GEORGE POLYA

MATHEMATICAL DISCOVERY

Volume 1

MATHEMATICAL DISCOVERY

On understanding, learning, and teaching problem solving

VOLUME I

GEORGE POLYA

PROFESSOR EMERITUS OF MATHEMATICS, STANFORD UNIVERSITY

PREFACE

A method of solution is perfect if we can foresee from the start, and even prove, that following that method we shall attain our aim.
LEIBNITZ: *Opuscules*, p. 161.

1. Solving a problem means finding a way out of a difficulty, a way around an obstacle, attaining an aim which was not immediately attainable. Solving problems is the specific achievement of intelligence, and intelligence is the specific gift of mankind: solving problems can be regarded as the most characteristically human activity. The aim of this work is to understand this activity, to propose means to teach it, and, eventually, to improve the problem-solving ability of the reader.

2. This work consists of two parts; let me characterize briefly the role of these two parts.

Solving problems is a practical art, like swimming, or skiing, or playing the piano: you can learn it only by imitation and practice. This book cannot offer you a magic key that opens all the doors and solves all the problems, but it offers you good examples for imitation and many opportunities for practice: if you wish to learn swimming you have to go into the water, and if you wish to become a problem solver you have to solve problems.

If you wish to derive the most profit from your effort, look out for such features of the problem at hand as may be useful in handling the problems to come. A solution that you have obtained by your own effort or one that you have read or heard, but have followed with real interest and insight, may become a *pattern* for you, a model that you can imitate with advantage in solving similar problems. The aim of Part One is to familiarize you with a few useful patterns.

It may be easy to imitate the solution of a problem when solving a closely similar problem; such imitation may be more difficult or scarcely possible if the similarity is not so close. Yet there is a deep-seated human desire for more: for some device, free of limitations, that could solve all problems. This desire may remain obscure in many of us, but it becomes manifest in a few fairy tales and in the writings of a few philosophers. You may remember the tale about the magic word that opens all the doors. Descartes meditated upon a universal method for solving all problems, and Leibnitz very clearly formulated the idea of a perfect method. Yet the quest for a universal perfect method has no more succeeded than did the quest for the philosopher's stone which was supposed to change base metals into gold; there are great dreams that must remain dreams. Nevertheless, such unattainable ideals may influence people: nobody has attained the North Star, but many have found the right way by looking at it. This book cannot offer you (and no book will ever be able to offer you) a universal perfect method for solving problems, but even a few small steps toward that unattainable ideal may clarify your mind and improve your problem-solving ability. Part Two outlines some such steps.

3. I wish to call *heuristic* the study that the present work attempts, the study of means and methods of problem solving. The term heuristic, which was used by some philosophers in the past, is half-forgotten and half-discredited nowadays, but I am not afraid to use it.

In fact, most of the time the present work offers a down-to-earth practical aspect of heuristic: I am trying, by all the means at my disposal, to entice the reader to do problems and to think about the means and methods he uses in doing them.

In most of the following chapters, the greater part of the text is devoted to the broad presentation of the solution of a few problems. The presentation may appear too broad to a mathematician who is not interested in methodical points. In fact, what is presented here are not merely solutions but *case histories* of solutions. Such a case history describes the sequence of essential steps by which the solution has been eventually discovered, and tries to disclose the motives and attitudes prompting these steps. The aim of such a careful description of a particular case is to suggest some general advice, or pattern, which may guide the reader in similar situations. The explicit formulation of such advice or such a pattern is usually reserved for a separate section, although tentative first formulations may be interspersed between the incidents of the case history.

Each chapter is followed by examples and comments. The reader who does the examples has an opportunity to apply, clarify, and amplify the methodical remarks offered in the text of the chapter. The comments interspersed between the examples give extensions, more technical or more subtle points, or incidental remarks.

How far I have succeeded I cannot know, but I have certainly tried hard to enlist the reader's participation. I have tried to fix on the printed page whatever modes of oral presentation I found most effective in my classes.

By the case histories, I have tried to familiarize the reader with the atmosphere of research. By the choice, formulation, and disposition of the proposed problems (formulation and disposition are much more important and cost me much more labor than the uninitiated could imagine) I have tried to challenge the reader, awake his curiosity and initiative, and give him ample opportunity to face a variety of research situations.

