CHAPTER 3

RECURSION

3.1. The story of a little discovery

There is a traditional story about the little Gauss who later became the great mathematician Carl Friedrich Gauss. I particularly like the following version which I heard as a boy myself, and I do not care whether it is authentic or not.

"This happened when little Gauss still attended primary school. One day the teacher gave a stiff task: To add up the numbers 1, 2, 3, and so on, up to 20. The teacher expected to have some time for himself while the boys were busy doing that long sum. Therefore, he was disagreeably surprised as the little Gauss stepped forward when the others had scarcely started working, put his slate on the teacher's desk, and said, 'Here it is.' The teacher did not even look at little Gauss's slate, for he felt quite sure that the answer must be wrong, but decided to punish the boy severely for this piece of impudence. He waited till all the other boys had piled their slates on that of little Gauss, and then he pulled it out and looked at it. What was his surprise as he found on the slate just one number and it was the right one! What was the number and how did little Gauss find it?"

Of course, we do not know exactly how little Gauss did it and we shall never be able to know. Yet we are free to imagine something that looks reasonable. Little Gauss was, after all, just a child, although an exceptionally intelligent and precocious child. It came to him probably more naturally than to other children of his age to grasp the purpose of a question, to pay attention to the essential point. He just represented to himself more clearly and distinctly than the other youngsters *what is required*: to find the sum

1
2
3
and so on
.
.
.
20
—

He must have seen the problem differently, more completely, than the others, perhaps with some variations as the successive diagrams *A*, *B*, *C*, *D*, and *E* of Fig. 3.1 indicate. The original statement of the problem emphasizes the beginning of the series of numbers that should be added (*A*). Yet we could also emphasize the end (*B*) or, still better, emphasize the beginning and the end *equally* (*C*). Our attention may attach itself to the two extreme numbers, the very first and the very last, and we may observe some particular relation between them (*D*). Then the idea appears (*E*). Yes, numbers equally removed from the extremes add up all along to the same sum

$$1 + 20 = 2 + 19 = 3 + 18 = \cdots = 10 + 11 = 21$$

and, therefore, the grand total of the whole series is

$$10 \times 21 = 210$$

Did little Gauss really do it this way? I am far from asserting that. I say only that it would be natural to solve the problem in some such way. How did we solve it? Eventually we understood the situation (*E*), we "saw the truth clearly and distinctly," as Descartes would say, we saw a convenient, effortless, well-adapted manner of doing the required sum. How did we reach this final stage? At the outset, we hesitated between two opposite ways of conceiving the problem (*A* and *B*) which we finally

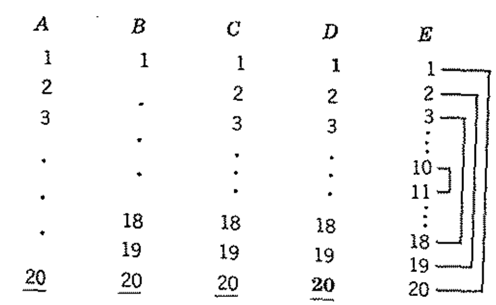

Fig. 3.1. Five phases of a discovery.

succeeded in merging into a better balanced conception (C). The original antagonism resolved into a happy harmony and the transition (D) to the final idea was quite close. Was little Gauss's final idea the same? Did he arrive at it passing through the same stages? Or did he skip some of them? Or did he skip all of them? Did he jump right away at the final conclusion? We cannot answer such questions. Usually a bright idea emerges after a longer or shorter period of hesitation and suspense. This happened in our case, and some such thing may have happened in the mind of little Gauss.

Let us generalize. Starting from the problem just solved and substituting the general positive integer n for the particular value 20, we arrive at the problem: *Find the sum S of the first n positive integers.*

Thus we seek the sum

$$S = 1 + 2 + 3 + \cdots + n$$

The idea developed in the foregoing (which might have been that of little Gauss) was to pair off the terms: a term that is at a certain distance from the beginning is paired with another term at the same distance from the end. If we are somewhat familiar with algebraic manipulations, we are easily led to the following modification of this scheme.

We write the sum twice, the second time reversing the original order:

$$\begin{aligned} S &= 1 + 2 + 3 + \cdots + (n-2) + (n-1) + n \\ S &= n + (n-1) + (n-2) + \cdots + 3 + 2 + 1 \end{aligned}$$

The terms paired with each other by the foregoing solution appear here conveniently aligned, one written under the other. Adding the two equations we obtain

$$\begin{aligned} 2S &= (n+1) + (n+1) + (n+1) + \cdots + (n+1) + (n+1) + (n+1) \\ 2S &= n(n+1) \\ S &= \frac{n(n+1)}{2} \end{aligned}$$

This is the general formula. For $n = 20$ it yields little Gauss's result, which is as it should be.

3.2. Out of the blue

Here is a problem similar to that solved in the foregoing section: *Find the sum of the first n squares.*

Let S stand for the required sum (we are no longer bound by the notation of the foregoing section) so that now

$$S = 1 + 4 + 9 + 16 + \cdots + n^2$$

The evaluation of this sum is not too obvious. Human nature prompts us to repeat a procedure that has succeeded before in a similar situation; and so, remembering the foregoing section, we may attempt to write the sum twice, reversing the order the second time:

$$\begin{aligned} S &= 1 + 4 + 9 + \cdots + (n-2)^2 + (n-1)^2 + n^2 \\ S &= n^2 + (n-1)^2 + (n-2)^2 + \cdots + 9 + 4 + 1 \end{aligned}$$

Yet the addition of these two equations, which was so successful in the foregoing case, leads us nowhere in the present case: our attempt fails, we undertook it with more optimism than understanding, our servile imitation of the chosen pattern was, let us confess, silly. (It was an overdose of mental inertia: our mind persevered in the same course, although this course should have been changed by the influence of circumstances.) Yet even such a misconceived trial need not be quite useless; it may lead us to a more adequate appraisal of the proposed problem: yes, it seems to be more difficult than the problem in the foregoing section.

