How To Solve It

A New Aspect of Mathematical Method

G. POLYA
Stanford University

SECOND EDITION

Doubleday Anchor Books
Doubleday & Company, Inc.
Garden City, New York

Contents

More examples

PART II. HOW TO SOLVE IT

PART III. SHORT DICTIONARY OF HEURISTIC

† Contains only cross-references.

† Contains only cross-references.

PART IV. PROBLEMS, HINTS, SOLUTIONS

COVER BY GEORGE GIUSTI

TYPOGRAPHY BY EDWARD GOREY

How To Solve It was originally published by Princeton University Press in 1945. The Anchor Books edition is published by arrangement with Princeton University Press.

Anchor Books edition: 1957

From the Preface to the First Printing

A great discovery solves a great problem but there is a grain of discovery in the solution of any problem. Your problem may be modest; but if it challenges your curiosity and brings into play your inventive faculties, and if you solve it by your own means, you may experience the tension and enjoy the triumph of discovery. Such experiences at a susceptible age may create a taste for mental work and leave their imprint on mind and character for a lifetime.

Thus, a teacher of mathematics has a great opportunity. If he fills his allotted time with drilling his students in routine operations he kills their interest, hampers their intellectual development, and misuses his opportunity. But if he challenges the curiosity of his students by setting them problems proportionate to their knowledge, and helps them to solve their problems with stimulating questions, he may give them a taste for, and some means of, independent thinking.

Also a student whose college curriculum includes some mathematics has a singular opportunity. This opportunity is lost, of course, if he regards mathematics as a subject in which he has to earn so and so much credit and which he should forget after the final examination as quickly as possible. The opportunity may be lost even if the student has some natural talent for mathematics because he, as everybody else, must discover his talents and tastes; he cannot know that he likes raspberry pie if he has never tasted raspberry pie. He may manage to find out, however, that a mathematics problem may be as much fun as a crossword puzzle, or that vigorous mental work may be an exercise as desirable as a fast game of tennis. Having tasted the pleasure in mathematics he will not forget it easily and then there is a good chance that mathematics will become something for him: a hobby, or a tool of his profession, or his profession, or a great ambition.

The author remembers the time when he was a student himself, a somewhat ambitious student, eager to understand a little mathematics and physics. He listened to lectures, read books, tried to take in the solutions and facts presented, but there was a question that disturbed him again and again: "Yes, the solution seems to work, it appears to be correct; but how is it possible to invent such a solution? Yes, this experiment seems to work, this appears to be a fact; but how can people discover such facts? And how could I invent or discover such things by myself?" Today the author is teaching mathematics in a university; he thinks or hopes that some of his more eager students ask similar questions and he tries to satisfy their curiosity. Trying to understand not only the solution of this or that problem but also the motives and procedures of the solution, and trying to explain these motives and procedures to others, he was finally led to write the present book. He hopes that it will be useful to teachers who wish to develop their students' ability to solve problems, and to students who are keen on developing their own abilities.

Although the present book pays special attention to the requirements of students and teachers of mathematics, it should interest anybody concerned with the ways and means of invention and discovery. Such interest may be more widespread than one would assume without reflection. The space devoted by popular newspapers and magazines to crossword puzzles and other riddles seems to show that people spend some time in solving unpractical problems. Behind the desire to solve this or that problem that confers no material advantage, there may be a deeper curiosity, a desire to understand the ways and means, the motives and procedures, of solution.

The following pages are written somewhat concisely, but as simply as possible, and are based on a long and serious study of methods of solution. This sort of study, called *heuristic* by some writers, is not in fashion nowadays but has a long past and, perhaps, some future.

Studying the methods of solving problems, we perceive another face of mathematics. Yes, mathematics has two faces; it is the rigorous science of Euclid but it is also something else. Mathematics presented in the Euclidean way appears as a systematic, deductive science; but mathematics in the making appears as an experimental, inductive science. Both aspects are as old as the science of mathematics itself. But the second aspect is new in one respect; mathematics "in statu nascendi," in the process of being invented, has never before been presented in quite this manner to the student, or to the teacher himself, or to the general public.

The subject of heuristic has manifold connections; mathematicians, logicians, psychologists, educationalists, even philosophers may claim various parts of it as belonging to their special domains. The author, well aware of the possibility of criticism from opposite quarters and keenly conscious of his limitations, has one claim to make: he has some experience in solving problems and in teaching mathematics on various levels.

The subject is more fully dealt with in a more extensive book by the author which is on the way to completion.

Stanford University, August 1, 1944

From the Preface to the Seventh Printing

I am glad to say that I have now succeeded in fulfilling, at least in part, a promise given in the preface to the first printing: The two volumes *Induction and Analogy in Mathematics* and *Patterns of Plausible Inference* which constitute my recent work *Mathematics and Plausible Reasoning* continue the line of thinking begun in *How to Solve It.*

Zurich, August 30, 1954

Preface to the Second Edition

The present second edition adds, besides a few minor improvements, a new fourth part, "Problems, Hints, Solutions."

As this edition was being prepared for print, a study appeared (Educational Testing Service, Princeton, N.J.; *cf. Time,* June 18, 1956) which seems to have formulated a few pertinent observations—they are not new to the people in the know, but it was high time to formulate them for the general public—: " . . . mathematics has the dubious honor of being the least popular subject in the curriculum . . . Future teachers pass through the elementary schools learning to detest mathematics . . . They return to the elementary school to teach a new generation to detest it."

I hope that the present edition, designed for wider diffusion, will convince some of its readers that mathematics, besides being a necessary avenue to engineering jobs and scientific knowledge, may be fun and may also open up a vista of mental activity on the highest level.

Zurich, June 30, 1956

The cover illustrates pp. 87–88 but also p. 208.

HOW TO SOLVE IT

UNDERSTANDING THE PROBLEM

First.

You have to *understand* the problem.

What is the unknown? What are the data? What is the condition?

Is it possible to satisfy the condition? Is the condition sufficient to determine the unknown? Or is it insufficient? Or redundant? Or contradictory?

Draw a figure. Introduce suitable notation.

Separate the various parts of the condition. Can you write them down?

DEVISING A PLAN

Second.

Find the connection between the data and the unknown. You may be obliged to consider auxiliary problems if an immediate connection cannot be found. You should obtain eventually a *plan* of the solution.

Have you seen it before? Or have you seen the same problem in a slightly different form?

Do you know a related problem? Do you know a theorem that could be useful?

Look at the unknown! And try to think of a familiar problem having the same or a similar unknown.

Here is a problem related to yours and solved before. Could you use it? Could you use its result? Could you use its method? Should you introduce some auxiliary element in order to make its use possible?

Could you restate the problem? Could you restate it still differently? Go back to definitions.

If you cannot solve the proposed problem try to solve first some related problem. Could you imagine a more accessible related problem? A more general problem? A more special problem? An analogous problem? Could you solve a part of the problem? Keep only a part of the condition, drop the other part; how far is the unknown then determined, how can it vary? Could you derive something useful from the data? Could you think of other data appropriate to determine the unknown? Could you change the unknown or the data, or both if necessary, so that the new unknown and the new data are nearer to each other?

Did you use all the data? Did you use the whole condition? Have you taken into account all essential notions involved in the problem?

CARRYING OUT THE PLAN

Third.

Carry out your plan.

Carrying out your plan of the solution, *check each step.* Can you see clearly that the step is correct? Can you prove that it is correct?

LOOKING BACK

Fourth.

Examine the solution obtained.

Can you *check the result?* Can you check the argument?

Can you derive the result differently? Can you see it at a glance?

Can you use the result, or the method, for some other problem?

How To Solve It

Introduction

The following considerations are grouped around the preceding list of questions and suggestions entitled "How to Solve It." Any question or suggestion quoted from it will be printed in *italics*, and the whole list will be referred to simply as "the list" or as "our list."

The following pages will discuss the purpose of the list, illustrate its practical use by examples, and explain the underlying notions and mental operations. By way of preliminary explanation, this much may be said: If, using them properly, you address these questions and suggestions to yourself, they may help you to solve your problem. If, using them properly, you address the same questions and suggestions to one of your students, you may help him to solve his problem.

The book is divided into four parts.

The title of the first part is "In the Classroom." It contains twenty sections. Each section will be quoted by its number in heavy type as, for instance, "section **7**." Sections **1** to **5** discuss the "Purpose" of our list in general terms. Sections **6** to **17** explain what are the "Main Divisions, Main Questions" of the list, and discuss a first practical example. Sections **18, 19, 20** add "More Examples."

The title of the very short second part is "How to Solve It." It is written in dialogue; a somewhat idealized teacher answers short questions of a somewhat idealized student.

