

Ch. VIII.1 by A. Gardiner
from 'The Princeton Companion to Mathematics'
edited by T. Gowers (2008)

Part VIII

Final Perspectives

VIII.1 The Art of Problem Solving

A. Gardiner

Where there are problems, there is life.

Zinoviev (1980)

In English the word "problem" has negative connotations, suggesting some unwanted and unresolved tension. Zinoviev's reminder is therefore important: problems are the stuff of life—and of mathematics. Good problems focus the mind: they challenge and frustrate; they cultivate ambition and humility; they show up the limitations of what we know, and highlight potential sources of more powerful ideas. By contrast, the word "solving" suggests a *release* of tension. The juxtaposition of these two words in the expression "problem solving" may encourage the naive to think that this unwelcome tension can be massaged away by means of some "magic formula" or process. It cannot; there is no magic formula.

Why don't we tell the truth? No one has the faintest idea how the process ... works, and in calling it a "process" we may be already making a dangerous assumption.

Gian-Carlo Rota, in Kac et al. (1986)

A "problem" is something that one wants to understand, to explain, or to solve, but which eludes one's initial attempts to classify it as being of some familiar "type." The experience of being confronted by such a "problem" is inevitably unsettling: it may eventually prove to be more familiar than one thought, but the would-be solver is initially dumped in terrain with few signposts or marked tracks. Some (such as Pólya and his recent followers) have tried to devise a universal "problem-solving meta-map." But in reality there is no easy alternative to that painful immersion so familiar to generations of postgraduate students.

Grand general principles can help to make sense of this experience, but are unlikely to take us very

far. Consider, for example, the four general principles formulated by DESCARTES [VI.11] in his *Discourse on Method*.

The first was never to accept anything for true which I did not clearly know to be such. The second, to divide each of the difficulties under examination into as many parts as possible, and as might be necessary for its adequate solution. The third, to conduct my thoughts in such order that by commencing with objects the simplest and easiest to know, I might ascend ... step by step to the knowledge of the more complex. And the last ... to make enumerations so complete ... that I might be assured that nothing was omitted.

Descartes's rules are worth pondering. But it is hard to accept that it was the systematic application of these four rules that led to Descartes's almost single-handed creation of analytic geometry as we know them today! In the detailed working out of the creative process, problem-specific "know-how" distilled from endless hands-on experience is likely to be far more important than any general principles. What then can one usefully say? To describe the "art of problem solving" in impressive-sounding detail would be irresponsible. But to say nothing would be misleading. Both options are unsatisfactory—yet these two responses are what students, teachers, and would-be mathematicians are most likely to meet! Attempts to teach "problem-solving" in schools often misconstrue mathematics as a kind of "subjective pattern-spotting." Instead of correcting this distortion at university level, mathematicians often maintain a discreet public silence about the very private matter of *how* serious mathematical problems actually get solved. Hence, in addressing the theme for readers with a mathematical bent, this article has to start largely from scratch, and to proceed slowly. So we begin with a warning. The subject of problem solving is well worth exploring, but we shall proceed obliquely and our conclusions will often remain implicit. Along the way we shall meet extracts from a number of sources—which may be viewed as an initial reading list for those

who wish to pursue the theme in greater detail, provided they never forget that the only way to gain true insight into a craft is *through practicing the craft itself*. Mathematics may be "the queen of the sciences," but the art of *doing* mathematics remains a *craft*, passed on in the ancient craft tradition, through painful initiation. A number of collections of problems at various levels—often using relatively elementary material—are listed in the references. Here we make do with a single example.

Problem. For all positive integers n and k , show that some triangular number is congruent to $k \pmod{2^n}$.

The reader is encouraged to explore this problem before reading on, noting any obvious stages along the way: from initial bewilderment, through an exploratory/organization phase, eventually culminating in a solution and an attempt to locate this isolated challenge in some broader mathematical context.