Well, here is a solution. We start from a particular case of a well-known formula:

$$(n+1)^3 = n^3 + 3n^2 + 3n + 1$$

from which follows

$$(n+1)^3 - n^3 = 3n^2 + 3n + 1$$

This is valid for any value of n; write it down successively for $n = 1, 2, 3, \ldots, n$:

$$\begin{aligned} 2^3 - 1^3 &= 3\cdot1^2 + 3\cdot1 + 1 \\ 3^3 - 2^3 &= 3\cdot2^2 + 3\cdot2 + 1 \\ 4^3 - 3^3 &= 3\cdot3^2 + 3\cdot3 + 1 \\ &\cdots\cdots \\ (n+1)^3 - n^3 &= 3n^2 + 3n + 1 \end{aligned}$$

What is the obvious thing to do with these n equations? Add them! Thanks to conspicuous cancellations, the left-hand side of the resulting equation will be very simple. On the right-hand side we have to add three columns. The first column brings in S, the desired sum of the squares—that's good! The last column consists of n units—that is easy. The column in the middle brings in the sum of the first n numbers—but we know this sum from the foregoing section. We obtain

$$(n+1)^3 - 1 = 3S + 3\frac{n(n+1)}{2} + n$$

and in this equation everything is known (that is, expressed in terms of n)

except S, and so we can determine S from the equation. In fact, we find by straightforward algebra

$$2(n^3 + 3n^2 + 3n) = 6S + 3(n^2 + n) + 2n$$

$$S = \frac{2n^3 + 3n^2 + n}{6}$$

or finally

$$S = \frac{n(n + 1)(2n + 1)}{6}$$

How do you like this solution?

I shall be highly pleased with the reader who is displeased with the foregoing solution provided that he gives the right reason for his displeasure. What is wrong with the solution?

The solution is certainly correct. Moreover, it is efficient, clear, and short. Remember that the problem appeared difficult—we cannot reasonably expect a much clearer or shorter solution. There is, as far as I can see, just one valid objection: the solution appears *out of the blue*, pops up from nowhere. It is like a rabbit pulled out of a hat. Compare the present solution with that in the foregoing section. There we could visualize to some extent how the solution was discovered, we could learn a little about the ways of discovery, we could even gather some hope that some day we shall succeed in finding a similar solution by ourselves. Yet the presentation of the present section gives no hint about the sources of discovery, we are just hit on the head with the initial equation from which everything follows, and there is no indication how we could find this equation by ourselves. This is disappointing; we came here to learn problem solving—how could we learn it from the solution just presented?[1]

3.3. We cannot leave this unapplied

Yes, we *can* learn something important about problem solving from the foregoing solution. It is true, the presentation was not enlightening: the source of the invention remained hidden and so the solution appeared as a trick, a cunning device. Do you wish to know what is behind the trick? Try to *apply that trick* yourself and then you may find out. The device was so successful that we really cannot afford to leave it unapplied.

Let us start by generalizing. We bring both problems considered in the foregoing (in sections 3.1 and 3.2) under the same viewpoint by considering the sum of the kth powers of the first n natural numbers

$$S_k = 1^k + 2^k + 3^k + \cdots + n^k$$

[1] Cf. MPR, vol. 2, pp. 146–148, the sections on "*deus ex machina*" and "heuristic justification."

We found in the foregoing section

$$S_2 = \frac{n(n + 1)(2n + 1)}{6}$$

and before that

$$S_1 = \frac{n(n + 1)}{2}$$

to which we may add the obvious, but perhaps not useless, extreme case

$$S_0 = n$$

Starting from the particular cases $k = 0, 1$, and 2 we may raise the general problem: express S_k similarly. Surveying those particular cases, we may even conjecture that S_k can be expressed as a polynomial of degree $k + 1$ in n.

It is natural to try on the general case the trick that served us so well in the case $k = 2$. Yet let us first examine the next particular case $k = 3$. We have to imitate what we have seen in sect. 3.2 on the next higher level—this cannot be very difficult.

In fact, we start by applying the binomial formula with the next higher exponent 4:

$$(n + 1)^4 = n^4 + 4n^3 + 6n^2 + 4n + 1$$

from which follows

$$(n + 1)^4 - n^4 = 4n^3 + 6n^2 + 4n + 1$$

This is valid for any value of n; write it down successively for $n = 1, 2, 3, \ldots, n$:

$$\begin{aligned}
2^4 - 1^4 &= 4\cdot 1^3 + 6\cdot 1^2 + 4\cdot 1 + 1\\
3^4 - 2^4 &= 4\cdot 2^3 + 6\cdot 2^2 + 4\cdot 2 + 1\\
4^4 - 3^4 &= 4\cdot 3^3 + 6\cdot 3^2 + 4\cdot 3 + 1\\
&\cdots\cdots\cdots\cdots\\
(n + 1)^4 - n^4 &= 4n^3 + 6n^2 + 4n + 1
\end{aligned}$$

As before, we add these n equations. There are conspicuous cancellations on the left-hand side. On the right-hand side, there are four columns to add, and each column involves a sum of like powers of the first n integers; in fact, each column introduces another particular case of S_k:

$$(n + 1)^4 - 1 = 4S_3 + 6S_2 + 4S_1 + S_0$$

Yet we can already express S_2, S_1, and S_0 in terms of n, see above. Using those expressions, we transform our equation into

$$(n + 1)^4 - 1 = 4S_3 + 6\,\frac{n(n + 1)(2n + 1)}{6} + 4\,\frac{n(n + 1)}{2} + n$$

and in this equation everything is expressed in terms of n except S_3. What is needed now to determine S_3 is merely a little straightforward algebra:

$$4S_3 = (n+1)^4 - (n+1) - 2n(n+1) - n(n+1)(2n+1)$$
$$= (n+1)[n^3 + 3n^2 + 3n - 2n - n(2n+1)]$$
$$S_3 = \left[\frac{n(n+1)}{2}\right]^2$$

We have arrived at the desired result, and even the route seems instructive: having used the trick a second time, we may foresee a general outline. Remember that dictum of a famous pedagogue: "A method is a device which you use twice."[2]

3.4. Recursion

What was the salient feature of our work in the preceding sect. 3.3? In order to obtain S_3, we went back to the previously determined S_2, S_1, and S_0. This illuminates the "trick" of sect. 3.2 where we obtained S_2 by recurring to the previously determined S_1 and S_0.

In fact, we could use the same scheme to derive S_1 which we obtained in sect. 3.1 by a quite different method. By a most familiar formula

$$(n+1)^2 = n^2 + 2n + 1$$
$$(n+1)^2 - n^2 = 2n + 1$$

We list particular cases:

$$2^2 - 1^2 = 2\cdot 1 + 1$$
$$3^2 - 2^2 = 2\cdot 2 + 1$$
$$4^2 - 3^2 = 2\cdot 3 + 1$$
$$\cdots\cdots\cdots$$
$$(n+1)^2 - n^2 = 2n + 1$$

By adding we obtain

$$(n+1)^2 - 1 = 2S_1 + S_0$$

Of course, $S_0 = n$ and so

$$S_1 = \frac{(n+1)^2 - 1 - n}{2} = \frac{n(n+1)}{2}$$

which is the final result of sect. 3.1.