4. This book deals most of the time with mathematical problems. Nonmathematical problems are rarely mentioned, but they are always present in the background. In fact, I have carefully taken them into consideration and have tried to treat mathematical problems in a way that sheds light on the treatment of nonmathematical problems whenever possible.

This book deals most of the time with elementary mathematical problems. More advanced mathematical problems, however, although seldom referred to, led me to the conception of the material included. In fact, my main source was my own research, and my treatment of many an elementary problem mirrors my experience with advanced problems which could not be included in this book.

5. This book combines its theoretical aim, the study of heuristic, with a concrete, urgent, practical aim: to improve the preparation of high school mathematics teachers.

I have had excellent opportunity to make observations and form opinions on the preparation of high school mathematics teachers, for all my classes have been devoted to such teachers in the last few years. I hope to be a comparatively unprejudiced observer, and as such I can have but one opinion: *the preparation of high school mathematics teachers is insufficient.* Furthermore, I think that all responsible organizations must share the blame, and that especially both the schools of education and the departments of mathematics in the colleges should very carefully revise their offerings to teachers if they wish to improve the present situation.

What courses should the colleges offer to prospective high school teachers? We cannot reasonably answer this question, unless we first answer the related question: *What should the high schools offer to their students?*

Yet this question is of little help, you may think, because it is too controversial; it seems impossible to give an answer that would command sufficient consensus. This is unfortunately so; but there is an aspect of this question about which at least the experts may agree.

Our knowledge about any subject consists of *information* and of *know-how*. If you have genuine *bona fide* experience of mathematical work on any level, elementary or advanced, there will be no doubt in your mind that, in mathematics, know-how is much more important than mere possession of information. Therefore, in the high school, as on any other level, we should impart, along with a certain amount of information, a certain degree of *know-how* to the student.

What is know-how in mathematics? The ability to solve problems—not merely routine problems but problems requiring some degree of independence, judgment, originality, creativity. Therefore, the first and foremost duty of the high school in teaching mathematics is to emphasize *methodical work in problem solving*. This is my conviction; you may not go along with it all the way, but I assume that you agree that problem solving deserves some emphasis—and this will do for the present.

The teacher should know what he is supposed to teach. He should show his students how to solve problems—but if he does not know, how can he show them? The teacher should develop his students' know-how, their ability to reason; he should recognize and encourage creative thinking—but the curriculum he went through paid insufficient attention to his mastery of the subject matter and no attention at all to his know-how, to his ability to reason, to his ability to solve problems, to his creative thinking. Here is, in my opinion, the worst gap in the present preparation of high school mathematics teachers.

To fill this gap, the teachers' curriculum should make room for *creative work on an appropriate level.* I attempted to give opportunity for such work by conducting seminars in problem solving. The present work contains the material I collected for my seminars and directions to use it; see the " Hints to Teachers, and to Teachers of Teachers " at the end of this volume, pp. 209–212. This will, I hope, help to improve the mathematics teacher's preparation; at any rate, this is the practical aim of the present work.

I believe that constant attention to both aims mentioned, the theoretical and the practical, made me write a better book. I believe too that there is no conflict between the interests of the various prospective readers (some concerned with problem solving in general, others with improving their own ability, and still others with improving the ability of their students). What matters to one type of reader has a good chance to be of consequence to the others.

6. The present work is the continuation of two earlier ones, *How to Solve It* and *Mathematics and Plausible Reasoning*; the two volumes of the latter have separate titles: *Induction and Analogy in Mathematics* (vol. 1) and *Patterns of Plausible Inference* (vol. 2). These books complete each other without essential overlapping. A topic considered in one may be reconsidered in another, but then the treatment is different: other examples, other details, or other aspects are offered. And so it does not matter much which one is read first and which one is read later.

For the convenience of the reader, the three works will be compared and corresponding passages listed in a cumulative index at the end of the second volume of this book, *Mathematical Discovery*.

7. To publish the first part of a book when the second part is not yet available entails certain risks. (There is a German proverb: " Don't show a half-built house to a fool.") These risks are not negligible; yet, in the interest of the practical aim of this work, I decided not to delay the publication of this volume; see p. 210.

