9
CHAPTER

Some of Our Own Reflections

How should mathematics be done? Of course, the answer to this question cannot take the form of a set of precise prescriptions which guarantee, if they are followed, that we will be successful every time we undertake a piece of mathematics—no such guarantee is ever available, even for the most expert and talented performers on the mathematical stage. Nevertheless, we believe it may be helpful to our readers to formulate certain principles which can be discussed by them, and which may prove of practical use in improving their success rate when actually doing mathematics—and we emphasize that the readers of this book should be expecting to get a lot of satisfaction from doing mathematics (one of our principles is that mathematics is something we *do*, not just something we read and try to learn).

So we will provide a set of principles here which may be useful to you. In fact, we provide *two* sets. The first set of principles is of a general nature, and refers to the overall approach we should take to doing mathematics. The second set of principles is more specific and, we hope, will be found useful when you are actually involved in some particular piece of mathematics. For those of you who have a special interest in teaching mathematics we include a short section on principles of mathematical pedagogy.

9.1 General Principles

1. Mathematics is only done effectively if the experience is enjoyable

Certainly, mathematics is a "serious" subject, in the sense that it is very useful and important. But that does not mean that a mathematical problem is to be tackled in a spirit of grim earnestness. Let us be frank—mathematicians are very happy that mathematics is important—but, for very few mathematicians, is this its prime attraction? We might say it is the reason why they are paid to do mathematics, but not really the reason why they do it. They do it because it brings them joy, fulfillment, and excitement; they find it fun even when it is deep, difficult, and demanding—indeed, this is a large part of its appeal. Our readers may like to read [1] on this theme.

2. Mathematics usually evolves out of communication between like-minded people

Mathematics is not, by its nature, a solitary activity; it is only our obsession with traditional tests and the determination of (some of) us to ensure the honesty of our students which have led us to insist so much on mathematics always being done by students working on their own. Like any other exciting and engrossing activity, mathematics is something to be talked about among friends, to be discussed informally, so that insights and ideas can be shared and developed, and so that many can enjoy the sense of achievement, even triumph, that the successful solution of a challenging problem brings. Much of the most important work that mathematicians do is done at conferences and in the staffroom over coffee.[1]

This crucial kind of social mathematical activity requires that we feel free to chat about mathematical ideas without the restraining necessity to be absolutely precise—we will have more to say about this later. It is one of the (many) advantages that human beings have over machines that they can communicate informally without having to be pedantic. Indeed,

[1]A great contemporary mathematician, Paul Erdös, has defined a mathematician as a device for turning coffee into mathematics.

a stronger statement is broadly true—*human beings can only communicate interesting ideas informally*,[2] while *machines can only communicate* (with each other or with human beings) ***pedantically***.

Finally, let us admit that there is a stage, in the solution of a mathematical problem or the successful completion of a piece of mathematical research, when solitude and silence are essential. The working out of detail can only be accomplished under really tranquil conditions. But mathematics is more than the working out of detail.

3. Never be pedantic; sometimes, but by no means always, be precise

When a new idea is introduced—and especially when a new idea is presented to students or colleagues—it is necessary to make that idea precise. If, for example, we want to discuss whether an equation has a solution, we must specify what kind of numbers we allow. If we want to know the quotient in a division problem, we need to know if we are referring to the *division algorithm* (involving both a quotient and a remainder) or *exact division* (involving only a quotient)—or even some other version of division.

However, once the precise idea has been conveyed, it is no longer necessary, or desirable, to insist on precision in subsequent discussion. We benefit, in all our conversations, from informality—without it there can be no ease of communication, no ready exchange of ideas. If you are talking to your friends about a dog called Jack, you need, at the outset, to specify to which "Jack" you are referring; once you've made that clear, subsequent references can—and should be—simply to "Jack." We refer to this refinement of the original principle as the

Principle of Licensed Sloppiness.

Our readers may find many examples of this principle in the earlier chapters of this book.

They may also be interested in the following logical consequence of the statement of the principle in its original form.

[2]Unfortunately, when we communicate our results in research journals we are compelled to write very formally.