Mathematics is a largely unexplored "mental universe," whose initial exploration and charting, subsequent colonization, routine traverse, and efficient administration correspond, in many ways, to the real-world adventures of geographical explorers in former centuries. To strike out beyond the security of the old-world coastline, to imagine and explore something new, takes intellectual courage.

Most prominent among these mathematical explorers are the "system builders," who identify new mathematical continents, or who uncover profound and unexpected bridges joining known lands. Their initial motivation may stem from a specific problem, whose analysis provides hints of the outline of previously undiscerned structures; but the system builder's focus then switches to the bigger picture: trying to identify, and to clarify, connections between the structures that underlie "mathematics in the large." Such ventures often end up with little to show for them—they may come close to discovering some mathematical El Dorado, but they lack the gold to prove it. Some of these explorers may later be singled out as major prophets or discoverers, but such recognition can be fickle: those so honored may not have been the first to see their particular promised land; they may not have appreciated the significance of what they had stumbled upon, or of how it would eventually be seen to link known mathematical lands; their success may have depended on earlier attempts by others; and their bounty may not have impressed their contemporaries as deeply as we now imagine.

Each triumph of the system builders is rooted in detailed knowledge of "mathematics in the small," which may derive from work in a very different mathematical style—such as that of the mathematical beachcomber, who is most at home exploring the *known* mathematical shoreline, using some sixth sense to spot suspicious-looking rocks, under which are hidden intricate, and totally unexpected, microworlds on our very doorstep. While great explorers range further and further afield, they leave behind annoying gaps, or unsolved problems, which represent significant lacunas in our understanding—gaps that some future beachcomber may one day explain, so opening the way for some new synthesis.

The system builder and the beachcomber represent very different mental styles; but their contributions complement each other. In our evolving picture of the mathematical universe, insights on a small scale and on a large scale must somehow fit together. Hence the beachcomber's chance discoveries may contribute in unexpected ways to our future conception of the large-scale mathematical universe.

Such differing styles should be borne in mind as we strive to make our introductory comments more specific. Our first attempt is based on a version of Alain Connes's three levels of mathematical activity.

The first [level] is defined by the faculty of calculation—being able to apply a given algorithm rapidly and reliably.... The second level begins when the actual method of calculation is adapted to, and criticized in the context of, a particular problem.... In mathematics this is what often makes it possible to solve problems that aren't too difficult or that don't require any new ideas.... The third level [is] the level at which the mind, or rather conscious thought, is occupied with another task while the problem in question is being solved... subconsciously.... At [the third] level it isn't only a matter of solving a given problem; it is also possible to discover... a part of mathematics to which the [previously] existing corpus gives no direct access.

Alain Connes, in Changeux and Connes (1995)

Connes's first level focuses on the development of *robust technique*—that is, fluency, accuracy, and confidence in using given procedures in relatively standard ways. We say no more about work on this level except to stress its importance! Discussion about the "art of problem solving" presupposes, and only makes sense in the context of, appropriate robust technique.

Connes's second level includes most, but by no means all, of the serious mathematics that mathematicians engage in on a daily basis. Genuine problems

occur on this level in different guises, ranging from (i) challenges designed to stretch the young would-be mathematician (in high school geometry, in puzzle books and problem-solving journals, in Olympiads, etc., whose material is designed to force the would-be solver to select, to adapt, and to combine known methods in unexpected ways), to (ii) genuine research problems that can be tackled and largely solved by selecting, adapting, and combining known methods in a suitably imaginative way.

In our problem about triangular numbers, the first level includes the immediate translation from words into symbols to obtain the congruence $m(m-1)/2 \equiv k \pmod{2^n}$, or $m(m-1) \equiv 2k \pmod{2^{n+1}}$, which, for arbitrary given $n \geq 1$, has to be solved for all $k \geq 1$. The second level might then include a systematic attempt to make sense of what happens for small values of n , leading to the formulation of simple conjectures whose proofs would solve the problem, followed by moves to devise the necessary proofs.