After having worked the scheme in the particular cases $k = 1$, 2, and 3,

[2] HSI, The traditional mathematics professor, p. 208.

we apply it without hesitation to the general sum S_k. We now need the binomial formula with the exponent $k + 1$:

$$(n+1)^{k+1} = n^{k+1} + \binom{k+1}{1}n^k + \binom{k+1}{2}n^{k-1} + \cdots + 1$$
$$(n+1)^{k+1} - n^{k+1} = (k+1)n^k + \binom{k+1}{2}n^{k-1} + \cdots + 1$$

We list particular cases:

$$2^{k+1} - 1^{k+1} = (k+1)1^k + \binom{k+1}{2}1^{k-1} + \cdots + 1$$
$$3^{k+1} - 2^{k+1} = (k+1)2^k + \binom{k+1}{2}2^{k-1} + \cdots + 1$$
$$4^{k+1} - 3^{k+1} = (k+1)3^k + \binom{k+1}{2}3^{k-1} + \cdots + 1$$
$$\cdots\cdots\cdots$$
$$(n+1)^{k+1} - n^{k+1} = (k+1)n^k + \binom{k+1}{2}n^{k-1} + \cdots + 1$$

By adding we obtain

$$(n+1)^{k+1} - 1 = (k+1)S_k + \binom{k+1}{2}S_{k-1} + \cdots + S_0$$

From this equation we can determine (express in terms of n) S_k provided that we have previously determined $S_{k-1}, S_{k-2}, \ldots, S_1$ and S_0. For example, as we have obtained in the foregoing expressions for S_0, S_1, S_2, and S_3, we could derive an expression for S_4 by straightforward algebra. Having obtained S_4, we could proceed to S_5, and so on.[3]

Thus, by following up the "trick" of sect. 3.2, which appeared "out of the blue," we have arrived at a pattern which deserves to be formulated and remembered with a view to further applications. When we are facing a well-ordered sequence (such as $S_0, S_1, S_2, S_3, \ldots, S_k, \ldots$) there is a chance to evaluate the terms of the sequence one at a time. We need two things.

First, the initial term of the sequence should be known somehow (the evaluation of S_0 was obvious).

Second, there should be some relation linking the general term of the sequence to the foregoing terms (S_k is linked to $S_0, S_1, \ldots, S_{k-1}$ by the final equation of the present section, foreshadowed by the "trick" of sect. 3.2).

Then we can find the terms one after the other, successively, *recursively*,

[3] This method is due to Pascal; see *Œuvres de Blaise Pascal*, edited by L. Brunschvicg and P. Boutroux, vol. 3, pp. 341–367.

by going back or recurring each time to the foregoing terms. This is the important pattern of *recursion*.

3.5. Abracadabra

The word "abracadabra" means something like "complicated nonsense." We use the word contemptuously today, but there was a time when it was a magic word, engraved on amulets in mysterious forms (like Fig. 3.2 in some respect), and people believed that such an amulet would protect the wearer from disease and bad luck.

In how many ways can you read the word "abracadabra" in Fig. 3.2? It is understood that we begin with the uppermost A (the north corner) and read down, passing each time to the next letter (southeast or southwest) till we reach the last A (the south corner).

The question is curious. Yet your interest may be really aroused if you notice that there is something familiar behind it. It may remind you of walking or driving in a city. Think of a city that consists of perfectly square blocks, where one-half of the streets run from northwest to southeast and the other streets (or avenues) crossing the former run from northeast to southwest. Reading the magic word of Fig. 3.2 corresponds to a zigzag path in the network of such streets. When you walk along the zigzag path emphasized in Fig. 3.3, you walk ten blocks from the initial A to the final A. There are several other paths which are ten blocks long between these two endpoints in this network of streets, but there is no path that would be shorter. *Find the number of the different shortest paths in the network between the given endpoints*—this is the general, really

```
          A
        B   B
      R   R   R
    A   A   A   A
  C   C   C   C   C
A   A   A   A   A   A
  D   D   D   D   D
    A   A   A   A
      B   B   B
        R   R
          A
```

Fig. 3.2. A magic word.

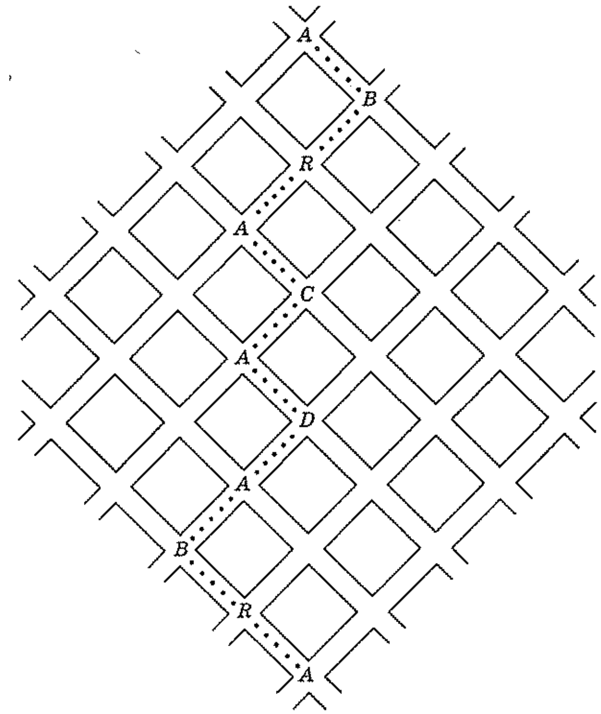

Fig. 3.3. The zigzag path is the shortest.

interesting, problem behind the curious particular problem about the magic word of Fig. 3.2.

A general formulation may have various advantages. It sometimes suggests an approach to the solution, and this happens in our case. *If you cannot solve the proposed problem* about Fig. 3.2 (probably you cannot), *try first to solve some simpler related problem.* At this point the general formulation may help: it suggests trying simpler cases that fall under it. In fact, if the two given corners are close enough to each other in the network (closer than the extreme A's in Fig. 3.3) it is easy to count the different zigzag paths between the two: you can draw each one after the other and survey all of them. Listen to this suggestion and pursue it systematically. Start from the point A and go downward. Consider first the points that you can reach by walking one block, then those to which you have to walk two blocks, then those which are three or four or more blocks away. Survey and count for each point the shortest zigzag paths

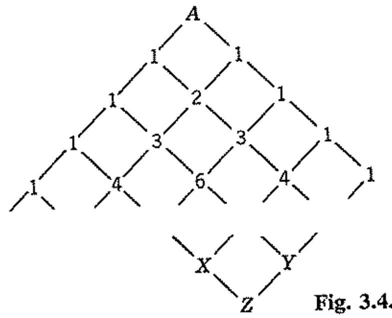

Fig. 3.4. Count the number of shortest zigzag paths.

that connect it with A. In Fig. 3.4 a few numbers so obtained are marked (but you should have obtained these numbers and a few more by yourself—check them at least). Observe these numbers—do you notice something?