Corollary. *Precision and pedantry are different things.*

4. Elementary arithmetic goes from question to answer; but genuine mathematics also, and importantly, goes from answer to question[3]

What we mean to imply by this slogan is that a false picture of the true nature of mathematics in action is conveyed by our earliest contact with elementary arithmetic, when we are given addition, subtraction, multiplication, or division problems to do, and have to provide exactly correct answers to score maximum points. The questions are uninteresting (and completely standard); and they are not *our* questions. We provide the equally uninteresting answers by carrying out an uninteresting algorithm.

In genuine mathematical activity answers suggest new questions, so that, in an important sense, mathematical work is never complete. It is, moreover, to be thought of as investigation and inquiry rather than the mere execution of mechanical processes. Here again we see a vital difference between human beings and machines—and we see that elementary arithmetic is fit for machines and not for human beings! Our readers should find many examples of this principle in action in the earlier chapters of this book.

5. Algorithms are first resorts for machines, but last resorts for human beings

Consider the problem 31×29. A calculating machine tackles this problem by recognizing it as a multiplication problem involving the product of two positive integers and applies its programmed algorithm for solving such problems. The intelligent human being may reason as follows:

$$31 \times 29 = (30 + 1)(30 - 1) = 30^2 - 1 = 900 - 1 = 899.$$

(But he, or she, probably does these steps mentally.) For the human being this is a great gain in simplicity, and hence a significant saving of time and

[3]The perceptive reader may object that there had to be a question in the first place to generate the answer. We would respond that "In the beginning there was elementary arithmetic."

effort, compared with the standard hand-algorithm for multiplication; for the machine it would be absurd to look for a short cut, when it can do the routine calculation in a flash.

Thus the natural procedures of machines and human beings when faced with an arithmetical or algebraic problem are diametrically opposed. The machine identifies the problem as belonging to a certain class, and then applies an algorithm suitable for solving any problem in that class. Intelligent human beings look for special features of the problem which make it possible to avoid the use of a universal algorithm, that is, to employ a short cut.

Students need to understand why the traditional algorithms work, but it is absurd to drill them so that they can use them ever more accurately and faster. For the machine will always be much more accurate and much faster.

6. Use particular but not special cases

This principle has many applications, both in teaching mathematics and in doing mathematics. When we want to think about a mathematical situation, it is usually a good idea to think about examples of this situation which are particular but typical. For example, if we are asked for the sum of the coefficients in the binomial expansion of $(1+x)^n$, we might look at the cases $n=3$, $n=4$. Then

$$(1+x)^3 = 1+3x+3x^2+x^3, \qquad 1+3+3+1=8,$$
$$(1+x)^4 = 1+4x+6x^2+4x^3+x^4, \qquad 1+4+6+4+1=16.$$

We might well be led to the conjecture that, in general, the sum of the coefficients is 2^n; and our experiments with $(1+x)^3$, $(1+x)^4$ should suggest to us how to prove this conjecture. Notice that we do *not* experiment with $(1+x)^0$; the case $n=0$ is too special.

The technique is, as we have said, a very good way of detecting patterns, and hence of generating conjectures. It is also a useful tool in understanding mathematical statements and arguments, and thus in explaining them to others. Of course, in testing conjectures already formulated, special cases may be used—but particular, nonspecial cases are usually more reliable indicators.

7. Geometry plays a special role in mathematics

We really have in mind here geometry at the secondary and undergraduate levels. We claim that it is a serious mistake to regard geometry as just one more topic in mathematics, like algebra, trigonometry, differential calculus, and so on. In fact, geometry and algebra, the two most important aspects of mathematics at these levels, play essentially complementary roles. Geometry is a source of questions, algebra is a source of answers. Geometry provides ideas, inspiration, insight; algebra provides clarification and systematic solution.

Thus it is particularly absurd to teach geometry and algebra in separate watertight compartments, as is so often done in the United States. Geometry without algebra leaves the student with questions without answers, and hence creates frustration; algebra without geometry provides the student with answers to questions nobody would ask, and hence creates boredom and disillusion. Together, however, they form the basis of a very rich curriculum, involving both discrete and continuous mathematics.