This first volume contains Part One of the work, Patterns, and two chapters of Part Two, Toward a General Method.

The four chapters of Part One have more extensive collections of problems than the later chapters. In fact, Part One is in many ways similar to a collection of problems in analysis by G. Szegö and the author (see the Bibliography). There are, however, obvious differences: in the present volume the problems proposed are much more elementary, and methodical points are not only suggested but explicitly formulated and discussed.

The second chapter of Part Two is inspired by a recent work of Werner Hartkopf (see the Bibliography). I present here only some points of Hartkopf's work which seem to me the most engaging, and I present them as they best fit my conception of heuristic, with suitable examples and additional remarks.

8. The Committee on the Undergraduate Program in Mathematics supported the preparation of the manuscript of this book by funds granted by the Ford Foundation. I wish to express my thanks, and I wish to thank the Committee also for its moral support. I wish to thank the editor of the *Journal of Education of the Faculty and College of Education, Vancouver and Victoria*, for permission to incorporate parts of an article into the present work. I also wish to thank Professor Gerald Alexanderson, Santa Clara, California, and Professor Alfred Aeppli, Zurich, Switzerland, for their efficient help in correcting the proofs.

GEORGE POLYA

Zurich, Switzerland
December 1961

HINTS TO THE READER

Section 5 of chapter 2 is quoted as sect. 2.5, subsection (3) of section 5 of chapter 2 as sect. 2.5(3), example 61 of chapter 3 as ex. 3.61.

HSI and MPR are abbreviations for titles of books by the author which will be frequently quoted; see the Bibliography.

Iff. The abbreviation " iff " stands for the phrase " if and only if."

†. The sign † is prefixed to examples, comments, sections, or shorter passages that require more than elementary mathematical knowledge (see the next paragraph). This sign, however, is not used when such a passage is very short.

Most of the material in this book requires only *elementary mathematical knowledge*, that is, as much geometry, algebra, " graphing " (use of coordinates), and (sometimes) trigonometry as is (or ought to be) taught in a good high school.

The problems proposed in this book seldom require knowledge beyond the high school level, but, with respect to difficulty, they are often a little above the high school level. The solution is fully (although concisely) presented for some problems, only a few steps of the solution are indicated for other problems, and sometimes only the result is given.

Hints that may facilitate the solution are added to some problems (in parentheses). The surrounding problems may provide hints. Especial attention should be paid to the introductory lines prefixed to the examples (or to certain groups of examples) in some chapters.

The reader who has spent serious effort on a problem may benefit from the effort even if he does not succeed in solving the problem. For example, he may look at some part of the solution, try to extract some helpful information, and then put the book aside and try to work out the rest of the solution by himself.

The best time to think about methods may be when the reader has finished solving a problem, or reading its solution, or reading a case history. With his task accomplished and his experience still fresh in mind, the reader, in *looking back* at his effort, can profitably explore the nature of the difficulty he has just overcome. He may ask himself many useful questions: " What was the decisive point? What was the main difficulty? What could I have done better? I failed to see this point: which item of knowledge, which attitude of mind should I have had to see it? Is there some trick worth learning, one that I could use the next time in a similar situation? " All these questions are good, and there are many others—but the best question is the one that comes spontaneously to mind.

CONTENTS

PART ONE
PATTERNS, 1

PART TWO
TOWARD A GENERAL METHOD, 115

PART ONE

PATTERNS

Each problem that I solved became a rule
which served afterwards to solve other problems.

DESCARTES: *Œuvres*, vol. VI, pp. 20–21; Discours de la Méthode.

If I found any new truths in the sciences,
I can say that they all follow from, or depend on,
five or six principal problems which I succeeded
in solving and which I regard as so many battles
where the fortune of war was on my side.

DESCARTES: *op. cit.*, p. 67.

CHAPTER 1

THE PATTERN OF TWO LOCI

1.1. Geometric constructions

Describing or constructing figures with ruler and compasses has a traditional place in the teaching of plane geometry. The simplest constructions of this kind are used by draftsmen, but otherwise the practical importance of geometric constructions is negligible and their theoretical importance not too great. Still, the place of such constructions in the curriculum is well justified: they are most suitable for familiarizing the beginner with geometric figures, and they are eminently appropriate for acquainting him with the ideas of problem solving. It is for this latter reason that we are going to discuss geometric constructions.