If you have enough previous knowledge you may notice many things. Yet even if you have never before seen this array of numbers displayed by Fig. 3.4 you may notice an interesting relation: any number in Fig. 3.4 that is different from 1 is the sum of two other numbers in the array, of its northwest and northeast neighbors. For instance,

$$4 = 1 + 3, \qquad 6 = 3 + 3$$

You may discover this law by observation as a naturalist discovers the laws of his science by observation. Yet, after having discovered it, you should ask yourself: Why is that so? What is the reason?

The reason is simple enough. Consider three corners in your network, the points X, Y, and Z, the relative position of which is shown by Fig. 3.4: X is the northwest neighbor and Y the northeast neighbor of Z. If we wish to reach Z coming from A along a shortest path in the network, we must pass either through X or through Y. Once we have reached X, we can proceed hence to Z in just one way, and the same is true for proceeding from Y to Z. Therefore, the *total number of shortest paths from A to Z* is a sum of two terms: it *equals the number of shortest paths from A to X added to the number of those from A to Y.* This explains fully our observation and proves the general law.

Having clarified this basic point, we can extend the array of numbers in Fig. 3.4 by simple additions till we obtain the larger array in Fig. 3.5, the south corner of which yields the desired answer: we can read the magic word in Fig. 3.2 in exactly 252 different ways.

3.6. The Pascal triangle

By now the reader has probably recognized the numbers and their

peculiar configuration which we have examined in the foregoing section. The numbers in Fig. 3.4 are *binomial coefficients* and their triangular arrangement is usually called the *Pascal triangle*. (Pascal himself called it the "arithmetical triangle.") Further lines can be added to the triangle of Fig. 3.4 and, in fact, it can be extended indefinitely. The array in Fig. 3.5 is a square piece cut out of a larger triangle.

Some of the binomial coefficients and their triangular arrangement can be found in the writings of other authors before Pascal's *Traité du triangle arithmétique*. Still, the merits of Pascal in this matter are quite sufficient to justify the use of his name.

(1) We have to introduce a suitable *notation* for the numbers contained in the Pascal triangle; this is a step of major importance. For us each number attached to a point of this triangle has a geometric meaning: it indicates the number of different shortest zigzag paths from the apex of the triangle to that point. Each of these paths passes along the same number of blocks, let us say n blocks. Moreover, all these paths agree in the number of blocks described in the southwesterly direction and in the number of those in the southeasterly direction. Let l and r stand for these numbers, respectively (l to the left and r to the right—of course, downward in both cases). Obviously

$$n = l + r$$

If we give any two of the three numbers n, l, and r, the third is fully determined and so is the point to which they refer. (In fact, l and r are the rectangular coordinates of the point with respect to a system the origin of which is the apex of the Pascal triangle; one of the axes points southwest,

1
1 1
1 2 1
1 3 3 1
1 4 6 4 1
1 5 10 10 5 1
6 15 20 15 6
21 35 35 21
56 70 56
126 126
252

Fig. 3.5. A square from a triangle.

the other southeast.) For instance, for the last A of the path shown in Fig. 3.3

$$l = 5, \qquad r = 5, \qquad n = 10$$

and for the second B of the same path

$$l = 5, \qquad r = 3, \qquad n = 8$$

We shall denote by $\binom{n}{r}$ (this notation is due to Euler) the number of shortest zigzag paths from the apex of the Pascal triangle to the point specified by n (total number of blocks) and r (blocks to the right downward). For instance, see Fig. 3.5,

$$\binom{8}{3} = 56, \qquad \binom{10}{5} = 252$$

The symbols for the numbers contained in Fig. 3.4 are assembled in Fig. 3.6. The symbols with the same number upstairs (the same n) are horizontally aligned (along the nth "base"—the base of a right triangle). The symbols with the same number downstairs (the same r) are obliquely aligned (along the rth "avenue"). The fifth avenue forms one of the sides of the square in Fig. 3.5—the opposite side is formed by the 0th avenue (but you may call it the borderline, or Riverside Drive, if you prefer to do so). The fourth base is emphasized in Fig. 3.4.

(2) Besides the geometric aspect, the Pascal triangle also has a computational aspect. All the numbers along the boundary (0th street, 0th avenue, and their common starting point) are equal to 1 (it is obvious that

$$\begin{array}{ccccccccc} & & & & \binom{0}{0} & & & & \\ & & & \binom{1}{0} & & \binom{1}{1} & & & \\ & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\ & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\ \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} \end{array}$$

$$\begin{array}{ccc} \binom{n}{r-1} & & \binom{n}{r} \\ & \binom{n+1}{r} & \end{array}$$

Fig. 3.6. Symbolic Pascal triangle.

there is just one shortest path to these street corners from the starting point). Therefore,

$$\binom{n}{0} = \binom{n}{n} = 1$$

It is appropriate to call this relation the *boundary condition* of the Pascal triangle.

Any number inside the Pascal triangle is situated along a certain horizontal row, or base. We compute a number of the $(n + 1)$th base by going back, or recurring, to two neighboring numbers of the nth base:

$$\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1}$$

see Fig. 3.6. It is appropriate to call this equation the *recursion formula* of the Pascal triangle.

From the computer's standpoint the numbers $\binom{n}{r}$ are determined (or defined, if you wish) by the recursion formula and the boundary condition of the Pascal triangle.

3.7. Mathematical induction

When we compute a number in the Pascal triangle by using the recursion formula, we have to rely on the previous knowledge of two numbers of the foregoing base. It would be desirable to have a scheme of computation independent of such previous knowledge. There is a well-known formula, which we shall call the *explicit formula* for binomial coefficients, that yields such an independent computation:

$$\binom{n}{r} = \frac{n(n-1)(n-2)\cdots(n-r+1)}{1 \cdot 2 \cdot 3 \cdots r}$$

Pascal's treatise contains the explicit formula (stated in words, not in our modern notation). Pascal does not say how he has discovered it and we shall not speculate too much how he might have discovered it. (Perhaps he just guessed it first—we often find such things by observation and tentative generalization of the observed; see the remark in the solution of ex. 3.39.) Yet Pascal gives a remarkable proof for the explicit formula and we wish to devote our full attention to his method of proof.[4]

We need a preliminary remark. The explicit formula does not apply,

[4] Cf. Pascal's *Œuvres l.c.* footnote 3, pp. 455–464, especially pp. 456–457. The following presentation takes advantage of modern notation and modifies less essential details.

as it stands, to the case $r = 0$. Yet we lay down the rule that, if $r = 0$, it should be interpreted as

$$\binom{n}{0} = 1$$

The explicit formula does apply to the case $r = n$ and yields

$$\binom{n}{n} = \frac{n(n-1)\cdots\ 2\cdot 1}{1\cdot 2\ \cdots(n-1)n} = 1$$

which is the correct result. Therefore, we have to prove the explicit formula only for $0 < r < n$, that is, in the interior of the Pascal triangle where we can use the recursion formula. Now, we quote Pascal, with unessential modifications some of which will be included in square brackets [ ].