It is tempting to think of Connes's third level as "inscrutable," in the spirit of the following extract:

In science, as well as in other fields of human endeavor, there are two kinds of geniuses: the "ordinary" and the "magicians." An ordinary genius is a fellow that you or I would be just as good as if we were only many times better [than we are]. There is no mystery as to how his mind works. Once we understand what he has done, we feel certain that we, too, could have done it. It is different with the magicians. They are... in the orthogonal complement of where we are and the working of their minds is for all intents and purposes incomprehensible. Even after we understand what they have done, the process by which they have done it is completely in the dark. They seldom if ever have students because they cannot be emulated and it must be terribly frustrating to cope with the mysterious ways in which a magician's mind works.

Kac (1985)

However, one would then expect activity on this level to be so idiosyncratic as to be irrelevant to ordinary mortals. In fact, the most valuable insights we have into "the art of problem solving" derive from personal testimony about work on this level by precisely such "magicians" as POINCARÉ [VI.61], which suggests that there are clear parallels between the experience of the very best mathematicians on Connes's third level and what happens when ordinary students, or mathematicians, operate "out of their depth" when tackling more mundane problems; that is, when their own fumbling requires them to work in regions

to which *their own* "existing corpus gives no direct access." In our problem about triangular numbers this might occur when a solver who has never met "congruences for binomial coefficients" manages to adapt the naive proof for $\binom{m}{2} \pmod{2^n}$ to cover the slightly more awkward $\binom{m}{3} \pmod{2^n}$, and realizes that, even though this naive approach does not extend to $\binom{m}{4} \pmod{2^n}$, something more general may be lurking in the darkness.

Thus we use the word "problem" to refer to a *serious mathematical challenge on at least Connes's second level*, where this is to be interpreted in the spirit of activity on Connes's second and third levels. So any analysis of the art of mathematical problem solving must somehow reflect experience on these two higher levels. By contrast, the educational assumptions that underpin most attempts to bring "problem-solving" to the classroom generally try to reduce this subtle process to a set of rules *in the spirit of Connes's first level!*

A problem is much more than just a hard exercise. Consider the question, When is a "problem" *not* a problem? One answer is clearly, When it is too *easy*! However, many students and teachers are tempted to reject unfamiliar or mildly confusing problems because they appear to be *too hard*. This is an understandable reaction only where mathematics is limited to a succession of predictable exercises.

Most of us learn mathematics as a collection of standard techniques, which we use to solve standard problems in predictable contexts (Connes's first level). Like the athlete or musician, the mathematician student needs to develop technique. However, as the athlete trains in order to *compete*, and the musician practices in order to *make music*, so the mathematician needs technique in order to *make mathematics* by tackling challenging problems. Each new piece of printed music may initially strike the beginner as a confusing array of black blobs. But as they work on the piece, phrase by phrase, it slowly takes on a shape of its own, revealing internal connections that may previously have been overlooked. Much the same is true when we confront an unfamiliar mathematical problem. At first sight we may not even understand the question. But as we struggle to make sense of the problem, we regularly find that, little by little, the fog begins to lift.

Two rats fell into a can of milk. After swimming for a time one of them realized his hopeless fate and drowned. The other persisted, and at last the milk was turned to butter and he could get out.

In the first part of the war, Miss Cartwright and I got drawn into van der Pol's equation.... [W]e went on and on... with no earthly prospect of "results": suddenly the entire vista of the dramatic fine structure of solutions stared us in the face.

Littlewood (1986)

In 1923 HARDY [VI.73] and LITTLEWOOD [VI.79] made a conjecture about the number of arithmetic progressions (APs) of length k among the primes. One potential corollary was that the prime numbers must contain *arbitrarily long* APs. Faced with such a claim it is natural to start looking for APs which consist entirely of primes! But if you try, you will soon approach the limits of what is known: the first three odd primes, 3, 5, 7, form a very familiar AP of length *three*, but longer APs are surprisingly elusive (in 2004 the record for an AP of distinct primes had length twenty-three, with both the primes themselves and the step size being astronomical). Despite this unpromising lack of evidence, in 2004 Ben Green and Terence Tao proved that the set of prime numbers does indeed contain arbitrarily long APs. Their proof is a fine example of the way in which significant progress often combines a detailed reevaluation of known results (in this case a deep result of Szemerédi), lateral thinking (they embed the primes not in the integers, but in a natural but sparser set of "almost primes" of which the primes constitute a *nonzero* fraction), and the determination and ingenuity to make such ideas deliver the goods.