Some may argue that there are methods in geometry (while not disputing that method is the characteristic of algebra). By "in geometry" they mean "in synthetic geometry" and refer to the method of proof by exploiting symmetry, similar triangles, properties of the circle, etc. Of course, it is true that one does exploit these key ideas in any form of geometrical reasoning, but, practically always, one needs a clever trick—for example, an ingenious construction—to complete the argument by "Euclidean" means. Roughly speaking, each geometrical problem, if solved by purely geometric methods, requires its own special idea. And none of us is bright enough to function, in any aspect of our lives, with such an enormous idea-to-problem ratio; we have to make a good idea go a long way. Fortunately, in mathematics, it does! So remember

All Good Ideas in Mathematics Show Up
in a
Variety of Mathematical and Real-World Contexts

8. Symmetry is a pervasive idea in mathematics

It is not only in geometry that we should look for opportunities to exploit symmetry—though the importance of the idea of symmetry in understanding geometrical situations and solving geometrical problems cannot be overemphasized. Symmetry also plays a very important role in algebra—consider, for example, the problem of determining the coefficient of a^3b^7 in the binomial expansion of $(a + b)^{10}$; whatever the answer is, considerations of symmetry show that the coefficient of a^7b^3 must be the same. You will find symmetry much exploited in Chapters 5 and 6. You will also find it playing a very significant role in Chapter 4, both in the geometrical and in the number-theoretical topics of that chapter. We doubt, indeed, if it is absent from *any* chapter. Our recommendation is—***always look for symmetry***.

There are, of course, many other characteristic properties of mathematics, but we will not go into detail about them here. We might mention two such properties, however. First, we often know something can be done without knowing how to do it. There's a wonderful example of this in Chapter 4 where we quote Gauss's discovery of which regular convex polygons can be constructed with straightedge and compass; but his argument gives no rule for carrying out the constructions. There is another example in Chapter 2, where we show that every residue modulo m, which is prime to m, has an order—but our proof provides no means of calculating the order. A second property characteristic of mathematics is that new concepts are introduced to help us to obtain results about already familiar concepts, but play no part in the statement of those results. There are lovely examples of this in Chapter 3, where we introduce certain irrational numbers α and β in order to establish identities connecting Fibonacci and Lucas numbers, which are, of course, integers; and in Chapter 6 where we use *pseudo-Eulerian coefficients*, just introduced, to prove identities relating binomial coefficients.

However, rather than listing these characteristic properties of mathematics systematically, we prefer to turn to certain more specific principles. Of course, the distinction between these and the principles above, which we have called "general", it not absolute. The reader should think of the principles above as relating more to the general *strategy* of doing mathematics, while those that follow relate more to the *tactics* to be used in trying to solve a particular problem.

9.2 Specific Principles

1. Use appropriate notation, and make it as simple as possible

It is obviously important to use good and clear definitions in thinking about a mathematical problem. But it is remarkable the extent to which we can simplify our thinking by using the appropriate notation. Consider, for example, the summation notation where

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

is replaced by the more concise expression

$$\sum_{i=1}^{n} i,$$

which is read "the sum from $i = 1$ to n of i." This notation obviously simplifies algebraic expressions; and, where we may even regard the range of the summation as understood and omit it, we achieve further valuable simplification.

Good notation also often results in reducing the strain on our overburdened memories. For example, in considering polynomials it is usually much better to use the notation

$$a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$$

rather than a notation which begins

$$a + bx + cx^2 + \cdots .$$

(What is then the coefficient of x^n?) To take this point further, in discussing the product of two polynomials (compare the discussion of the third principle, below), it is very nice to write the general rule in the form

$$\left(\sum a_nx^n\right)\left(\sum b_nx^n\right) = \sum c_nx^n,$$

where $c_n = \sum_{r+s=n} a_rb_s$.