Although this proposition [the explicit formula] contains infinitely many cases I shall give for it a very short proof, supposing two lemmas.

The first lemma asserts that the proposition holds for the first base, which is obvious. [The explicit formula is valid for $n = 1$, because, in this case, all possible values of r, $r = 0$ and $r = 1$, fall under the preliminary remark.]

The second lemma asserts this: if the proposition happens to be valid for any base [for any value n] it is necessarily valid for the next base [for $n + 1$].

We see hence that the proposition holds necessarily for all values of n. For it is valid for $n = 1$ by virtue of the first lemma; therefore, for $n = 2$ by virtue of the second lemma; therefore, for $n = 3$ by virtue of the same, and so on *ad infinitum*.

And so nothing remains but to prove the second lemma.

In accordance with the statement of the second lemma, we assume that the explicit formula is valid for the nth base, that is, for a certain value of n and all compatible values of r (for $r = 0, 1, 2, \ldots, n$). In particular, along with

$$\binom{n}{r} = \frac{n(n-1)\cdots(n-r+2)(n-r+1)}{1\cdot 2\ \cdots\ (r-1)\ \cdot\ r}$$

we also have (if $r \geqq 1$)

$$\binom{n}{r-1} = \frac{n(n-1)\cdots(n-r+2)}{1\cdot 2\ \cdots\ (r-1)}$$

Adding these two equations and using the recursion formula, we derive as a necessary consequence

$$\binom{n+1}{r} = \binom{n}{r} + \binom{n}{r-1} = \frac{n(n-1)\cdots(n-r+2)}{1\cdot 2\ \cdots\ (r-1)}\left[\frac{n-r+1}{r} + 1\right]$$

$$= \frac{n(n-1)\cdots(n-r+2)}{1\cdot 2\ \cdots\ (r-1)}\cdot\frac{n+1}{r}$$

$$= \frac{(n+1)n(n-1)\cdots(n-r+2)}{1\cdot 2\cdot 3\ \cdots\ r}$$

That is, the validity of the explicit formula for a certain value of n involves its validity for $n + 1$. This is precisely what the second lemma asserts—we have proved it.

The words of Pascal which we have quoted are of historic importance because his proof is the first example of a fundamental pattern of reasoning which is usually called *mathematical induction.*

This pattern of reasoning deserves further study.[5] If carelessly introduced, reasoning by mathematical induction may puzzle the beginner; in fact, it may appear as a devilish trick.

You know, of course, that the devil is dangerous: if you give him the little finger, he takes the whole hand. Yet Pascal's second lemma does exactly this: by admitting the first lemma you give just one finger, the case $n = 1$. Yet then the second lemma also takes your second finger (the case $n = 2$), then the third finger ($n = 3$), then the fourth, and so on, and finally takes all your fingers even if you happen to have infinitely many.

3.8. Discoveries ahead

After the work in the three foregoing sections, we now have three different approaches to the numbers in the Pascal triangle, the binomial coefficients.

(1) *Geometrical approach.* A binomial coefficient is the number of the different shortest zigzag paths between two given corners in a network of streets.

(2) *Computational approach.* The binomial coefficients can be defined by their recursion formula and their boundary condition.

(3) *Explicit formula.* We have proved it, by Pascal's method, in sect. 3.7.

The name of the numbers considered reminds us of another approach.

(4) *Binomial theorem.* For indeterminate (or variable) x and any nonnegative integer n we have the identity

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$$

For a proof, see ex. 3.1.

There are still other approaches to the numbers in the Pascal triangle which play, in fact, a role in a great many interesting questions and possess a great many interesting properties. "This table of numbers has eminent

[5] HSI, Induction and mathematical induction, pp. 114–121; MPR, vol. 1, pp. 108–120.

and admirable properties" wrote Jacob Bernoulli in his *Ars Conjectandi* (Basle 1713; see Second Part, Chapter III, p. 88). "We have just shown that the essence of combinations is concealed in it [see ex. 3.22–3.27] but those who are more intimately acquainted with Geometry know also that capital secrets of all Mathematics are hidden in it." Times have changed and many things hidden in Bernoulli's time are clearly seen today. Still, the reader who wants instructive, and perhaps fascinating, exercise has an excellent opportunity: he has an excellent chance to discover something by observing the numbers in the Pascal triangle and combining his observations with one or the other or several approaches. There are so many possibilities—some of them should be favorable.

By the way, we have broached another subject in the first four sections of the chapter (sum of like powers of the first n integers). Moreover, we have encountered two important general patterns (recursion and mathematical induction) which we still should apply to more examples if we wish to understand them thoroughly. And so there are still more prospects ahead.

3.9. Observe, generalize, prove, and prove again

Let us return to our starting point and have another look at it.

(1) We started from the magic word of Fig. 3.2 and Fig. 3.3, or rather from a problem concerning that word. What was the unknown? The number of shortest zigzag paths in that network of streets from the first A to the last A, that is, from the north corner of the square to its south corner. Such a zigzag path must cross somewhere the horizontal diagonal of the square. There are six possible crossing points (street corners, A's) along the horizontal diagonal. There are, therefore, six different kinds of zigzag paths in our problem—how many paths are there of each kind? We have here a *new problem*.

Let us be specific. Take a definite crossing point on that horizontal diagonal, for instance the third point from the left ($l = 3$, $r = 2$, $n = 5$ in the notation of sect. 3.6). A zigzag path crossing this chosen point consists of two sections: the upper section starts from the north corner of the square and ends in the chosen point, the lower section starts from the chosen point and ends in the south corner; see Fig. 3.3. We have found before (see Fig. 3.5) the number of the different upper sections; it is

$$\binom{5}{2} = 10$$

The number of the different lower sections is the same. Now any upper

section can be combined with any lower section to form a full path [as suggested by Fig. 3.7(III)]. Therefore, the number of such paths is

$$\binom{5}{2}^2 = 100$$

Of course, the number of zigzag paths crossing the horizontal diagonal at any other given point can be similarly computed. Hence we find a new solution of our original problem: we can read the magic word of Fig. 3.2 in exactly

$$1 + 25 + 100 + 100 + 25 + 1$$

different ways. This sum must agree with the result found at the end of sect. 3.5; in fact, it equals 252.