It remains a serious challenge to capture the essence of Littlewood's experience (where the fog suddenly lifts) in a form that is suitable for relative beginners, whether through time-constrained problems (see Barbeau 1989; Gardiner 1997; Lovasz 1979), or through structured investigations (see Gardiner 1987; Ringel 1974). In the year in which Green and Tao announced their proof, the British Mathematical Olympiad posed the following problem, which readers are encouraged to tackle.

Problem. In an AP of seven distinct primes, what is the smallest possible value of the largest prime?

This challenge could enliven any introductory number theory course, as well as providing a natural link to recent developments. For the novice it is far from obvious how to begin, but the basic idea is elementary and should be "known" (in some sense), and can be used to generate natural APs of lengths 4, 5, 6, 7, 8, provided that one accepts the value of carrying out extensive computations quickly and intelligently.

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.

From the preface to the first printing of Pólya (2004)

Pólya is, if anything, too reticent here. The important distinction is not between that which is "known" and that which is truly "original," but rather between mathematical activity in the spirit of Connes's first level and mathematical activity in the spirit of Connes's second and third levels. Any introduction to this distinction is inevitably through problems whose solution is *known to someone*, so we should collect and use good "modest problems," not apologize for them. Ulam puts it more directly.

I learned chess from my father.... The moves of the knight fascinated me, especially the way two enemy pieces can be threatened simultaneously with one knight. Although it is a simple stratagem, I thought it was marvelous, and I have loved the game ever since.

Could the same process apply to the talent for mathematics? A child by chance has some satisfying experiences with numbers; then he experiments further and enlarges his memory by building up a store of experiences.

Ulam (1991)

Children also find delight—if less profound and more short-lived—in the discovery that one can set up a "corner move" in the children's game noughts and crosses (tic-tac-toe) so as to *simultaneously* threaten to complete two lines-of-three, at most one of which can be countered. This delight in a double-edged strategy, which points in two directions at once, has much in common with the pleasure we derive from (i) puns and *double entendres* in ordinary language, in humor, and in poetry, (ii) the almost physical response when we recognize subtle variations on a theme in music, and (iii) the more cerebral appreciation we feel when we meet counting methods based on unanticipated isomorphisms, or the essentially two-faced idea of "proof by contradiction" in mathematics. This enjoyment of hidden ambiguities and double meanings is related to the evident (but poorly understood) way in which *analogy* guides, and delights, mathematicians of all ages.

Banach once told me, "Good mathematicians see analogies between theorems or theories; the very best ones see analogies between analogies."

Ulam (1991)

Koestler, in his thought-provoking book *The Act of Creation* (1976), shows how scientific and literary "creativity" often flows from the identification and exploitation of "double meanings with a built-in tension." (Koestler calls them *bisociations*: "the perceiving of a situation or idea *L* in two self-consistent but habitually incompatible frames of reference... the event *L* is made to vibrate simultaneously on two different wavelengths, as it were.") His study begins with an analysis in precisely this vein of the human response to humor, both comic and tragic, including a selection of jokes attributed to VON NEUMANN [VI.91]!

Ulam's innocent-sounding question (in the extract before last) challenges us not only to provide children with "satisfying experiences with numbers," but also to identify other quintessential aspects of mathematics and to ensure that they are experienced memorably at school (and undergraduate) level. In particular, insofar as there is such a thing as an "art of problem solving," we need to learn how to convey it faithfully and effectively through the medium of classical elementary mathematics to those who are near the beginning of their mathematical studies, or who may not yet have any commitment to mathematics.

It is often claimed that Pólya's little book *How to Solve It* provides an answer. It does not. Pólya was a pioneer who sought to provoke a debate among mathematicians about "heuristics." This debate never really got started. Instead his first low-level attempt at a theoretical framework has been embraced uncritically.