Let us see what we have achieved. First, by adopting the summation notation $\sum$, we have been free to write only one typical term to denote a general polynomial. Second, by adopting the convention that $a_n = 0$ if the term x^n doesn't occur, we have not had to worry about the degree of the polynomial; and, since we have a coefficient for each n, we also don't have to stipulate the limits of the summation. Third, the rule for calculating c_n becomes completely general and does not have to take account of the degrees of the polynomials being multiplied. Just compare the resulting simplicity with how we would have to state the rule for multiplying a cubic polynomial by a quadratic polynomial[4] in the "bad old notation." Our polynomials would be

$$a_0 + a_1x + a_2x^2, \quad b_0 + b_1x + b_2x^2 + b_3x^3.$$

(This notation is not at all as bad as that found in many algebra texts!) and the rule would be

$$(a_0 + a_1x + a_2x^2)(b_0 + b_1x + b_2x^2 + b_3x^3) = c_0 + c_1x + c_2x^2 + c_3x^3 + c_4x^4 + c_5x^5,$$

where

$$\begin{aligned}
c_0 &= a_0b_0 \\
c_1 &= a_0b_1 + a_1b_0 \\
c_2 &= a_0b_2 + a_1b_1 + a_2b_0 \\
c_3 &= a_0b_3 + a_1b_2 + a_2b_1 \qquad (\text{remember } a_3 = 0) \\
c_4 &= a_1b_3 + a_2b_2 \\
c_5 &= a_2b_3.
\end{aligned}$$

And even then we would only have a rule for multiplying a cubic by a quadratic, whereas our statement $c_n = \sum_{r+s=n} a_rb_s$ is completely general and even applies to the multiplication of power series.

As another example of good notation, consider the study of quadratic equations. Such an equation is usually presented as $ax^2 + bx + c = 0$ with solutions

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

How much better to take the general equation as

$$ax^2 + 2bx + c = 0$$

with solutions

$$x = \frac{-b \pm \sqrt{b^2 - ac}}{a}.$$

Compare the two formulae, for example, when you want to solve the equation

$$x^2 - 10x + 16 = 0.$$

[4]You may view this argument as an application of our General Principle 6.

Those of you who have already become familiar with the quadratic formula in its first form may decide the effort to change is not worthwhile. We would not argue with that—but you should look out for trouble when you come to study quadratic forms later.

Similarly, the general equation of a circle should be given as

$$x^2 + y^2 + 2gx + 2fy + c = 0$$

so that the center is $(-g, -f)$ and the radius $\sqrt{g^2 + f^2 - c}$; yet many texts in the United States give the equation of a circle as $x^2 + y^2 + ax + by + c = 0$, so that the center is $\left(-\frac{a}{2}, -\frac{b}{2}\right)$, and the radius $\frac{1}{2}\sqrt{a^2 + b^2 - 4c}$. ***UGH***!

There is one small price to pay for adopting good notation. Over the course of an entire book, we must often use the same symbol to refer to quite different mathematical objects. We cannot reserve the symbol F_k for the kth Fibonacci number if we want also to talk about the kth Fermat number and about a sequence $\{F_k\}$ of geometric figures converging to the fractal F. We depend—as we do in everyday life—on the context to make clear which meaning the symbol has. Thus, in ordinary conversation, if somebody says "Can Jack join us on our picnic?," it is presumably clear to which Jack she refers. There are not enough names for us to be able to reserve a unique name for each human being—and there are not enough letters, even if we vary the alphabet and the font, to reserve a unique symbol for each mathematical concept. Thus the fear of repetition should never deter us from adopting the best notation.

2. Be optimistic!

This may strike the reader as banal—surely one should always be optimistic in facing any of life's problems. But we have a special application of this principle to mathematics in mind. When you are trying to prove something in algebra, assume that you are making progress, and keep in mind what you are trying to prove. Suppose, for example, that we wish to prove that

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

We argue by induction (the formula is clearly correct for $n = 1$) and find ourselves needing to show that

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

At this stage we should not just "slog out" the LHS of (9.1). We should note that we are hoping it will equal the RHS, so we take out the factor $\left(\frac{n+1}{2}\right)^2$, present in all three terms of (1), getting, for the LHS,

$$\left(\frac{n+1}{2}\right)^2 (n^2 + 4n + 4); \qquad (2)$$

and this plainly equals the RHS.

You will find that when you have reached the stage of applying this principle automatically you will be much more confident about your ability to do mathematics successfully—such confidence is crucial.