(2) *Generalization.* One side of the square considered in Fig. 3.3 consists of five blocks. In generalizing (passing from 5 to n) we find that

$$\binom{n}{0}^2 + \binom{n}{1}^2 + \binom{n}{2}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}$$

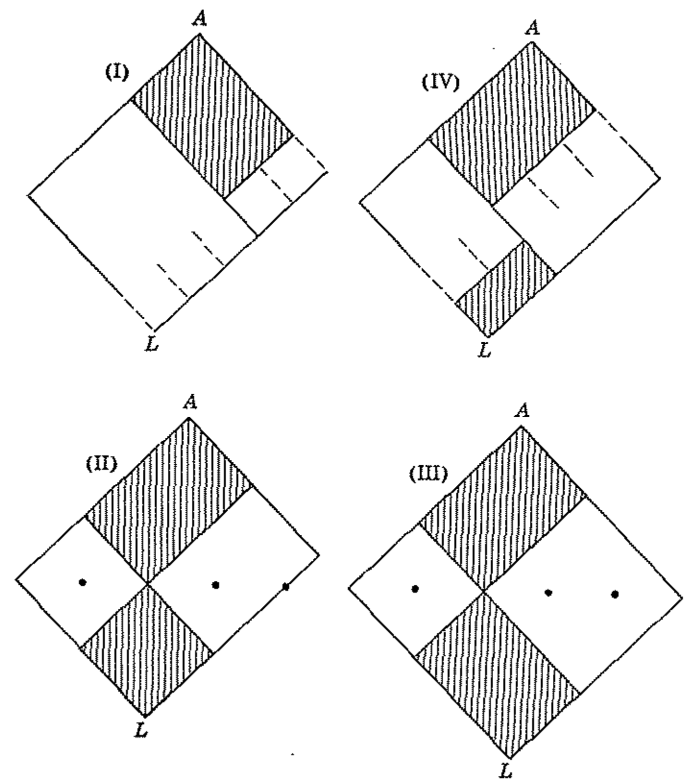

Fig. 3.7. Suggestions.

"The sum of the squares of the numbers in the nth base of the Pascal triangle is equal to the number in the middle of the $2n$th base." Our reasoning under (1) essentially proves this general statement. It is true, we have considered the special case $n = 5$ (we have even considered a special point of the fifth base) but there is no particular virtue (and no misleading peculiarity) in the special case considered. And so our reasoning is generally valid. Yet it may be a useful exercise for the reader to repeat the reasoning with special attention to its generality—he has to say n instead of 5.[6]

(3) *Another approach.* Still, the result is surprising. We would understand it better if we could attain it from another side.

Surveying the various approaches listed in sect. 3.8, we may try to link our result to the binomial formula. There is, in fact, a connection:

$$\begin{aligned}(1+x)^{2n} &= \cdots + \binom{2n}{n}x^n + \cdots \\ &= (1+x)^n(1+x)^n \\ &= \left[\binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n\right]\cdot \\ &\quad \left[\binom{n}{n} + \cdots + \binom{n}{2}x^{n-2} + \binom{n}{1}x^{n-1} + \binom{n}{0}x^n\right].\end{aligned}$$

Let us focus the coefficient of x^n. On the right-hand side of the first line the coefficient of x^n is the right-hand side of the general equation given under (2) for which we are seeking a second proof. Now let us turn to the product of the two factors which are displayed on the last two lines; in writing them we made use of the symmetry of the binomial coefficients:

$$\binom{n}{r} = \binom{n}{n-r}$$

Now, in this product, the coefficient of x^n is obviously the left-hand side of the equation under (2) which we are about to prove. And here is the proof: the coefficient of x^n must be the same in both cases since we have here an identity in x.

Examples and Comments on Chapter 3

First Part

The examples and comments of this first part are connected with the first four sections.

[6] We have here a *representative* special case; see MPR, vol. 1, p. 25, ex. 10.

3.1. Prove the binomial theorem stated in sect. 3.8(4) (and used in sect. 3.4). (Use mathematical induction. Which one of the first three approaches mentioned in sect. 3.8 appears the most appropriate for the present purpose?)

3.2. *A particular case equivalent to the general case.* The identity asserted in sect. 3.8(4) and proved in ex. 3.1 follows as a particular case ($a = 1, b = x$) from the more general identity

$$(a+b)^n = \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n}b^n$$

Show that, conversely, the general identity also follows from that particular case.[7]

3.3. In the first three sections of this chapter we have computed S_k (defined in sect. 3.3) for $k = 1, 2, 3$; the case $k = 0$ is obvious. Comparing these expressions, we may be led to the general theorem: *S_k is a polynomial in n of degree $k + 1$ and the coefficient of its highest term is $1/(k + 1)$.*

This theorem which asserts that

$$S_k = \frac{n^{k+1}}{k+1} + \cdots$$

(where the dots indicate terms of lower degree in n) played an important role in the history of the integral calculus.

Prove the theorem; use mathematical induction.

3.4. We can guess an expression for S_4 by computing numerically the ratio S_4/S_2 for a few small values of n. In fact, for

$$n = 1, \quad 2, \quad 3, \quad 4, \quad 5$$

$$\frac{S_4}{S_2} = 1, \quad \frac{17}{5}, \quad 7, \quad \frac{59}{5}, \quad \frac{89}{5}$$

For the sake of uniformity we write rather

$$\frac{5}{5}, \quad \frac{17}{5}, \quad \frac{35}{5}, \quad \frac{59}{5}, \quad \frac{89}{5}$$

The numerators are close to multiples of 6; in fact, they are

$$6\cdot 1 - 1, \quad 6\cdot 3 - 1, \quad 6\cdot 6 - 1, \quad 6\cdot 10 - 1, \quad 6\cdot 15 - 1$$

You should recognize the numbers

$$1, \quad 3, \quad 6, \quad 10, \quad 15$$

If you succeed in constructing an expression for S_4, prove it, independently of sect. 3.4, by mathematical induction.[8]

3.5. Compute S_4, independently of ex. 3.4, by the method indicated in sect. 3.4.

[7] Such equivalence of the particular and the general may seem bewildering to the philosopher or to the beginner, but is, in fact, quite usual in mathematics; see MPR, vol. 1, p. 23, ex. 3 and ex. 4.

[8] For broader discussion of a very similar simpler case see MPR, vol. 1, pp. 108–110.

3.6. Show that

$$n = S_0$$
$$n^2 = 2S_1 - S_0$$
$$n^3 = 3S_2 - 3S_1 + S_0$$
$$n^4 = 4S_3 - 6S_2 + 4S_1 - S_0$$

and generally

$$n^k = \binom{k}{1}S_{k-1} - \binom{k}{2}S_{k-2} + \binom{k}{3}S_{k-3} - \cdots + (-1)^{k-1}\binom{k}{k}S_0$$

(This is similar to, but different from, the principal formula of sect. 3.4.)