The third and most extensive part is a "Short Dictionary of Heuristic"; we shall refer to it as the "Dictionary."

It contains sixty-seven articles arranged alphabetically. For example, the meaning of the term HEURISTIC (set in small capitals) is explained in an article with this title on page 112. When the title of such an article is referred to within the text it will be set in small capitals. Certain paragraphs of a few articles are more technical; they are enclosed in square brackets. Some articles are fairly closely connected with the first part to which they add further illustrations and more specific comments. Other articles go somewhat beyond the aim of the first part of which they explain the background. There is a key-article on MODERN HEURISTIC. It explains the connection of the main articles and the plan underlying the Dictionary; it contains also directions how to find information about particular items of the list. It must be emphasized that there is a common plan and a certain unity, because the articles of the Dictionary show the greatest outward variety. There are a few longer articles devoted to the systematic though condensed discussion of some general theme; others contain more specific comments, still others cross-references, or historical data, or quotations, or aphorisms, or even jokes.

The Dictionary should not be read too quickly; its text is often condensed, and now and then somewhat subtle. The reader may refer to the Dictionary for information about particular points. If these points come from his experience with his own problems or his own students, the reading has a much better chance to be profitable.

The title of the fourth part is "Problems, Hints, Solutions." It proposes a few problems to the more ambitious reader. Each problem is followed (in proper distance) by a "hint" that may reveal a way to the result which is explained in the "solution."

We have mentioned repeatedly the "student" and the "teacher" and we shall refer to them again and again. It

may be good to observe that the "student" may be a high school student, or a college student, or anyone else who is studying mathematics. Also the "teacher" may be a high school teacher, or a college instructor, or anyone interested in the technique of teaching mathematics. The author looks at the situation sometimes from the point of view of the student and sometimes from that of the teacher (the latter case is preponderant in the first part). Yet most of the time (especially in the third part) the point of view is that of a person who is neither teacher nor student but anxious to solve the problem before him.

hesitation between shame and pretension. See WHY PROOFS?

If you cannot solve the proposed problem do not let this failure afflict you too much but try to find consolation with some easier success, *try to solve first some related problem;* then you may find courage to attack your original problem again. Do not forget that human superiority consists in going around an obstacle that cannot be overcome directly, in devising some suitable auxiliary problem when the original one appears insoluble.

Could you imagine a more accessible related problem? You should now *invent* a related problem, not merely *remember* one; I hope that you have tried already the question: *Do you know a related problem?*

The remaining questions in that paragraph of the list which starts with the title of the present article have a common aim, the VARIATION OF THE PROBLEM. There are different means to attain this aim as GENERALIZATION, SPECIALIZATION, ANALOGY, and others which are various ways of DECOMPOSING AND RECOMBINING.

Induction and mathematical induction. Induction is the process of discovering general laws by the observation and combination of particular instances. It is used in all sciences, even in mathematics. Mathematical induction is used in mathematics alone to prove theorems of a certain kind. It is rather unfortunate that the names are connected because there is very little logical connection between the two processes. There is, however, some practical connection; we often use both methods together. We are going to illustrate both methods by the same example.

1. We may observe, by chance, that

$$1 + 8 + 27 + 64 = 100$$

and, recognizing the cubes and the square, we may give to the fact we observed the more interesting form:

$$1^3 + 2^3 + 3^3 + 4^3 = 10^2.$$

How does such a thing happen? Does it often happen that such a sum of successive cubes is a square?

In asking this we are like the naturalist who, impressed by a curious plant or a curious geological formation, conceives a general question. Our general question is concerned with the sum of successive cubes

$$1^3 + 2^3 + 3^3 + \cdots + n^3.$$

We were led to it by the "particular instance" $n = 4$.

What can we do for our question? What the naturalist would do; we can investigate other special cases. The special cases $n = 2, 3$ are still simpler, the case $n = 5$ is the next one. Let us add, for the sake of uniformity and completeness, the case $n = 1$. Arranging neatly all these cases, as a geologist would arrange his specimens of a certain ore, we obtain the following table:

$$\begin{array}{lcrcl}
1 & = & 1 & = & 1^2 \\
1 + 8 & = & 9 & = & 3^2 \\
1 + 8 + 27 & = & 36 & = & 6^2 \\
1 + 8 + 27 + 64 & = & 100 & = & 10^2 \\
1 + 8 + 27 + 64 + 125 & = & 225 & = & 15^2.
\end{array}$$

It is hard to believe that all these sums of consecutive cubes are squares by mere chance. In a similar case, the naturalist would have little doubt that the general law suggested by the special cases heretofore observed is correct; the general law is almost proved by *induction.* The mathematician expresses himself with more reserve although fundamentally, of course, he thinks in the same fashion. He would say that the following theorem is strongly suggested by induction:

The sum of the first n *cubes is a square.*

2. We have been led to conjecture a remarkable, somewhat mysterious law. Why should those sums of successive cubes be squares? But, apparently, they are squares.

What would the naturalist do in such a situation? He would go on examining his conjecture. In so doing, he may follow various lines of investigation. The naturalist may accumulate further experimental evidence; if we wish to do the same, we have to test the next cases, $n = 6, 7, \ldots$. The naturalist may also reexamine the facts whose observation has led him to his conjecture; he compares them carefully, he tries to disentangle some deeper regularity, some further analogy. Let us follow this line of investigation.

Let us reexamine the cases $n = 1, 2, 3, 4, 5$ which we arranged in our table. Why are all these sums squares? What can we say about these squares? Their bases are 1, 3, 6, 10, 15. What about these bases? Is there some deeper regularity, some further analogy? At any rate, they do not seem to increase too irregularly. How do they increase? The difference between two successive terms of this sequence is itself increasing,

$$3 - 1 = 2, \quad 6 - 3 = 3, \quad 10 - 6 = 4, \quad 15 - 10 = 5.$$

Now these differences are conspicuously regular. We may see here a surprising analogy between the bases of those squares, we may see a remarkable regularity in the numbers 1, 3, 6, 10, 15:

$$\begin{aligned} 1 &= 1 \\ 3 &= 1 + 2 \\ 6 &= 1 + 2 + 3 \\ 10 &= 1 + 2 + 3 + 4 \\ 15 &= 1 + 2 + 3 + 4 + 5. \end{aligned}$$

If this regularity is general (and the contrary is hard to

believe) the theorem we suspected takes a more precise form:

It is, for $n = 1, 2, 3, \ldots$

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

3. The law we just stated was found by induction, and the manner in which it was found conveys to us an idea about induction which is necessarily one-sided and imperfect but not distorted. Induction tries to find regularity and coherence behind the observations. Its most conspicuous instruments are generalization, specialization, analogy. Tentative generalization starts from an effort to understand the observed facts; it is based on analogy, and tested by further special cases.

We refrain from further remarks on the subject of induction about which there is wide disagreement among philosophers. But it should be added that many mathematical results were found by induction first and proved later. Mathematics presented with rigor is a systematic deductive science but mathematics in the making is an experimental inductive science.

4. In mathematics as in the physical sciences we may use observation and induction to discover general laws. But there is a difference. In the physical sciences, there is no higher authority than observation and induction but in mathematics there is such an authority: rigorous proof.

After having worked a while experimentally it may be good to change our point of view. Let us be strict. We have discovered an interesting result but the reasoning that led to it was merely plausible, experimental, provisional, heuristic; let us try to establish it definitively by a rigorous proof.

We have arrived now at a "problem to prove": to prove or to disprove the result stated before (see 2, above).

There is a minor simplification. We may know that

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

At any rate, this is easy to verify. Take a rectangle with sides n and $n+1$, and divide it in two halves by a zigzag line as in Fig. 15a which shows the case $n = 4$. Each of the halves is "staircase-shaped" and its area has the expression $1 + 2 + \cdots + n$; for $n = 4$ it is $1 + 2 + 3 + 4$, see Fig. 18b. Now, the whole area of the rectangle is $n(n+1)$ of which the staircase-shaped area is one half; this proves the formula.