Much of what Pólya writes about specific problems in *How to Solve It* makes sense; but his general conclusions on "how to help students solve problems" are less convincing. As a result, much of the book's general theorizing needs to be read extremely carefully. For example, Pólya's suggestion that "when the teacher solves a problem before the class, he should dramatize his ideas a little and he should put to himself the same questions which he uses when helping students" is spot on. But alarm bells should start ringing when he confidently concludes that "[thanks] to such guidance, the student will eventually... acquire something that is more important than the knowledge of any particular mathematical fact." In the right setting the claim may occasionally be true; but as a statement about the effect on students in general it is false.

Similar claims have been widely used to justify the introduction of a whole new branch of school mathematics called "problem-solving" (see NCTM (1980) and www.pisa.oecd.org), which has grown at the expense of mastery of the "particular mathematical facts" on which the activity itself depends.

Pólya and others were right to insist that school mathematics should include a regular diet of good problems, and that educators have a duty to convey not just the techniques and inner logical structure of the subject, but also the experience of struggling to uncover the mathematics hidden in multistep problems and carefully structured investigations. Fortunately, the four volumes that Pólya wrote to illustrate this broader thesis remain in print (Pólya 1981, 1990). There the focus is on mathematics, and the rhetoric is more restrained:

[L]et us learn proving, but also *let us learn guessing*.... I do not believe there is a foolproof method to learn guessing. At any rate, if there is such a method, I do not know it, and quite certainly I do not pretend to offer it in the following pages.... [P]lausible reasoning is a practical skill and it is learned, as any other practical skill, by imitation and practice.

Pólya (1990, volume 1)

These four books should be compulsory reading for all serious mathematics educators, graduate students, and mathematics lecturers. However, Pólya and others failed to show how problem solving could be developed *within* the standard school mathematics curriculum. Instead they concentrated on proposing general rules that might "help students become better problem solvers." What is needed is to clarify (i) which aspects of elementary mathematics have the potential to captivate young minds—not because they are more "enjoyable" in some superficial sense, but because they are more "pregnant with meaning"; and (ii) *how to teach* such material so as to convey this deeper meaning on an elementary level. This is not the place for a detailed analysis, but we suspect such an analysis would *strengthen* the position of many traditionally important topics and themes, encouraging them to be taught in such a way as to bring out their inherent richness, while recognizing that these goals depend on prior mastery of certain basic techniques without which this richness can scarcely be appreciated. In contrast, recent "reforms," whose declared intention was to *enrich* school mathematics, have regularly *reduced* both the emphasis on, and the time available for, serious elementary mathematics.

Those who want good problems to enrich school mathematics often fail to recognize that well-intentioned "reforms" are usually unstable under the kind of distortions that routinely affect large-scale educational change (where the cultivation of professional competence, sensitivity, independence, and responsibility among teachers is regularly replaced by centralized control via a fragmented list of separate "outcomes," which are then assessed in ways that actively discourage good teaching).

Small-scale experiments can also have unintended side effects! As a little-known example of a radical attempt to cultivate the art of problem solving at school level we offer Eisenstein's account of his own education at lower secondary school (1833-37).

[E]ach student had to prove the theorems consecutively. No lecture took place at all. No one was allowed to tell his solutions to anybody else and each student received the next theorem to prove, independent of the other students, as soon as he had proved the preceding one correctly, and as long as he had understood the reasoning.... While my peers were still struggling with the eleventh or twelfth, I had already proved the hundredth.... [T]his method... can probably not be adapted.... One does not obtain that overview of the whole subject, which can only be achieved by a good lecture.... In the end, the best mathematical genius cannot discover alone what has been discovered by the collaboration of many outstanding minds.... For students this method is only practicable if it deals with small fields of easily understandable knowledge, especially geometric theorems, which do not require new insights and ideas.