3.7. Show that

$$S_1 = S_1$$
$$2S_1^2 = 2S_3$$
$$4S_1^3 = 3S_5 + S_3$$
$$8S_1^4 = 4S_7 + 4S_5$$

and generally, for $k = 1, 2, 3, \ldots$,

$$2^{k-1}S_1^k = \binom{k}{1}S_{2k-1} + \binom{k}{3}S_{2k-3} + \binom{k}{5}S_{2k-5} + \cdots$$

The last term on the right-hand side is S_k or kS_{k+1} according as k is odd or even.

(This is similar to ex. 3.6 where, in fact, we could substitute $S_0{}^k$ for n^k.)

3.8. Show that

$$3S_2 = 3S_2$$
$$6S_2S_1 = 5S_4 + S_2$$
$$12S_2S_1^2 = 7S_6 + 5S_4$$
$$24S_2S_1^3 = 9S_8 + 14S_6 + S_4$$

and generally, for $k = 1, 2, 3, \ldots$,

$$3\cdot 2^{k-1}S_2S_1{}^{k-1} = \left[\binom{k}{0} + 2\binom{k}{1}\right]S_{2k} + \left[\binom{k}{2} + 2\binom{k}{3}\right]S_{2k-2} + \cdots$$

the last term on the right-hand side is $(k + 2)S_{k+1}$ or S_k according as k is odd or even.

3.9. Show that

$$S_3 = S_1^2$$
$$S_5 = S_1^2(4S_1 - 1)/3$$
$$S_7 = S_1^2(6S_1^2 - 4S_1 + 1)/3$$

and generally that S_{2k-1} is a polynomial in $S_1 = n(n + 1)/2$, of degree k, divisible by S_1^2 provided that $2k - 1 \geqq 3$. (This generalizes the result of sect. 3.3.)

3.10. Show that

$$S_4 = S_2(6S_1 - 1)/5$$
$$S_6 = S_2(12S_1^2 - 6S_1 + 1)/7$$
$$S_8 = S_2(40S_1^3 - 40S_1^2 + 18S_1 - 3)/15$$

and generally that S_{2k}/S_2 is a polynomial in S_1 of degree $k - 1$. (This generalizes a result encountered in the solution of ex. 3.4.)

3.11. We introduce the notation

$$1^k + 2^k + 3^k + \cdots + n^k = S_k(n)$$

which is more explicit (or specific) than the one introduced in sect. 3.3; k stands for a non-negative integer and n for a positive integer.

We now extend the range of n (but not the range of k): we let $S_k(x)$ denote the polynomial in x of degree $k + 1$ that coincides with $S_k(n)$ for $x = 1, 2, 3, \ldots$; for example,

$$S_3(x) = \frac{x^2(x + 1)^2}{4}$$

Prove that for $k \geqq 1$ (not for $k = 0$)

$$S_k(-x - 1) = (-1)^{k-1}S_k(x)$$

3.12. Find $1 + 3 + 5 + \cdots + (2n - 1)$, the sum of the first n odd numbers. (List as many different approaches as you can.)

3.13. Find $1 + 9 + 25 + \cdots + (2n - 1)^2$.

3.14. Find $1 + 27 + 125 + \cdots + (2n - 1)^3$.

3.15. (Continued) Generalize.

3.16. Find $2^2 + 5^2 + 8^2 + \cdots + (3n - 1)^2$

3.17. (Continued) Generalize.

3.18. Find a simple expression for

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

(Of course, you should try to use suitable points from the foregoing work. What has better prospects to be usable: the results or the method?)

3.19. Consider the $\frac{n(n-1)}{2}$ differences

$$\begin{array}{lllll} 2-1, & & & & \\ 3-1, & 3-2 & & & \\ 4-1, & 4-2, & 4-3 & & \\ \cdot\ \cdot & \cdot\ \cdot & \cdot\ \cdot & & \\ n-1, & n-2, & n-3 & ,\ldots, & n-(n-1) \end{array}$$

and compute (*a*) their sum, (*b*) their product, and (*c*) the sum of their squares.

3.20. Define $E_1, E_2, E_3 \ldots$ by the identity

$$x^n - E_1x^{n-1} + E_2x^{n-2} - \cdots + (-1)^nE_n = (x - 1)(x - 2)(x - 3)\cdots(x - n)$$

Show that

$$E_1 = \frac{n(n+1)}{2}$$

$$E_2 = \frac{(n-1)n(n+1)(3n+2)}{24}$$

$$E_3 = \frac{(n-2)(n-1)n^2(n+1)^2}{48}$$

$$E_4 = \frac{(n-3)(n-2)(n-1)n(n+1)(15n^3+15n^2-10n-8)}{5760}$$

and show in general that E_k [which should rather be denoted by $E_k(n)$ since it depends on n] is a polynomial of degree $2k$ in n.

[The knowledge of a certain proposition of algebra may be a great help; E_k is the so-called kth elementary symmetric function of the first n integers the sum of the kth powers of which is $S_k = S_k(n)$. Check $E_k(k) = k!$]

3.21. *Two forms of mathematical induction.* A typical proposition A that is accessible to proof by mathematical induction has an infinity of cases $A_1, A_2, A_3, \ldots, A_n, \ldots$; in fact, A is equivalent to the simultaneous assertion of $A_1, A_2, A_3, \ldots$. For instance, if A is the binomial theorem, A_n asserts the validity of the identity.

$$(1+x)^n = \binom{n}{0} + \binom{n}{1}x + \binom{n}{2}x^2 + \cdots + \binom{n}{n}x^n$$

see ex. 3.1; the binomial theorem asserts, in fact, that this identity holds for $n = 1, 2, 3, 4, \ldots$.

Let us consider three statements about the sequence of propositions $A_1, A_2, A_3, \ldots$:

(I) A_1 is true.

(II*a*) A_n implies A_{n+1}.

(II*b*) $A_1, A_2, A_3, \ldots A_{n-1}$ and A_n jointly imply A_{n+1}.

Now we can distinguish two procedures.

(*a*) We can conclude from (I) and (II*a*) that A_n is true generally, for $n = 1, 2, 3, \ldots$; we drew this conclusion, with Pascal, in sect. 3.7.

(*b*) We can conclude from (I) and (II*b*) that A_n is true generally, for $n = 1, 2, 3, \ldots$; we proceeded so in the solution of ex. 3.3.

You may feel that the difference between the procedures (*a*) and (*b*) is more in the form than in the essence. Could you clarify this feeling and propose a clear argument?