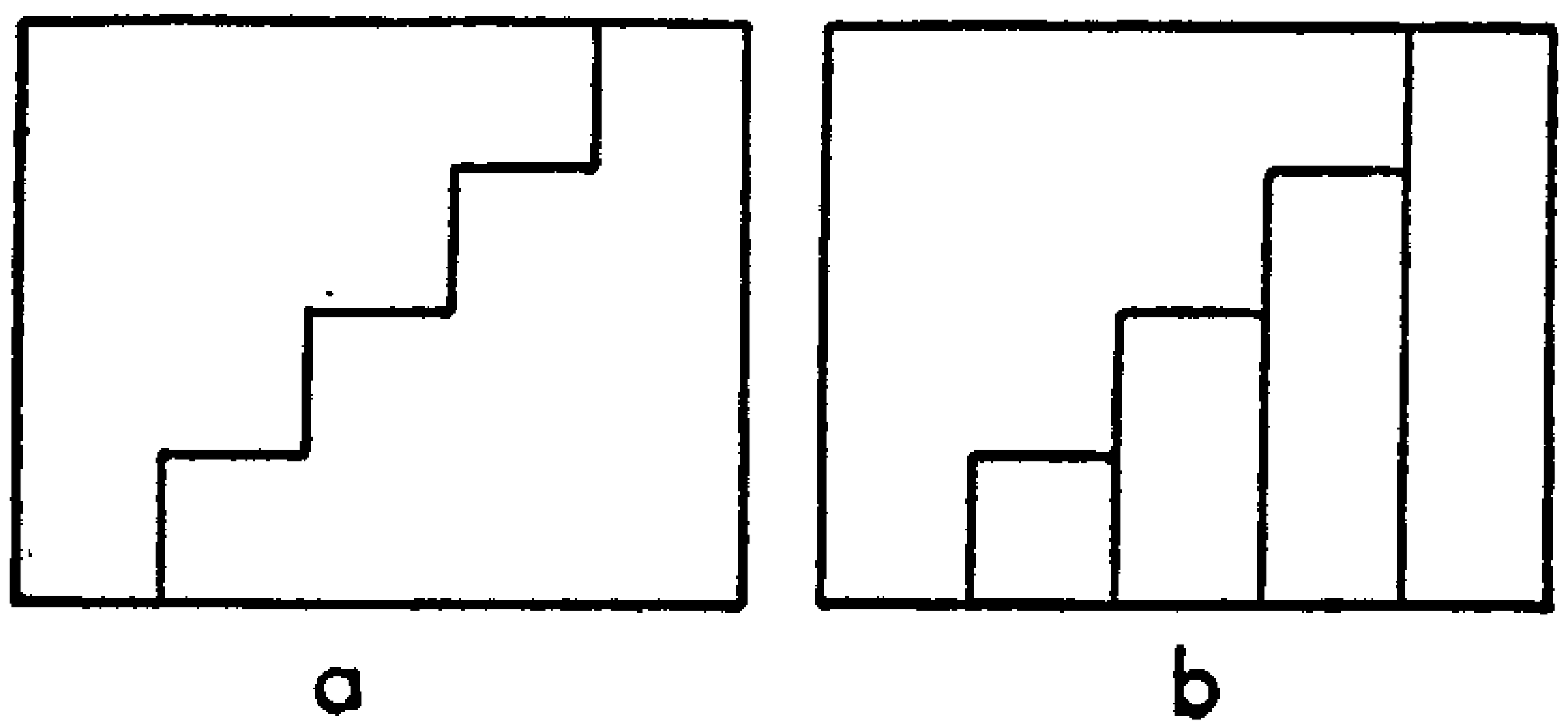

FIG. 18

We may transform the result which we found by induction into

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$$

5. If we have no idea how to prove this result, we may at least test it. Let us test the first case we have not tested yet, the case $n = 6$. For this value, the formula yields

$$1 + 8 + 27 + 64 + 125 + 216 = \left(\frac{6 \times 7}{2}\right)^2$$

and, on computation, this turns out to be true, both sides being equal to 441.

We can test the formula more effectively. The formula is, very likely, generally true, true for all values of n. Does it remain true when we pass from any value n to the next value $n+1$? Along with the formula as written above (p. 118) we should also have

$$1^3 + 2^3 + 3^3 + \cdots + n^3 + (n+1)^3 = \left(\frac{(n+1)(n+2)}{2}\right)^2.$$

Now, there is a simple check. Subtracting from this the formula written above, we obtain

$$(n+1)^3 = \left(\frac{(n+1)(n+2)}{2}\right)^2 - \left(\frac{n(n+1)}{2}\right)^2.$$

This is, however, easy to check. The right hand side may be written as

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

Our experimentally found formula passed a vital test.

Let us see clearly what this test means. We verified beyond doubt that

$$(n+1)^3 = \left(\frac{(n+1)(n+2)}{2}\right)^2 - \left(\frac{n(n+1)}{2}\right)^2.$$

We do not know yet whether

$$1^3 + 2^3 + 3^3 + \cdots + n^3 = \left(\frac{n(n+1)}{2}\right)^2$$

is true. But *if* we knew that this *was* true we could infer, by adding the equation which we verified beyond doubt, that

$$1^3 + 2^3 + 3^3 + \cdots + n^3 + (n+1)^3 = \left(\frac{(n+1)(n+2)}{2}\right)^2$$

is *also* true which is the same assertion for the next integer $n+1$. Now, we actually know that our conjecture is true for $n = 1, 2, 3, 4, 5, 6$. By virtue of what we have just said, the conjecture, being true for $n = 6$, must also be true for $n = 7$; being true for $n = 7$ it is true for $n = 8$; being true for $n = 8$ it is true for $n = 9$; and so on. It holds for all n, it is proved to be true generally.

6. The foregoing proof may serve as a pattern in many similar cases. What are the essential lines of this pattern?

The assertion we have to prove must be given in advance, in precise form.

The assertion must depend on an integer n.

The assertion must be sufficiently "explicit" so that we have some possibility of testing whether it remains true in the passage from n to the next integer $n+1$.

If we succeed in testing this effectively, we may be able to use our experience, gained in the process of testing, to conclude that the assertion must be true for $n+1$ provided it is true for n. When we are so far it is sufficient to know that the assertion is true for $n = 1$; hence it follows for $n = 2$; hence it follows for $n = 3$, and so on; passing from any integer to the next, we prove the assertion generally.

This process is so often used that it deserves a name. We could call it "proof from n to $n+1$" or still simpler "passage to the next integer." Unfortunately, the accepted technical term is "mathematical induction." This name results from a random circumstance. The precise assertion that we have to prove may come from any source, and it is immaterial from the logical viewpoint what the source is. Now, in many cases, as in the case we discussed here in detail, the source is induction, the assertion is found experimentally, and so the proof appears as a mathematical complement to induction; this explains the name.

7. Here is another point, somewhat subtle, but important to anybody who desires to find proofs by himself. In the foregoing, we found two different assertions by observation and induction, one after the other, the first under 1, the second under 2; the second was more precise than the first. Dealing with the second assertion, we found a possibility of checking the passage from n to $n+1$, and so we were able to find a proof by "mathematical induction." Dealing with the first assertion, and ignoring the precision added to it by the second one, we should scarcely have been able to find such a proof. In fact, the first assertion is less precise, less "explicit," less "tangible," less accessible to testing and checking than the second one. Passing from the first to the second, from the less precise to the more precise statement, was an important preparative for the final proof.

This circumstance has a paradoxical aspect. The second assertion is stronger; it implies immediately the first, whereas the somewhat "hazy" first assertion can hardly imply the more "clear-cut" second one. Thus, the stronger theorem is easier to master than the weaker one; this is the INVENTOR'S PARADOX.

Inventor's paradox. The more ambitious plan may have more chances of success.

This sounds paradoxical. Yet, when passing from one problem to another, we may often observe that the new, more ambitious problem is easier to handle than the original problem. More questions may be easier to answer than just one question. The more comprehensive theorem may be easier to prove, the more general problem may be easier to solve.

The paradox disappears if we look closer at a few examples (GENERALIZATION, 2; INDUCTION AND MATHEMATICAL INDUCTION, 7). The more ambitious plan may have

The questions and suggestions of our list cannot work magic. They cannot give us the solution of all possible puzzles without any effort on our part. If the reader wishes to find the word, he must keep on trying and thinking about it. What the questions and suggestions of the list can do is to "keep the ball rolling." When, discouraged by lack of success, we are inclined to drop the problem, they may suggest to us a new trial, a new aspect, a new variation of the problem, a new stimulus; they may keep us thinking.

For another example see DECOMPOSING AND RECOMBINING, 8.

Reductio ad absurdum and indirect proof are different but related procedures.

Reductio ad absurdum shows the falsity of an assumption by deriving from it a manifest absurdity. "Reduction to an absurdity" is a mathematical procedure but it has some resemblance to irony which is the favorite procedure of the satirist. Irony adopts, to all appearance, a certain opinion and stresses it and overstresses it till it leads to a manifest absurdity.