Eisenstein (1975)

Eisenstein was a remarkable mathematician. Yet at the tender age of twenty, on the threshold of the mathematical world that he longed to inhabit, he could see the limitations of this approach—even for students such as himself.

Problems that cultivate a taste for problem solving tend to incorporate certain characteristic features, such as simplicity, rhythm, naturalness, elegance, and surprise; and their solutions are often double-edged. But their most important feature is that, while their solution should be within reach of those in the target audience, the statement of the problem should convey no direct hint as to how to begin. Indeed, a good problem may continue to frustrate the would-be solver for a disturbingly long period.

A tacit rite of passage for the mathematician is the first sleepless night caused by an unsolved problem.

Reznick (1994)

The role of sleep and sleeplessness in creative problem solving is well documented (if poorly understood). It often features within the "incubation" phase of HADAMARD's [VI.65] "four phases" (discussed below), which summarize the process through which the initial experience of helplessness and leaden frustration is sometimes transmuted into golden success.

Such success is neither mechanical, nor the result of pure chance. In solving a good problem—as with a good puzzle—there is no magic problem-solving method that might relieve us of the need to struggle: the struggle may sometimes be fruitless, but it is an important part of the process. Thus, a successful outcome generally presupposes a certain kind of preparatory hard work. When asked how he made his discoveries, GAUSS [VI.26] is said to have answered, "Durch planmässiges Tattonieren," that is, through systematic and persistent groping around!

Having discovered a way into a problem, one may realize that it "should have been obvious" where to begin; but things are often obvious only in retrospect. One learns by experience how a certain kind of persistence can cause the fog that initially surrounds an unfamiliar problem to magically evaporate; what was at first invisible then stands out so clearly that one can scarcely understand how it could ever have been missed.

When faced with an unfamiliar mathematical problem, the mathematician, young or old, is like someone who is trying to open some fiendishly difficult Chinese puzzle box with a hopelessly small bunch of keys. At first glance the surface seems totally smooth, without a single visible crack. If you were not convinced that it was indeed a Chinese puzzle box, and that it could in fact be opened, you would soon give up. Knowing (or rather *believing*) that it can be opened, you may be willing to keep searching until you eventually begin to discern the slightest hint of a crack here and there. You may still have no idea how the pieces are meant to move, or which of your "keys" may help you to open up the first layer of the puzzle, but by trying the most appropriate-looking keys in the most promising cracks you eventually stumble on one that fits exactly, and the pieces begin to move. The job is certainly not done; but the mood has changed and you feel you are well on the way.

As we have already seen, this experience of initial confusion, giving way as one grapples with a problem to unexpected insight, is in no way confined to beginners. It is part of the very nature of mathematics and the way human beings do mathematics. If a problem

unfamiliar, its solution may require persistence, faith, and much time. So one should never give up too easily, and should always be prepared to look back after solving a problem to see what one could perhaps have done differently.

It is most important in creative science not to give up. If you are an optimist you will be willing to "try" more than if you are a pessimist. It is the same in games like chess. A really good chess player tends to believe (sometimes mistakenly) that he holds a better position than his opponent. This, of course, helps to keep the game moving and does not increase the fatigue that self-doubt engenders. Physical and mental stamina are of crucial importance in chess and also in creative scientific work.

Ulam (1991)

Persistence is of course easier to sustain if one has a degree of optimism about the likely outcome, or if one has cultivated the sheer "bloody-mindedness" that makes one refuse to give up (as with Littlewood's surviving rat). However, there are dangers.

I learned, subconsciously, from Mazur how to control my inborn optimism and how to verify details. I learned to go more slowly over intermediate steps with a skeptical mind and not to let myself be carried away.

Ulam (1991)

At the International Congress of Mathematicians in Paris in 1900, HILBERT [VI.63] presented twenty-three major research problems, which he judged would be important for the development of mathematics in the twentieth century. These problems seemed very hard; yet in bringing them to the attention of his fellow mathematicians Hilbert felt the need to stress that this should not be used as an excuse for putting off trying to solve them.