Second Part

3.22. Ten boys went camping together, Bernie, Ricky, Abe, Charlie, Al, Dick, Alex, Bill, Roy, and Artie. In the evening they divided into two teams of five boys each: one team put up the tent, the other team cooked the supper. In how many different ways is such a division into two teams possible? (Can a magic word help you?)

3.23. Show that a set of n individuals has $\binom{n}{r}$ different subsets of r individuals. [In more traditional language: the number of *combinations* of n objects taken r at a time is $\binom{n}{r}$.]

3.24. Given n points in the plane in "general position" so that no three points lie on the same straight line. How many straight lines can you draw by joining two given points? How many triangles can you form with vertices chosen among the given points?

3.25. (Continued) Formulate and solve an analogous problem in space.

3.26. Find the number of the diagonals of a convex polygon with n sides.

3.27. Find the number of intersections of the diagonals of a convex polygon of n sides. Consider only points of intersection inside the polygon, and assume that the polygon is "general" so that no three diagonals have a common point.

3.28. A polyhedron has six faces. (We may consider the polyhedron as irregular so that no two of its faces are congruent.) The faces should be painted, one red, two blue, and three brown. In how many different ways can this be done?

3.29. A polyhedron has n faces (no two of which are congruent.) Of these faces, r should be painted red, s sapphire, and t tan; we suppose that $r + s + t = n$. In how many different ways can this be done?

3.30. (Continued) Generalize.

Third Part

In solving some of the following problems, the reader may consider, and choose between, several approaches. (See sect. 3.8; the combinatorial interpretation of the binomial coefficients, cf. ex. 3.23, provides one more access.) The importance of approaching the same problem from several sides was emphasized by Leibnitz. Here is a free translation of one of his remarks: "In comparing two different expressions of the same quantity, you may find an unknown; in comparing two different derivations of the same result, you may find a new method."

3.31. Show in as many ways as you can that

$$\binom{n}{r} = \binom{n}{n-r}$$

3.32. Consider the sum of the numbers along a base of the Pascal triangle:

$$\begin{aligned} 1 &= 1 \\ 1+1 &= 2 \\ 1+2+1 &= 4 \\ 1+3+3+1 &= 8 \end{aligned}$$

These facts seem to suggest a general theorem. Can you guess it? Having guessed it, can you prove it? Having proved it, can you devise another proof?

3.33. Observe

$$\begin{aligned} 1-1 &= 0 \\ 1-2+1 &= 0 \\ 1-3+3-1 &= 0 \\ 1-4+6-4+1 &= 0 \end{aligned}$$

generalize, prove, and prove again.

3.34. Consider the sum of the first six numbers along the third avenue of the Pascal triangle:

$$1 + 4 + 10 + 20 + 35 + 56 = 126$$

Locate this sum in the Pascal triangle, try to observe analogous facts, generalize, prove, and prove again.

3.35. Add the thirty-six numbers displayed in Fig. 3.5, try to locate their sum in the Pascal triangle, formulate a general theorem, and prove it. (Adding so many numbers is a boring task—in doing it cleverly, you may easily catch the essential idea.)

3.36. Try to recognize and locate in the Pascal triangle the numbers involved in the following relation:

$$1\cdot 1 + 5\cdot 4 + 10\cdot 6 + 10\cdot 4 + 5\cdot 1 = 126$$

Observe (or remember) analogous cases, generalize, prove, prove again.

3.37. Try to recognize and locate in the Pascal triangle the numbers involved in the following relation:

$$6\cdot 1 + 5\cdot 3 + 4\cdot 6 + 3\cdot 10 + 2\cdot 15 + 1\cdot 21 = 126$$

Observe (or remember) analogous cases, generalize, prove, prove again.

3.38. Fig. 3.8 shows the first four from an infinite sequence of figures each of which is an assemblage of equal circles into an equilateral triangular shape. Any circle that is not on the rim of the assemblage touches six surrounding circles. In the nth figure there are n circles aligned along each side of the triangular assemblage and the total number of circles in this nth figure is termed the nth *triangular number*. Express the nth triangular number in terms of n and locate it in the Pascal triangle.

3.39. Replace in Fig. 3.8 each circle by a sphere (a marble) of which the circle forms the equator. Fix 10 marbles arranged as in Fig. 3.8 on a horizontal plane, place 6 marbles on top (they fit neatly into the interstices) as a second layer, add 3 marbles on top of these as a third layer and place finally 1 marble on the very top. This configuration of

$$1 + 3 + 6 + 10 = 20$$

marbles is so related to a regular tetrahedron as each of the assemblages of circles shown by Fig. 3.8 is related to a certain equilateral triangle: 20 is the fourth *pyramidal number*. Express the nth pyramidal number in terms of n and locate it in the Pascal triangle.

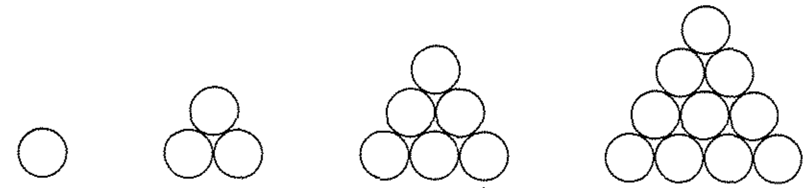

Fig. 3.8. The first four triangular numbers.

3.40. You can build a pyramidal pile of marbles in another manner: begin with a layer of n^2 marbles, arranged in a square as in Fig. 3.9, place on top of it a second layer of $(n-1)^2$ marbles, then $(n-2)^2$ marbles, and so on, and finally just one marble on the very top. How many marbles does the pile contain?

3.41. Interpret the product

$$\binom{n_1}{r_1}\binom{n_2}{r_2}\binom{n_3}{r_3}\cdots\binom{n_h}{r_h}$$

as the number of a certain set of zigzag paths in a network of streets.

3.42. All the shortest zigzag paths from the apex of the Pascal triangle to the point specified by n (the total number of blocks) and r (blocks to the right downward) have a point in common with the line of symmetry of the Pascal triangle (from the first A to the last A in Fig. 3.3) namely their common initial

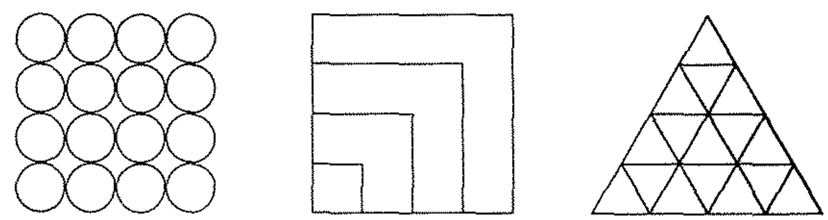

Fig. 3.9. The fourth square number.