Indirect proof establishes the truth of an assertion by showing the falsity of the opposite assumption. Thus, indirect proof has some resemblance to a politician's trick of establishing a candidate by demolishing the reputation of his opponent.

Both "reductio ad absurdum" and indirect proof are effective tools of discovery which present themselves naturally to an intent mind. Nevertheless, they are disliked by a few philosophers and many beginners, which is understandable; satirical people and tricky politicians do not appeal to everybody. We shall first illustrate the effectiveness of both procedures by examples and discuss objections against them afterwards.

1. *Reductio ad absurdum.* Write numbers using each of the ten digits exactly once so that the sum of the numbers is exactly 100.

We may learn something by trying to solve this puzzle whose statement demands some elucidation.

What is the unknown? A set of numbers; and by numbers we mean here, of course, ordinary integers.

What is given? The number 100.

What is the condition? The condition has two parts. First, writing the desired set of numbers, we must use each of the ten digits, 0, 1, 2, 3, 4, 5, 6, 7, 8 and 9, just once. Second, the sum of all numbers in the set must be 100.

Keep only a part of the condition, drop the other part. The first part alone is easy to satisfy. Take the set 19, 28, 37, 46, 50; each figure occurs just once. But, of course, the second part of the condition is not satisfied; the sum of these numbers is 180, not 100. We could, however, do better. "Try, try again." Yes,

$$19 + 28 + 30 + 7 + 6 + 5 + 4 = 99.$$

The first part of the condition is satisfied, and the second part is almost satisfied; we have 99 instead of 100. Of course, we can easily satisfy the second part if we drop the first:

$$19 + 28 + 31 + 7 + 6 + 5 + 4 = 100.$$

The first part is not satisfied: the figure 1 occurs twice, and 0 not at all; the other figures are all right. "Try, try again."

After a few unsuccessful trials, however, we may be led to suspect that it is not possible to obtain 100 in the manner required. Eventually the problem arises: *Prove that it is impossible to satisfy both parts of the proposed condition at the same time.*

Quite good students may find that this problem is above their heads. Yet the answer is easy enough if we have the right attitude. *We have to examine the hypothetical situation in which both parts of the condition are satisfied.*

We suspect that this situation cannot actually arise and our suspicion, based on the experience of our unsuccessful trials, has some foundation. Nevertheless, let us keep an open mind and let us face the situation in which hypothetically, supposedly, allegedly both parts of the condition are satisfied. Thus, let us imagine a set of numbers whose sum is 100. They must be numbers with one or two figures. There are ten figures, and these ten figures must be all different, since each of the figures, 0, 1, 2, . . . 9 should occur just once. Thus, the sum of all ten figures must be

$$0+1+2+3+4+5+6+7+8+9=45.$$

Some of these figures denote units and others tens. It takes a little sagacity to hit upon the idea that the *sum of the figures denoting tens* may be of some importance. In fact, let t stand for this sum. Then the sum of the remaining figures, denoting units, is $45-t$. Therefore, the sum of all numbers in the set must be

$$10t+(45-t)=100.$$

We have here an equation to determine t. It is of the first degree and gives

$$t=\frac{55}{9}.$$

Now, there is something that is definitely wrong. The value of t that we have found is not an integer and t should be, of course, an integer. Starting from the supposition that both parts of the proposed condition can

be simultaneously satisfied, we have been led to a manifest absurdity. How can we explain this? Our original supposition must be wrong; both parts of the condition *cannot* be satisfied at the same time. And so we have attained our goal, we have succeeded in proving that the two parts of the proposed condition are incompatible.

Our reasoning is a typical "reductio ad absurdum."

2. *Remarks.* Let us look back at the foregoing reasoning and understand its general trend.

We wish to prove that it is impossible to fulfill a certain condition, that is, that the situation in which all parts of the condition are simultaneously satisfied can never arise. But, if we have proved nothing yet, we have to face the possibility that the situation could arise. Only by facing squarely the hypothetical situation and examining it closely can we hope to perceive some definitely wrong point in it. And we must lay our hand upon some definitely wrong point if we wish to show conclusively that the situation is impossible. Hence we can see that the procedure that was successful in our example is reasonable in general: We have to examine the hypothetical situation in which all parts of the condition are satisfied, *although such a situation appears extremely unlikely.*

The more experienced reader may see here another point. The main step of our procedure consisted in setting up an equation for t. Now, we could have arrived at the same equation without suspecting that something was wrong with the condition. If we wish to set up an equation, we have to express in mathematical language that all parts of the condition are satisfied, *although we do not know yet whether it is actually possible to satisfy all these parts simultaneously.*

Our procedure is "open-minded." We may hope to find the unknown satisfying the condition, or we may hope to show that the condition cannot be satisfied. It matters little in one respect: the investigation, if it is well conducted, starts in both cases in the same way, examining the hypothetical situation in which the condition is fulfilled, and shows only in its later course which hope is justified.

Compare FIGURES, 2. Compare also PAPPUS; an analysis which ends in disproving the proposed theorem, or in showing that the proposed "problem to find" has no solution, is actually a "reductio ad absurdum."

3. *Indirect proof.* The prime numbers, or primes, are the numbers 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, . . . which cannot be resolved into smaller factors, although they are greater than 1. (The last clause excludes the number 1 which, obviously, cannot be resolved into smaller factors, but has a different nature and should not be counted as a prime.) The primes are the "ultimate elements" into which all integers (greater than 1) can be decomposed. For instance,

$$630 = 2 \cdot 3 \cdot 3 \cdot 5 \cdot 7$$

is decomposed into a product of five primes.

Is the series of primes infinite or does it end somewhere? It is natural to suspect that the series of primes never ends. If it ended somewhere, all integers could be decomposed into a finite number of ultimate elements and the world would appear "too poor" in a manner of speaking. Thus arises the problem of proving the existence of an infinity of prime numbers.

This problem is very different from elementary mathematical problems of the usual kind and appears at first inaccessible. Yet, as we said, it is extremely unlikely that there should be a last prime, say P. Why is it so unlikely?

Let us face squarely the unlikely situation in which, hypothetically, supposedly, allegedly, there is a last prime P. Then we could write down the complete series of primes 2, 3, 5, 7, 11, . . . P. Why is this so unlikely? What is wrong with it? Can we point out anything that is definitely wrong? Indeed, we can. We can construct the number

$$Q = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \ldots P) + 1.$$

This number Q is greater than P and therefore, allegedly, Q cannot be a prime. Consequently, Q must be divisible by a prime. Now, all primes at our disposal are, supposedly, the numbers 2, 3, 5, . . . P but Q, divided by any of these numbers, leaves the rest 1; and so Q is not divisible by any of the primes mentioned which are, hypothetically, all the primes. Now, there is something that is definitely wrong; Q must be either a prime or it must be divisible by some prime. Starting from the supposition that there is a last prime P we have been led to a manifest absurdity. How can we explain this? Our original supposition must be wrong; there cannot be a last prime P. And so we have succeeded in proving that the series of prime numbers never ends.

Our proof is a typical indirect proof. (It is a famous proof too, due to Euclid; see Proposition 20 of Book IX of the Elements.)

We have established our theorem (that the series of primes never ends) by disproving its contradictory opposite (that the series of primes ends somewhere) which we have disproved by deducing from it a manifest absurdity. Thus we have combined indirect proof with "reductio ad absurdum"; this combination is also very typical.

4. *Objections.* The procedures which we are studying encountered considerable opposition. Many objections have been raised which are, possibly, only various forms of the same fundamental objection. We discuss here a "practical" form of the objection, which is on our level.

To find a not obvious proof is a considerable intellectual achievement but to learn such a proof, or even to understand it thoroughly costs also a certain amount of mental effort. Naturally enough, we wish to retain some benefit from our effort, and, of course, what we retain in our memory should be true and correct and not false or absurd.

But it seems difficult to retain something true from a "reductio ad absurdum." The procedure starts from a false assumption and derives from it consequences which are equally, but perhaps more visibly, false till it reaches a last consequence which is manifestly false. If we do not wish to store falsehoods in our memory we should forget everything as quickly as possible which is, however, not feasible because all points must be remembered sharply and correctly during our study of the proof.

The objection to indirect proofs can be now stated very briefly. Listening to such a proof, we are obliged to focus our attention all the time upon a false assumption which we should forget and not upon the true theorem which we should retain.

If we wish to judge correctly of the merits of these objections, we should distinguish between two uses of the "reductio ad absurdum," as a tool of research and as a means of exposition, and make the same distinction concerning the indirect proof.