However unapproachable these problems may seem to us and however helpless we stand before them, we have, nevertheless, the firm conviction that their solution must follow by a finite number of purely logical processes.... This conviction of the solvability of every mathematical problem is a powerful incentive to the worker. We hear within us the perpetual call: There is the problem. Seek its solution. You can find it by pure reason; for in mathematics there is no *ignorabimus*.

During the nineteenth century it became clear that the more that scientists discovered about nature, the more they realized *how little they knew*, and that one could never hope to discover "the whole truth." This realization was summed up by the physiologist Emil du Bois-

Reymond in the phrase "*ignoramus et ignorabimus*"—ignorant we are and ignorant we shall remain. As the new century dawned, Hilbert felt that it was important to state as clearly as he could that *mathematics is different*. In mathematics, he said, we can tackle problems with "the firm conviction that their solution must follow by a finite number of purely logical processes." As if to underline his assertion, one of his problems was solved almost immediately (though the most famous, the RIEMANN HYPOTHESIS [IV.2 §3], remains unresolved).

Hilbert was talking about mathematical *research*: but his principle applies even more strongly when tackling problems from textbooks, Olympiads, or university courses. When faced with an unfamiliar and apparently very difficult mathematical problem, we have little choice about how to proceed: we must either tackle the problem using the "bunch of keys" or mathematical techniques that we already know (no matter how limited they may be), or put off trying. Of course it is important to learn new tricks, and to revise old ones, as we go along. And of course there is always the temptation to imagine that the problem we face is simply *too* hard, that progress toward a solution requires some trick or technique that we have not yet learned and that the solution is therefore beyond our powers. This defeatist view is all the more plausible because it must sometimes be true! Mathematicians know perfectly well that, strictly speaking, the assumption that every problem can be solved is irrational (in that it cannot be justified logically, and is in general clearly false: we now know that some problems are intrinsically insoluble as stated). It is *nevertheless an invaluable working hypothesis*. Thus we should never let such doubts interfere with the basic hypothesis that *every problem we tackle has to be solved using essentially the techniques that we already know* (deployed with sufficient ingenuity!). Though strictly illogical, the assumption that every problem can be solved has justified itself so often in practice that it becomes a powerful conviction—a conviction that is psychologically invaluable each time we experience that feeling of helplessness when trying to get to grips with a hard mathematical problem.

Hilbert's judgment that his problems would play a central role in the mathematics of the twentieth century was remarkably astute. But the most interesting thing for us here is his rallying call: however unapproachable these problems may seem at first sight, and however helpless we stand before them, we have the firm conviction that their solution must be possible by

purely logical processes. "There is the problem. Seek its solution. You can find it by pure reason." As in most printed mathematics, Hilbert offered no psychological guidance on how to proceed. Those who took up Hilbert's challenge were expected to discover such things for themselves.

Like every social activity, mathematics has a "front" and a "back": the *front* is where the finished products are displayed for public consumption, while the *back* is where the real work is done in less presentable surroundings. A naive realist might view the *front* as a mere facade, insist that all serious "problem solving" goes on "out back," and declare this separation to be artificial.

Sometime, in a future that is knocking at our door, we shall have to retrain ourselves and our children to properly tell the truth. The exercise will be particularly painful in mathematics. The enrapturing discoveries of our field systematically conceal, like footprints erased in the sand, the analogical train of thought that is the authentic life of mathematics.... Until that day, however, the truths of mathematics will make only fleeting appearances, like shameful confessions whispered to a priest, to a psychiatrist, or to a wife.

In the nineteenth chapter of "The Betrothed," Manzoni describes as follows the one genuine moment in a conversation between astute Milanese diplomats: "It was as if, between acts in the performance of an opera, the curtain were to be raised too soon, and the spectators were given a glimpse of a half-dressed soprano screaming at the tenor."

Gian-Carlo Rota, in Kac et al. (1986)

However, the prospect of some mathematical equivalent of being obliged to witness "a half-dressed soprano screaming at the tenor" should cause us to hesitate before embracing Rota's vision of the future.