As so many other traditions in the teaching of mathematics, geometric constructions go back to Euclid in whose system they play an important role. The very first problem in Euclid's Elements, Proposition One of Book One, proposes "to describe an equilateral triangle on a given finite straight line." In Euclid's system there is a good reason for restricting the problem to the equilateral triangle but, in fact, the solution is just as easy for the following more general problem: *Describe* (or construct) *a triangle being given its three sides.*

Let us devote a moment to analyzing this problem.

In any problem there must be an *unknown*—if everything is known, there is nothing to seek, nothing to do. In our problem the unknown (the thing desired or required, the *quaesitum*) is a geometric figure, a triangle.

Yet in any problem something must be known or *given* (we call the given things the *data*)—if nothing is given, there is nothing by which we could recognize the required thing: we would not know it if we saw it. In our problem the data are three "finite straight lines" or line segments.

Finally, in any problem there must be a *condition* which specifies how the unknown is linked to the data. In our problem, the condition specifies that the three given segments must be the sides of the required triangle.

The condition is an essential part of the problem. Compare our problem with the following: "Describe a triangle being given its three altitudes." In both problems the data are the same (three line segments) and the unknown is a geometric figure of the same kind (a triangle). Yet the connection between the unknown and the data is different, the condition is different, and the problems are very different indeed (our problem is easier).

The reader is, of course, familiar with the solution of our problem. Let a, b, and c stand for the lengths of the three given segments. We lay down the segment a between the endpoints B and C (draw the figure yourself). We draw two circles, one with center C and radius b, the other with center B and radius c; let A be one of their two points of intersection. Then ABC is the desired triangle.

1.2. From example to pattern

Let us look back at the foregoing solution, and let us look for promising features which have some chance to be useful in solving similar problems.

By laying down the segment a, we have already located two vertices of the required triangle, B and C; just one more vertex remains to be found. In fact, by laying down that segment we have transformed the proposed problem into another problem equivalent to, but different from, the original problem. In this new problem

the unknown is a point (the third vertex of the required triangle);
the data are two points (B and C) and two lengths (b and c);
the condition requires that the desired point be at the distance b from the given point C and at the distance c from the given point B.

This condition consists of two parts, one concerned with b and C, the other with c and B. *Keep only one part of the condition, drop the other part; how far is the unknown then determined, how can it vary?* A point of the plane that has the given distance b from the given point C is neither completely determined nor completely free: it is restricted to a "locus"; it must belong to, but can move along, the periphery of the circle with center C and radius b. The unknown point must belong to two such loci and is found as their intersection.

We perceive here a pattern (the "pattern of two loci") which we can imitate with some chance of success in solving problems of geometric construction:

First, reduce the problem to the construction of ONE point.
Then, split the condition into TWO parts so that each part yields a locus for the unknown point; each locus must be either a straight line or a circle.

Examples are better than precepts—the mere statement of the pattern cannot do you much good. The pattern will grow in color and interest and value with each example to which you apply it successfully.

1.3. Examples

Almost all the constructions which traditionally belong to the high school curriculum are straightforward applications of the pattern of two loci.

(1) *Circumscribe a circle about a given triangle.* We reduce the problem to the construction of the center of the required circle. In the so reduced problem

the unknown is a point, say X;
the data are three points A, B, and C;
the condition consists in the equality of three distances:

$$XA = XB = XC$$

We split the condition into two parts:

First $XA = XB$
Second $XA = XC$

To each part of the condition corresponds a locus. The first locus is the perpendicular bisector of the segment AB, the second that of AC. The desired point X is the intersection of these two straight lines.

We could have split the condition differently: first, $XA = XB$, second, $XB = XC$. This yields a different construction. Yet can the result be different? Why not?

(2) *Inscribe a circle in a given triangle.* We reduce the problem to the construction of the center of the required circle. In the so reduced problem

the unknown is a point, say X;
the data are three (infinite) straight lines a, b, and c;
the condition is that the point X be at the same (perpendicular) distance from all three given lines.
We split the condition into two parts:
First, X is equidistant from a and b.
Second, X is equidistant from a and c.