It must be confessed that "reductio ad absurdum" as a means of exposition is not an unmixed blessing. Such a "reductio," especially if it is long, may become very painful indeed for the reader or listener. All the derivations which we examine in succession are correct but all the situations which we have to face are impossible. Even the verbal expression may become tedious if it insists, as it should, on emphasizing that everything is based on an initial assumption; the words "hypothetically," "supposedly," "allegedly" must recur incessantly, or some other device must be applied continually. We wish to reject and forget the situation as impossible but we have to retain and examine it as the basis for the next step, and this inner discord may become unbearable in the long run.

Yet it would be foolish to repudiate "reductio ad absurdum" as a tool of discovery. It may present itself naturally and bring a decision when all other means seem to be exhausted as the foregoing examples may show.

We need some experience to perceive that there is no essential opposition between our two contentions. Experience shows that usually there is little difficulty in converting an indirect proof into a direct proof, or in rearranging a proof found by a long "reductio ad absurdum" into a more pleasant form from which the "reductio ad absurdum" may even completely disappear (or, after due preparation, it may be compressed into a few striking sentences).

In short, if we wish to make full use of our capacities, we should be familiar both with "reductio ad absurdum" and with indirect proof. When, however, we have succeeded in deriving a result by either of these methods we should not fail to look back at the solution and ask: *Can you derive the result differently?*

Let us illustrate by examples what we have said.

5. *Rearranging a reductio ad absurdum.* We look back at the reasoning presented under 1. The reductio ad absurdum started from a situation which, eventually, turned out to be impossible. Let us however carve out a part of the argument which is independent of the initial false assumption and contains positive information. Reconsidering what we have done, we may perceive that this much is doubtless true: If a set of numbers with one

or two digits is written so that each of the ten figures occurs just once, then the sum of the set is of the form

$$10t + (45 - t) = 9\,(t + 5)\,.$$

Thus, this sum is divisible by 9. The proposed puzzle demands however that this sum should be 100. Is this possible? No, it is not, since 100 is not divisible by 9.

The "reductio ad absurdum" which led to the discovery of the argument vanished from our new presentation.

By the way, a reader acquainted with the procedure of "casting out nines" can see now the whole argument at a glance.

6. *Converting an indirect proof.* We look back at the reasoning presented under 3. Reconsidering carefully what we have done, we may find elements of the argument which are independent of any false assumption, yet the best clue comes from a reconsideration of the meaning of the original problem itself.

What do we mean by saying that the series of primes never ends? Evidently, just this: when we have ascertained any finite set of primes as 2, 3, 5, 7, 11, . . . P, where P is the last prime hitherto found, there is always one more prime. Thus, what must we do to prove the existence of an infinity of primes? We have to point out a way of finding a prime different from all primes hitherto found. Thus, our "problem to prove" is in fact reduced to a "problem to find": *Being given the primes* 2, 3, 5, . . . P, *find a new prime* N *different from all the given primes.*

Having restated our original problem in this new form, we have taken the main step. It is relatively easy now to see how to use the essential parts of our former argument for the new purpose. In fact, the number

$$Q = (2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \ldots P) + 1$$

is certainly divisible by a prime. Let us take—this is the idea—any prime divisor of Q (for instance, the smallest one) for N. (Of course, if Q happens to be a prime, then $N = Q$.) Obviously, Q divided by any of the primes 2, 3, 5, . . . P leaves the remainder 1 and, therefore, none of these numbers can be N which is a divisor of Q. But that is all we need: N is a prime, and different from all hitherto found primes 2, 3, 5, 7, 11, . . . P.

This proof gives a definite procedure of prolonging again and again the series of primes, without limit. Nothing is indirect in it, no impossible situation needs to be considered. Yet, fundamentally, it is the same as our former indirect proof which we have succeeded in converting.

Redundant. See CONDITION.

Routine problem may be called the problem to solve the equation $x^2 - 3x + 2 = 0$ if the solution of the general quadratic equation was explained and illustrated before so that the student has nothing to do but to substitute the numbers -3 and 2 for certain letters which appear in the general solution. Even if the quadratic equation was not solved generally in "letters" but half a dozen similar quadratic equations with numerical coefficients were solved just before, the problem should be called a "routine problem." In general, a problem is a "routine problem" if it can be solved either by substituting special data into a formerly solved general problem, or by following step by step, without any trace of originality, some well-worn conspicuous example. Setting a routine problem, the teacher thrusts under the nose of the student an immediate and decisive answer to the question: *Do you know a related problem?* Thus, the student needs nothing but a little care and patience in following a cut-and-dried precept, and he has no opportunity to use his judgment or his inventive faculties.

missing idea. If even the hint does not help, he may look at the solution, try to isolate the key idea, put the book aside, and then try to work out the solution.

PROBLEMS

1. A bear, starting from the point P, walked one mile due south. Then he changed direction and walked one mile due east. Then he turned again to the left and walked one mile due north, and arrived exactly at the point P he started from. What was the color of the bear?

2. Bob wants a piece of land, exactly level, which has four boundary lines. Two boundary lines run exactly north-south, the two others exactly east-west, and each boundary line measures exactly 100 feet. Can Bob buy such a piece of land in the U.S.?

3. Bob has 10 pockets and 44 silver dollars. He wants to put his dollars into his pockets so distributed that each pocket contains a different number of dollars. Can he do so?

4. To number the pages of a bulky volume, the printer used 2989 digits. How many pages has the volume?

5. Among Grandfather's papers a bill was found:

72 turkeys $_67.9_

The first and last digit of the number that obviously represented the total price of those fowls are replaced here by blanks, for they have faded and are now illegible.

What are the two faded digits and what was the price of one turkey?

6. Given a regular hexagon and a point in its plane. Draw a straight line through the given point that divides the given hexagon into two parts of equal area.

7. Given a square. Find the locus of the points from which the square is seen under an angle (a) of 90° (b) of 45°. (Let P be a point outside the square, but in the same plane. The smallest angle with vertex P containing the square is the "angle under which the square is seen" from P.) Sketch clearly both loci and give a full description.

8. Call "axis" of a solid a straight line joining two points of the surface of the solid and such that the solid, rotated about this line through an angle which is greater than 0° and less than 360° coincides with itself.

Find the axes of a cube. Describe clearly the location of the axes, find the angle of rotation associated with each. Assuming that the edge of the cube is of unit length, compute the arithmetic mean of the lengths of the axes.

9. In a tetrahedron (which is not necessarily regular) two opposite edges have the same length a and they are perpendicular to each other. Moreover they are each perpendicular to a line of length b which joins their midpoints. Express the volume of the tetrahedron in terms of a and b, and prove your answer.

10. The vertex of a pyramid opposite the base is called the *apex*. (a) Let us call a pyramid "isosceles" if its apex is at the same distance from all *vertices* of the base. Adopting this definition, prove that the base of an isosceles pyramid is *inscribed* in a circle the center of which is the foot of the pyramid's altitude.

(b) Now let us call a pyramid "isosceles" if its apex is at the same (perpendicular) distance from all sides of the base. Adopting this definition (different from the foregoing) prove that the base of an isosceles pyramid is *circumscribed* about a circle the center of which is the foot of the pyramid's altitude.

11. Find x, y, u, and v, satisfying the system of four equations

$$\begin{aligned} x + 7y + 3v + 5u &= 16 \\ 8x + 4y + 6v + 2u &= -16 \\ 2x + 6y + 4v + 8u &= 16 \\ 5x + 3y + 7v + u &= -16 \end{aligned}$$

(This may look long and boring: look for a short cut.)

12. Bob, Peter, and Paul travel together. Peter and Paul are good hikers; each walk p miles per hour. Bob has a bad foot and drives a small car in which two people can ride, but not three; the car covers c miles per hour. The three friends adopted the following scheme: They start together, Paul rides in the car with Bob, Peter walks. After a while, Bob drops Paul, who walks on; Bob returns to pick up Peter, and then Bob and Peter ride in the car till they overtake Paul. At this point they change: Paul rides and Peter walks just as they started and the whole procedure is repeated as often as necessary.

(a) How much progress (how many miles) does the company make per hour?

(b) Through which fraction of the travel time does the car carry just one man?

(c) Check the extreme cases $p = 0$ and $p = c$.

13. Three numbers are in arithmetic progression, three other numbers in geometric progression. Adding the corresponding terms of these two progressions successively, we obtain

$$85, \quad 76, \quad \text{and} \quad 84$$

respectively, and, adding all three terms of the arithmetic progression, we obtain 126. Find the terms of both progressions.

14. Determine m so that the equation in x

$$x^4 - (3m + 2)x^2 + m^2 = 0$$

has four real roots in arithmetic progression.

15. The length of the perimeter of a right triangle is 60 inches and the length of the altitude perpendicular to the hypotenuse is 12 inches. Find the sides.

