies conditions (Mult1) and (Mult2), then it must be distributive, associative, and commutative.

Distributivity. Consider the set

$$M = \{c \mid G(a, b + c) = G(a, b) + G(a, c) \ \forall a, b \in \mathcal{N}\}.$$

Since

$$G(a, b + 1) = G(a, s(b)) = G(a, b) + a = G(a, b) + G(a, 1),$$

we have $1 \in M$. Let $c \in M$. Then

$$\begin{aligned} G(a, b + s(c)) &= G(a, s(b) + c) = G(a, s(b)) + G(a, c) \\ &= (G(a, b) + a) + G(a, c) = G(a, b) + (G(a, c) + a) \\ &= G(a, b) + G(a, s(c)). \end{aligned}$$

Hence M is an inductive set, $M = \mathcal{N}$, and

$$G(a, b + c) = G(a, b) + G(a, c) \text{ for all } a, b, c \in \mathcal{N}.$$

Using Lemma 1.10, one proves the distributivity of multiplication with respect to addition in the first argument similarly:

$$G(a + b, c) = G(a, c) + G(b, c) \text{ for all } a, b, c \in \mathcal{N}.$$

Associativity. Define

$$M = \{c \mid G(a, G(b, c)) = G(G(a, b), c) \ \forall a, b \in \mathcal{N}\}.$$

Since $G(a, G(b, 1)) = G(a, b) = G(G(a, b), 1)$, we have $1 \in M$. If $c \in M$, then

$$\begin{aligned} G(G(a, b), s(c)) &= G(G(a, b), c) + G(a, b) \\ &= G(a, G(b, c)) + G(a, b) = G(a, G(b, c) + b) = G(a, G(b, s(c))), \end{aligned}$$

showing that $s(c) \in M$. This establishes that M is inductive, whence

$$G(a, G(b, c)) = G(G(a, b), c) \text{ for all } a, b, c \in \mathcal{N}.$$

Commutativity. Define $M = \{a \mid G(1, a) = a\}$. Clearly, $1 \in M$. Let $a \in M$. Then

$$G(1, s(a)) = G(1, a) + 1 = a + 1 = s(a).$$

Hence M is an inductive set, $M = \mathcal{N}$, and $G(1, a) = a = G(a, 1)$ for all $a \in \mathcal{N}$. Now consider the set

$$M = \{a \mid G(a, b) = G(b, a) \forall b \in \mathcal{N}\}.$$

We have just shown that $1 \in M$. Let $a \in M$. Since

$$G(s(a), b) = G(a, b) + b = G(b, a) + b = G(b, s(a)),$$

it follows that $s(a) \in M$, whence, once again, $M = \mathcal{N}$, so that

$$G(a, b) = G(b, a) \text{ for all } a, b \in \mathcal{N}.$$

1.8. Check to see which properties of addition and multiplication justify the following chain of equalities:

$$2 \times 2 = 2 \times (1 + 1) = 2 \times 1 + 2 = 2 + (1 + 1) = (2 + 1) + 1 = 3 + 1 = 4.$$

1.9. If $a < b$, then there is a number $k \in \mathbb{N}$, such that $b = a + k$. Then $bc = (a + k)c = ac + kc = ac + l$, where $l \in \mathbb{N}$. Hence $bc > ac$.

1.10. 1) It is not difficult to verify that the formula $F'(\Phi(x), \Phi(y)) = \Phi(x + y)$ defines an operation on \mathcal{N}' satisfying conditions (Add1) and (Add2). Hence by Theorem 1.5, the mapping F' defines a unique addition operation on \mathcal{N}' . The second claim is established similarly.

2) If $x < y$, then there is an element $k \in \mathcal{N}$, such that $y = x + k$. Hence $\Phi(y) = \Phi(x + k) = \Phi(x) + \Phi(k) > \Phi(x)$.

1.11. We first show that for any element $a \in I_k$, we have $I_k \setminus a \cong I_{k-1}$.¹⁵ This is trivially true if $a = k$. Suppose that $a < k$, and consider the mapping $I_{k-1} \rightarrow I_k \setminus a$, defined by

$$I_{k-1} \rightarrow I_k \setminus a : b \mapsto \begin{cases} b, & \text{if } b < a, \\ b + 1, & \text{if } b \geq a. \end{cases}$$

It is easy to verify that this mapping is a one-to-one correspondence, establishing our claim. We now reformulate the theorem slightly: We shall prove that by induction that $I_k \not\cong I_l$ for each $k < l$. This is easy if $l = 2$, since then $k = 1$. Suppose inductively that it holds for some l , but is false for $l + 1$, that is, that $I_k \cong I_{l+1}$ for some $k \leq l$. Let $f : I_k \rightarrow I_{l+1}$ be a one-to-one correspondence realizing this. Writing $a := f(k)$, we infer that f defines (by restriction) a one-to-one correspondence between I_{k-1} and $I_{l+1} \setminus a \cong I_l$, which by the inductive hypothesis is impossible since $k - 1 < l$.

1.12. Consider the subsets $A = \{\frac{1}{n} \mid n \in \mathbb{N}\}$ and $B = \{\frac{1}{n+1} \mid n \in \mathbb{N}\}$ of $[0, 1]$ and $[0, 1)$, respectively. Clearly, $[0, 1] \setminus A = [0, 1) \setminus B$. On the other hand, the mapping from A to B defined by

$$A \rightarrow B : \frac{1}{n} \mapsto \frac{1}{n+1}$$

is obviously a one-to-one correspondence.

Remark: Try to prove that if $f : [0, 1] \rightarrow [0, 1)$ is a one-to-one correspondence, then it cannot be continuous!

¹⁵Here, for sets A and B , $A \cong B$ means that there is a one-to-one correspondence, or bijection, between A and B . *Trans.*

1.13. If the set B consists of just one element, that is, $|B| = 1$, then there is an obvious one-to-one correspondence between the sets $A \times B$ and A . Write $|A| = a$. Then

$$H(a, 1) = H(|A|, |B|) = |A \times B| = |A| = a,$$

so that H satisfies condition (Mult1).

Now let $|A| = a$, $|B| = b$. Consider the set $B' = B \cup \{b'\}$, where $b' \notin B$. It was noted earlier that then $|B'| = s(b)$. On the other hand, $A \times B' = (A \times B) \cup (A \times b')$, whence it follows that

$$H(a, s(b)) = H(|A|, |B'|) = |A \times B| + |A \times b'| = H(a, b) + a.$$

Hence the mapping H also satisfies condition (Mult2).