There is, however, one other sense (at least) in which optimism is a valuable principle when one is actually making mathematics, that is, making conjectures and trying to prove them; and we feel we should mention it explicitly. One should be as ambitious as possible! This means that one should test the validity of strong rather than weak statements, and one should formulate conjectures in their most general (but reasonable) form. There are many examples of this aspect of optimism in Chapter 4. Thus we hope to find a means of constructing arbitrarily good approximation to *any* regular convex polygon, or even to *any* regular star polygon. Then, when we try to determine which convex polygons can be folded by the 2-period folding procedures, we quickly decide to look at rational numbers $\frac{t^{m+n}-1}{t^n-1}$, for *any* integer $t \geq 2$, instead of merely the numbers $\frac{2^{m+n}-1}{2^n-1}$ emerging from our paper-folding activities. The resulting gains, in both cases, are immense—not only do we get better results, but we find ourselves inventing new and fruitful concepts.

3. Employ reorganization as an algebraic technique

A simple example of such reorganization may be seen in the strategy of factorizing the polynomial $x^4 + 4$. Polynomials are usually organized, as in this case, by powers of the indeterminate x. However, factorization is often achieved by exhibiting the expression as a difference of two squares. We recognize $x^4 + 4$ as consisting of two "parts" of the expression for $(x^2 + 2)^2$.

Thus we reorganize the expression by writing

$$x^4 + 4 = (x^4 + 4x^2 + 4) - 4x^2$$

and hence achieve the factorization

$$x^4 + 4 = (x^2 + 2x + 2)(x^2 - 2x + 2).$$

(Let us insert here a small anecdote. At the meeting of the New Zealand Association of Mathematics Teachers in Christchurch, New Zealand, in September, 1993, which all three authors attended, the question was going around—Can you show that $n^4 + 4$ (where n is a positive integer) is only prime if $n = 1$? It is clear that this fact follows from the factorization of $x^4 + 4$ above. For this factorization shows that $n^4 + 4$ is not prime unless the smaller factor of $n^4 + 4$, namely, $n^2 - 2n + 2$, is equal to 1. But the equation $n^2 - 2n + 2 = 1$, with n a positive integer, has the single solution $n = 1$.)

4. Look for conceptual proofs

Conceptual proofs are not only almost always simpler than algebraic proofs; by their very nature, they also almost always give us better insight into the reason why a statement is true. An algebraic proof may compel belief in the truth of the statement being proved; but, so often, it does not convey the genuine understanding which makes the student confident in using the result. Let us give a couple of examples.

We claim that if you multiply together r consecutive positive integers, the result is divisible by $r!$ Here's a proof. Let the integers be n, $n - 1$, $n - 2, \ldots, n - r + 1$, with $n \geq r$ (this is perfectly general). Now

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \binom{n}{r},$$

the binomial coefficient. Since $\binom{n}{r}$ may be viewed as the number of ways of selecting r objects from n objects, it is, by its nature, an integer. This completes the proof. A purely arithmetical proof could have been given; but such a proof could not have been described, as this proof may fairly be described, as an *explanation*.

As a second example, also involving binomial coefficients, consider the Pascal Identity

$$\binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r}, \quad r \geq 1. \tag{3}$$

This could be proved by using the formula in the previous paragraph for $\binom{n}{r}$ and "slogging it out"; but you would not then be any the wiser as to why (3) is true. Instead, you could consider $(n+1)$ objects of which one is marked. A selection of r of these objects might, or might not, contain the marked object. We claim that $\binom{n}{r-1}$ of the selections contain the marked object and $\binom{n}{r}$ do not. (Do you see why?) This observation achieves a conceptual proof. Certainly there are also available arithmetical and algebraic proofs; all three types of proof are given in Chapter 6. In this case, the arithmetical and algebraic proofs have the advantage that they are easily extended to the case where n is *any real number* (while r remains a positive integer). However, while the arithmetical proof continues to provide no insight, the algebraic proof shows how the Pascal Identity may be viewed conceptually in an entirely different light, namely, as a special case of the rule for multiplying power series. Thus the conceptual viewpoint once again triumphs!

You should not think that conceptual proofs are always combinatorial. There are other kinds of conceptual proof, which some might label *abstract*, but which we prefer to characterize as *noncomputational.* Chapter 2, though it is all about numbers, is full of such conceptual proofs; among the advantages we get from using them one should especially mention the aesthetic satisfaction they bring.