16. From the peak of a mountain you see two points, A and B, in the plain. The lines of vision, directed to these points, include the angle γ. The inclination of the first line of vision to a horizontal plane is α, that of the second line β. It is known that the points A and B are on the same level and that the distance between them is c.

Express the elevation x of the peak above the common level of A and B in terms of the angles α, β, γ, and the distance c.

17. Observe that the value of

$$\frac{1}{2!} + \frac{2}{3!} + \frac{3}{4!} + \cdots + \frac{n}{(n+1)!}$$

is 1/2, 5/6, 23/24 for $n = 1,2,3$, respectively, guess the general law (by observing more values if necessary) and prove your guess.

18. Consider the table

$$\begin{aligned} 1 &= 1 \\ 3 + 5 &= 8 \\ 7 + 9 + 11 &= 27 \\ 13 + 15 + 17 + 19 &= 64 \\ 21 + 23 + 25 + 27 + 29 &= 125 \end{aligned}$$

Guess the general law suggested by these examples, express it in suitable mathematical notation, and prove it.

19. The side of a regular hexagon is of length n (n is an integer). By equidistant parallels to its sides the hexagon is divided into T equilateral triangles each of which has sides of length 1. Let V denote the number of vertices appearing in this division, and L the number of boundary lines of length 1. (A boundary line belongs to one or two triangles, a vertex to two or more triangles.) When

$n = 1$, which is the simplest case, $T = 6$, $V = 7$, $L = 12$. Consider the general case and express T, V, and L in terms of n. (Guessing is good, proving is better.)

20. In how many ways can you change one dollar? (The "way of changing" is determined if it is known how many coins of each kind—cents, nickels, dimes, quarters, half dollars—are used.)

HINTS

1. *What is the unknown?* The color of a bear—but how could we find the color of a bear from mathematical data? *What is given?* A geometrical situation—but it seems self-contradictory: how could the bear, after walking three miles in the manner described, return to his *starting* point?

2. *Do you know a related problem?*

3. If Bob had very many dollars, he would have obviously no difficulty in filling each of his pockets differently. *Could you restate the problem?* What is the minimum number of dollars that can be put in 10 pockets so that no two different pockets contain the same amount?

4. *Here is a problem related to yours:* If the book has exactly 9 numbered pages, how many digits uses the printer? (9, of course.) Here is another *problem related to yours:* If the book has exactly 99 numbered pages, how many digits does the printer use?

5. *Could you restate the problem?* What can the two faded digits be if the total price, expressed in cents, is divisible by 72?

6. *Could you imagine a more accessible related problem? A more general problem? An analogous problem?* (GENERALIZATION, 2.)

7. *Do you know a related problem?* The locus of the points from which a given segment of a straight line is

seen under a given angle consists of two circular arcs, ending in the extreme points of the segment, and symmetric to each other with respect to the segment.

8. I assume that the reader is familiar with the shape of the cube and has found certain axes just by inspection—but are they *all* the axes? *Can you prove* that your list of axes is exhaustive? Has your list a clear principle of classification?

9. *Look at the unknown!* The unknown is the volume of a tetrahedron—yes, I know, the volume of any pyramid can be computed when the base and the height are given (product of both, divided by 3) but in the present case neither the base nor the height is given. *Could you imagine a more accessible related problem?* (Don't you see a more accessible tetrahedron which is an aliquot part of the given one?)

10. *Do you know a related theorem?* Do you know a related . . . simpler . . . *analogous theorem?* Yes: the foot of the altitude is the mid-point of the base in an isosceles triangle. Here is a *theorem related to yours and proved before. Could you use . . . its method?* The theorem on the isosceles triangle is proved from congruent right triangles of which the altitude is a common side.

11. It is assumed that the reader is somewhat familiar with systems of linear equations. To solve such a system, we have to combine its equations in some way—look out for relations between the equations which could indicate a particularly advantageous combination.

12. *Separate the various parts of the condition. Can you write them down?* Between the start and the point where the three friends meet again there are three different phases:

(1) Bob rides with Paul
(2) Bob rides alone
(3) Bob rides with Peter.

Call t_1, t_2, and t_3 the durations of these phases, respectively. How could you split the condition into appropriate parts?

13. *Separate the various parts of the condition. Can you write them down?* Let

$$a-d, \qquad a, \qquad a+d$$

be the terms of the arithmetic progression, and

$$bg^{-1}, \qquad b, \qquad bg$$

be the terms of the geometric progression.

14. *What is the condition?* The four roots must form an arithmetic progression. Yet the equation has a particular feature: it contains only even powers of the unknown x. Therefore, if a is a root, $-a$ is also a root.

15. *Separate the various parts of the condition. Can you write them down?* We may distinguish three parts in the condition, concerning

(1) perimeter
(2) right triangle
(3) height to hypotenuse.

16. *Separate the various parts of the condition. Can you write them down?* Let a and b stand for the lengths of the (unknown) lines of vision, α and β for their inclinations to the horizontal plane, respectively. We may distinguish three parts in the condition, concerning

(1) the inclination of a
(2) the inclination of b
(3) the triangle with sides a, b, and c.

17. Do you *recognize* the denominators 2, 6, 24? *Do you know a related problem? An analogous problem?* (INDUCTION AND MATHEMATICAL INDUCTION.)

18. Discovery by induction needs observation. Observe the right-hand sides, the initial terms of the left-hand sides, and the final terms. What is the general law?

19. *Draw a figure.* Its observation may help you to discover the law inductively, or it may lead you to relations between T, V, L, and n.

20. *What is the unknown?* What are we supposed to seek? Even the aim of the problem may need some clarification. *Could you imagine a more accessible related problem? A more general problem? An analogous problem?* Here is a *very* simple analogous problem: In how many ways can you pay one cent? (There is just one way.) Here is a more general problem: In how many ways can you pay the amount of n cents using these five kinds of coins: cents, nickels, dimes, quarters, and half dollars. We are especially concerned with the particular case $n = 100$.

In the simplest particular cases, for small n, we can figure out the answer without any high-brow method, just by trying, by inspection. Here is a short table (which the reader should check).

n	4	5	9	10	14	15	19	20	24	25
E_n	1	2	2	4	4	6	6	9	9	13

The first line lists the amounts to be paid, generally called n. The second line lists the corresponding numbers of "ways of paying," generally called E_n. (Why I have chosen this notation is a secret of mine which I am not willing to give away at this stage.)

We are especially concerned with E_{100}, but there is little hope that we can compute E_{100} without some clear method. In fact the present problem requires a little more from the reader than the foregoing ones; he should *create* a little *theory*.

Our question is general (to compute E_n for general n), but it is "isolated." *Could you imagine a more accessible related problem? An analogous problem?* Here is a *very* simple analogous problem: Find A_n, the number of ways to pay the amount of n cents, using only cents. ($A_n = 1$.)

SOLUTIONS

1. You think that the bear was white and the point P is the North Pole? *Can you prove that this is correct?* As it was more or less understood, we idealize the question. We regard the globe as exactly spherical and the bear as a moving material point. This point, moving due south or due north, describes an arc of a *meridian* and it describes an arc of a *parallel* circle (parallel to the equator) when it moves due east. We have to distinguish two cases.

(1) If the bear returns to the point P along a meridian *different* from the one along which he left P, P is necessarily the North Pole. In fact the only other point of the globe in which two meridians meet is the South Pole, but the bear could leave this pole only in moving northward.

(2) The bear could return to the point P along the same meridian he left P if, when walking one mile due east, he describes a parallel circle exactly n times, where n may be 1, 2, 3 . . . In this case P is not the North Pole, but a point on a parallel circle very close to the South Pole (the perimeter of which, expressed in miles, is slightly inferior to $2\pi + 1/n$).

2. We represent the globe as in the solution of Problem 1. The land that Bob wants is bounded by two meridians and two parallel circles. Imagine two fixed meridians, and a parallel circle moving *away* from the equator: the arc on the moving parallel intercepted by the two fixed meridians is steadily shortened. The center of the land that Bob wants should be on the equator: he can *not* get it in the U.S.

3. The least possible number of dollars in a pocket is obviously 0. The next greater number is at least 1, the next greater at least 2 . . . and the number in the last (tenth) pocket is at least 9. Therefore, the number of dollars required is at least

$$0 + 1 + 2 + 3 + \ldots + 9 = 45$$

Bob cannot make it: he has only 44 dollars.

4. A volume of 999 pages needs

$$9 + 2 \times 90 + 3 \times 900 = 2889$$

digits. If the bulky volume in question has x pages

$$2889 + 4(x - 999) = 2989$$
$$x = 1024$$

This problem may teach us that a preliminary estimate of the unknown may be useful (or even necessary, as in the present case).

5. If _679_ is divisible by 72, it is divisible both by 8 and by 9. If it is divisible by 8, the number 79_ must be divisible by 8 (since 1000 is divisible by 8) and so 79_ must be 792: the last faded digit is 2. If _6792 is divisible by 9, the sum of its digits must be divisible by 9 (the rule about "casting out nines") and so the first faded digit must be 3. The price of one turkey was (in grandfather's time) $367.92 ÷ 72 = $5.11.

6. "*A point and a figure with a center of symmetry* (in the same plane) are given in position. Find a straight line that passes through the given point and bisects the area of the given figure." The required line passes, of course, through the center of symmetry. See INVENTOR'S PARADOX.

7. In any position the two sides of the angle must pass through two vertices of the square. As long as they pass through the same pair of vertices, the angle's vertex

moves along the same arc of circle (by the theorem underlying the hint). Hence each of the two loci required consists of several arcs of circle: of 4 semicircles in the case (a) and of 8 quarter circles in the case (b); see Fig. 31.

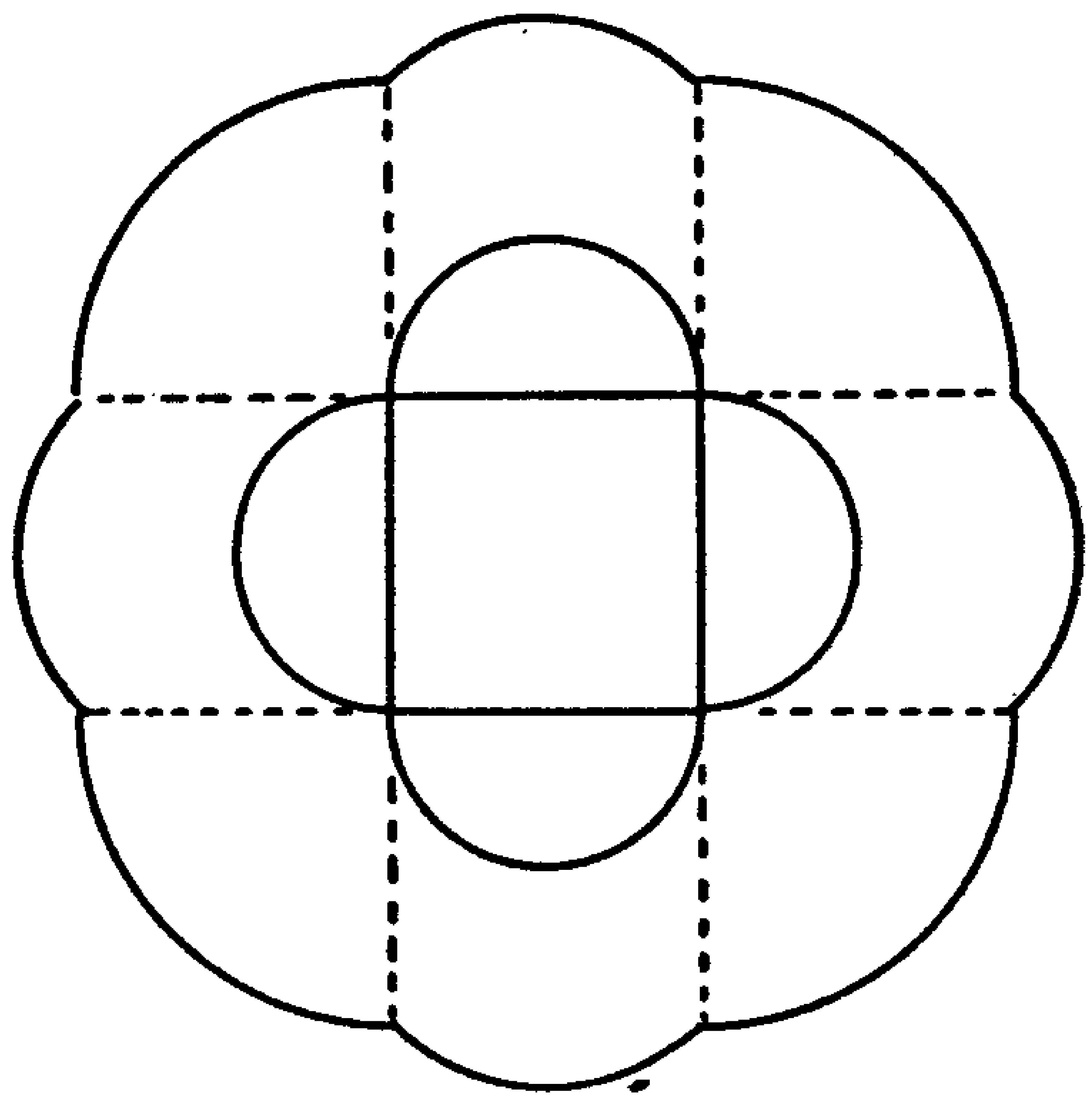

FIG. 31

8. The axis pierces the surface of the cube in some point which is either a vertex of the cube or lies on an edge or in the interior of a face. If the axis passes through a point of an edge (but not through one of its end-points) this point must be the midpoint: otherwise the edge could not coincide with itself after the rotation. Similarly, an axis piercing the interior of a face must pass through its center. Any axis must, of course, pass through the center of the cube. And so there are three kinds of axes:

(1) 4 axes, each through two opposite vertices; angles 120°, 240°

(2) 6 axes, each through the mid-points of two opposite edges; angle 180°

(3) 3 axes, each through the center of two opposite faces; angles 90°, 180°, 270°.

For the length of an axis of the first kind see section 12; the others are still easier to compute. The desired average is

$$\frac{4\sqrt{3} + 6\sqrt{2} + 3}{13} = 1.416.$$

(This problem may be useful in preparing the reader for the study of crystallography. For the reader sufficiently advanced in the integral calculus it may be observed that the average computed is a fairly good approximation to the "average width" of the cube, which is, in fact, $3/2 = 1.5$.)

9. The plane passing through one edge of length a and the perpendicular of length b divides the tetrahedron into two *more accessible* congruent tetrahedra, each with base $ab/2$ and height $a/2$. Hence the required volume

$$= 2 \cdot \frac{1}{3} \cdot \frac{ab}{2} \cdot \frac{a}{2} = \frac{a^2 b}{6}.$$

10. The base of the pyramid is a polygon with n sides. In the case (a) the n lateral edges of the pyramid are equal; in the case (b) the altitudes (drawn from the apex) of its n lateral faces are equal. If we draw the altitude of the pyramid and join its foot to the n vertices of the base in the case (a), but to the feet of the altitudes of the n lateral faces in the case (b), we obtain, in both cases, *n right triangles of which the altitude* (of the pyramid) *is a common side:* I say that these n right triangles are congruent. In fact the hypotenuse [a lateral edge in the case (a), a lateral altitude in the case (b)] is of the same length in each, according to the definitions

laid down in the proposed problem; we have just mentioned that another side (the altitude of the pyramid) and an angle (the right angle) are common to all. In the n congruent triangles the third sides must also be equal; they are drawn from the same point (the foot of the altitude) in the same plane (the base): they form n radii of a circle which is circumscribed about, or inscribed into, the base of the pyramid, in the cases (a) and (b), respectively. [In the case (b) it remains to show, however, that the n radii mentioned are perpendicular to the respective sides of the base; this follows from a well-known theorem of solid geometry on projections.]

It is most remarkable that a plane figure, the isosceles triangle, may have *two different analogues* in solid geometry.