Conceptual proofs also serve in the solution of geometrical problems, though, frequently, such nonanalytic proofs are hard to find. But the geometrical viewpoint may often provide the crucial insight into the strategy for proving a mathematical assertion which has no obvious geometrical content.

9.3 Appendix: Principles of Mathematical Pedagogy

We have included this Appendix because we invisage that our text may be read by actual and prospective teachers of mathematics. However, we very much hope that all our readers will find it useful.

It is clear that any principle for doing mathematics effectively will imply a principle of sound mathematical pedagogy; and it cannot be necessary for us, writing for readers of this book, to be explicit about how each principle we have enunciated *translates* into a pedagogical principle. It is surely sufficient to enunciate a very general but rather controversial

Basic Principle of Mathematical Instruction

Mathematics should be taught so that students have a chance of comprehending how and why mathematics is done by those who do it successfully.

However, we believe that there are certain pedagogical principles which do *not* follow directly from principles about doing mathematics—or, at least, from those principles which we have explicitly identified. With the understanding that we offer the following list tentatively as a basis for discussion, that we do *not* claim that it is complete, and that, emphatically, we do *not* claim any special insight as teachers, we append here a brief list of such principles.

$\mathbf{P}_1$. **Inculcate a dynamic approach.** Mathematics is something to be *done*, not merely something to be learnt, and certainly not something simply to be committed to memory.

$\mathbf{P}_2$. **Often adopt a historical approach.** Make it plain that, over the centuries, mathematics has been something which intelligent adults have chosen to do. Moreover, the mathematical syllabus was not engraved on the tablets Moses brought down from Mount Sinai. No piece of mathematics has always existed. Each piece has been invented in response to some stimulus, some need; and the best pieces have continued to be used.

$\mathbf{P}_3$. **Recognize the utility of mathematics, but do not underestimate the power of mathematics itself to attract students.** Thus, applications should be used, both as justification and as inspiration for mathematical ideas; but one should not always insist on dealing extensively with applications in presenting a mathematical topic.[5] We should remember that mathematics has its own natural internal dynamic which should guide its development and sequencing, so that it is often philosophically correct, as well as pedagogically sound, to "stay with the mathematics."

P_4. **Insist on the proper, and only the proper, use and design of tests and other evaluation instruments.** Tests are unacceptable unless they contribute to the learning process. Students must never be asked to do mathematics under conditions which no mathematician would tolerate; it follows, of course, that their ability to do mathematics must not be assessed by subjecting them to artificial conditions and restraints. As a minimal requirement, tests must be designed so as not to endanger the crucial relationship between teacher and student.

For further thoughts on tests, the reader may like to consult [2].

P_5. **Where there are at least two different ways of looking at a problem, discuss at least two.** Different students look at problems and ideas in different ways. What is clear to one may be far less clear to another, without this being a reflection of their overall mathematical ability. In particular, some students (and mathematicians) visualize discretely, others continuously. By giving attention to more than one approach, the teacher gives more students the chance to benefit, and enhances the prospect of new connections being made in the students' understanding.

In the context of this principle it is particularly important to insist that one must never cut short an explanation or exposition in order to complete an unrealistically inflated syllabus. Alas, how often have we heard a colleague say words to the effect, "I did not really expect the students to understand, but they will need the technique in their physics course next term!" There is no merit in the teacher completing the syllabus unless the students complete it too!

[5]We cannot support the concept, prevalent in the United States, of a 'problem-driven' curriculum, that is, a curriculum in which mathematical items are introduced as and when they are needed to solve problems coming from outside mathematics.

References

[1] Hilton, P., The joy of mathematics, *Coll. Math. J.*, **23**, 4, 1992, 274–281.

[2] Hilton, Peter, The tyranny of tests, *Amer. Math. Monthly* **100**, 4 (1993), 365–369.

http://www.springer.com/978-0-387-94770-9

Mathematical Reflections
In a Room with Many Mirrors
Hilton, P.; Holton, D.; Pedersen, J.
1997, XVI, 352 p., Hardcover
ISBN: 978-0-387-94770-9