11. Observe that the first equation is so related to the last as the second is to the third: the coefficients on the left-hand sides are the same, but in opposite order, whereas the right-hand sides are opposite. Add the first equation to the last and the second to the third:

$$6(x + u) + 10(y + v) = 0,$$
$$10(x + u) + 10(y + v) = 0.$$

This can be regarded as a system of two linear equations for two unknowns, namely for $x + u$ and $y + v$, and easily yields

$$x + u = 0, \qquad y + v = 0.$$

Substituting $-x$ for u and $-y$ for v in the first two equations of the original system, we find

$$-4x + 4y = 16$$
$$6x - 2y = -16.$$

This is a simple system which yields

$$x = -2, \qquad y = 2, \qquad u = 2, \qquad v = -2$$

12. Between the start and the meeting point each of the friends traveled the same distance. (Remember, distance = velocity × time.) We distinguish two parts in the condition:

Bob traveled as much as Paul:

$$ct_1 - ct_2 + ct_3 = ct_1 + pt_2 + pt_3,$$

Paul traveled as much as Peter:

$$ct_1 + pt_2 + pt_3 = pt_1 + pt_2 + ct_3.$$

The second equation yields

$$(c - p)t_1 = (c - p)t_3.$$

We assume, of course, that the car travels faster than a pedestrian, $c > p$. It follows

$$t_1 = t_3;$$

that is, Peter walks just as much as Paul. From the first equation, we find that

$$\frac{t_3}{t_2} = \frac{c + p}{c - p}$$

which is, of course, also the value for t_1/t_2. Hence we obtain the answers:

(a) $$\frac{c(t_1 - t_2 + t_3)}{t_1 + t_2 + t_3} = \frac{c(c + 3p)}{3c + p}$$

(b) $$\frac{t_2}{t_1 + t_2 + t_3} = \frac{c - p}{3c + p}$$

(c) In fact, $0 < p < c$. There are two extreme cases:

If $p = 0$ (a) yields $c/3$ and (b) yields $1/3$

If $p = c$ (a) yields c and (b) yields 0.

These results are easy to see without computation.

13. The condition is easily split into four parts expressed by the four equations

$$\begin{aligned} a - d + bg^{-1} &= 85 \\ a + b &= 76 \\ a + d + bg &= 84 \\ 3a &= 126. \end{aligned}$$

The last equation yields $a = 42$, then the second $b = 34$. Adding the remaining two equations (to eliminate d), we obtain

$$2a + b(g^{-1} + g) = 169.$$

Since a and b are already known, we have here a quadratic equation for g. It yields

$$g = 2, \quad d = -26 \quad \text{or} \quad g = 1/2, \quad d = 25.$$

The progressions are

68, 42, 16		17, 42, 67
	or	
17, 34, 68		68, 34, 17

14. If a and $-a$ are the roots having the least absolute value, they will stand next to each other in the progression which will, therefore, be of the form

$$-3a, \ -a, \ a, \ 3a.$$

Hence the left-hand side of the proposed equation must have the form

$$(x^2 - a^2)(x^2 - 9a^2).$$

Carrying out the multiplication and comparing coefficients of like powers, we obtain the system

$$\begin{aligned} 10a^2 &= 3m + 2, \\ 9a^4 &= m^2. \end{aligned}$$

Elimination of a yields

$$19m^2 - 108m - 36 = 0.$$

Hence $m = 6$ or $-6/19$.

15. Let a, b, and c denote the sides, the last being the hypotenuse. The three parts of the condition are expressed by

$$\begin{aligned} a + b + c &= 60 \\ a^2 + b^2 &= c^2 \\ ab &= 12c. \end{aligned}$$

Observing that

$$(a + b)^2 = a^2 + b^2 + 2ab$$

we obtain

$$(60 - c)^2 = c^2 + 24c.$$

Hence $c = 25$ and either $a = 15$, $b = 20$ or $a = 20$, $b = 15$ (no difference for the triangle).

16. The three parts of the condition are expressed by

$$\sin \alpha = \frac{x}{a},$$

$$\sin \beta = \frac{x}{b},$$

$$c^2 = a^2 + b^2 - 2ab \cos \gamma$$

The elimination of a and b yields

$$x^2 = \frac{c^2 \sin^2 \alpha \sin^2 \beta}{\sin^2 \alpha + \sin^2 \beta - 2 \sin \alpha \sin \beta \cos \gamma}.$$

17. We conjecture that

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

Following the pattern of INDUCTION AND MATHEMATICAL INDUCTION, we ask: Does the conjectured formula remain true when we pass from the value n to the next value $n = 1$? Along with the formula above we should have

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

Check this by subtracting from it the former:

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

which boils down to

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

and this last equation is obviously true for $n = 1, 2, 3, \ldots$ hence, by following the pattern referred to above, we can prove our conjecture.

18. In the nth line the right-hand side seems to be n^3 and the left-hand side a sum of n terms. The final term of this sum is the mth odd number, or $2m - 1$, where

$$m = 1 + 2 + 3 + \cdots + n = \frac{n(n+1)}{2};$$

see INDUCTION AND MATHEMATICAL INDUCTION, 4. Hence the final term of the sum on the left-hand side should be

$$2m - 1 = n^2 + n - 1.$$

We can derive hence the initial term of the sum considered in *two* ways: going back $n - 1$ steps from the final term, we find

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

whereas, advancing one step from the final term of the foregoing line, we find

$$[(n-1)^2 + (n-1) - 1] + 2$$

which, after routine simplification, boils down to the same: good! We assert therefore that

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

where the left-hand side indicates the sum of n successive terms of an arithmetic progression the difference of which is 2. If the reader knows the rule for the sum of such a progression (arithmetic mean of the initial term and the final term, multiplied by the number of terms), he can verify that

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

and so prove the assertion.

(The rule quoted can be easily proved by a picture little different from Fig. 18.)

19. The length of the perimeter of the regular hexagon with side n is $6n$. Therefore, this perimeter consists of $6n$ boundary lines of length 1 and contains $6n$ vertices. Therefore, in the transition from $n - 1$ to n, V increases by $6n$ units, and so

$$V = 1 + 6(1 + 2 + 3 + \cdots + n) = 3n^2 + 3n + 1;$$

see INDUCTION AND MATHEMATICAL INDUCTION, 4. By 3 diagonals through its center the hexagon is divided into 6 (large) equilateral triangles. By inspection of one of these

$$T = 6(1 + 3 + 5 + \cdots + 2n - 1) = 6n^2$$

(rule for the sum of an arithmetic progression, quoted in the solution of Problem 18). The T triangles have jointly $3T$ sides. In this total $3T$ each internal line of division of length 1 is counted twice, whereas the $6n$ lines along the perimeter of the hexagon are counted but once. Hence

$$2L = 3T + 6n, \qquad L = 9n^2 + 3n.$$

(For the more advanced reader: it follows from Euler's theorem on polyhedra that $T + V = L + 1$. Verify this relation!)

20. Here is a well-ordered array of analogous problems: Compute A_n, B_n, C_n, D_n and E_n. Each of these quantities represents the number of ways to pay the amount of n cents; the difference is in the coins used:

A_n only cents
B_n cents and nickels
C_n cents, nickels, and dimes
D_n cents, nickels, dimes, and quarters
E_n cents, nickels, dimes, quarters, and half dollars.

The symbols E_n (reason now clear) and A_n were used before.

All ways and manners to pay the amount of n cents with the five kinds of coin are enumerated by E_n. We may, however, distinguish two possibilities:

First. No half dollar is used. The number of such ways to pay is D_n, by definition.

Second. A half dollar (possibly more) is used. After the first half dollar is laid on the counter, there remains the amount of $n - 50$ cents to pay, which can be done in exactly E_{n-50} ways.

We infer that

$$E_n = D_n + E_{n-50}.$$

Similarly

$$D_n = C_n + D_{n-25},$$
$$C_n = B_n + C_{n-10},$$
$$B_n = A_n + B_{n-5}.$$

A little attention shows that these formulas remain valid if we set

$$A_0 = B_0 = C_0 = D_0 = E_0 = 1$$

(which obviously makes sense) and regard any one of the quantities $A_n, B_n \ldots E_n$ as equal to 0 when its subscript happens to be negative. (For example, $E_{25} = D_{25}$, as can be seen immediately, and this agrees with our first formula since $E_{25-50} = E_{-25} = 0$.)

Our formulas allow us to compute the quantities considered *recursively*, that is, by going back to lower values of n *or* to former letters of the alphabet. For example, we can compute C_{30} by simple addition if C_{20} and B_{30} are already known. In the table below the initial row, headed by A_n, and the initial column, headed by 0, contain only numbers equal to 1. (Why?) Starting from these initial numbers, we compute the others recursively, by simple additions: any other number of the table is equal either to the number above it or to the sum of two numbers: the number above it and another at the proper distance to the left. For example,

$$C_{30} = B_{30} + C_{20} = 7 + 9 = 16$$

The computation is carried through till $E_{50} = 50$: *you can pay 50 cents in exactly 50 different ways.* Carrying it further, the reader can convince himself that $E_{100} = 292$: *you can change a dollar in 292 different ways.*

n	0	5	10	15	20	25	30	35	40	45	50
A_n	1	1	1	1	1	1	1	1	1	1	1
B_n	1	2	3	4	5	6	7	8	9	10	11
C_n	1	2	4	6	9	12	16	20	25	30	36
D_n	1	2	4	6	9	13	18	24	31	39	49
E_n	1	2	4	6	9	13	18	24	31	39	50