The *front-back* metaphor is due to the sociologist Erving Goffman. One standard example is that of a restaurant. We tend to think of a restaurant in terms of what we see "out front," where the manners, food, and language are "all dressed up"; but everything we see out front is totally dependent on the raw heat, the steam and grease, the conflicts and curses "out back" in the kitchen—where the hard work is done to tight deadlines and in very different conditions.

The triumph of mathematics in the modern world has been largely due to the fact that these two worlds—the front and the back—have been deliberately and systematically separated. It may seem curious that we have no agreed way of discussing the dynamics of the mathematical kitchen; but mathematics has grown

largely because its practitioners have learned to separate its *objective* results, and the way they are validated and presented, from the intriguing, but inscrutable (and ultimately irrelevant!) *subjective* alchemy through which these mathematical results are conjured up. This formal separation has led to the adoption of a universally communicable format, which transcends personal taste and style, and which can therefore be comprehended, checked, and improved by anyone. Any move to pay greater attention to the mental, physical, and emotional dynamics that underlie mathematical problem solving must understand the need for this separation and respect the formal world of "objective" mathematics.

There are intriguing insights into the human dynamics of the mathematical kitchen scattered throughout the mathematical literature. One such insight is the fact that different mathematicians may have very different styles, even though most of these differences are rarely discussed. One example is the perceived role of *memory*. Some mathematicians value memory highly.

It seems to me that a good memory—at least for mathematicians and physicists—forms a large part of their talent. And what we call talent or perhaps genius itself depends to a large extent on the ability to use one's memory properly to find the analogies, past, present and future, which, as Banach said, are essential to the development of new ideas.

Ulam (1991)

Others have an excellent memory for anything *within their own field of interest*, but have considerable difficulty storing information from outside that domain in an easily retrievable form. And many would-be mathematicians are drawn to the subject precisely because they see it as requiring markedly less memorizing than most other disciplines. The important point would seem to be not *how much* one remembers, but *what* one makes automatic, and *how accessibly* this and other information is stored. It is clearly worth making a serious effort to organize in one's mind that material which is central to one's own work—so that it is available for instant use. It is also important, as we shall see, to collect a penumbra of possibly useful ideas, information, and examples—so that the mind is in a position to make incidental connections which might be fruitful. But it is not necessarily wise to learn in a uniform way everything that might conceivably be needed for the problem at hand: knowing slightly less sometimes forces the mind to get by on less, and hence to be more ingenious or inventive.

VIII.1. The Art of Problem Solving

Hadamard's Four Phases

Littlewood's (1986) numerous perceptive observations concerning his contemporaries highlight other differences in style—such as speed and working habits. Similar insights may be found in many of the livelier mathematical autobiographies, but Littlewood's remarks are especially valuable.

With a good deal of diffidence I will try to give some practical advice about research and the strategy it calls for. In the first place research work is of a different "order" from the learning process of pre-research education (essential as it is). The latter can easily be rote-memory, with little associative power: on the other hand, after a month's immersion in research the mind knows its problem as much as the tongue knows the inside of one's mouth. You must acquire the art of "thinking vaguely", an elusive idea that I can't elaborate in short form.... I should stress the importance of giving the subconscious every chance. There should be relaxed periods during the working day, profitably I say spent in walking.

Littlewood (1986)

At one stage Poincaré thought there might be just two main styles of mathematical thinking:

The one sort are above all preoccupied with logic.... The other sort are guided by intuition and at the first stroke make quick but sometimes precarious conquests.... [O]ne often says of the first that they are *analysts* and calls the others *geometers*.

Poincaré (1904)

But in identifying the label "logical" with that of "analyst," and the label "intuitive" with that of "geometer," he noticed that HERMITE [VI.47] constituted a counterexample—an "intuitive analyst"! Clearly, the range of mathematical styles is more complex (see Hadamard 1945, chapter VII). One consequence is that any analysis of the art of problem solving in general needs to be drawn with a broad brush. Despite this caveat, Hadamard's "four-phase" model of mathematical creativity has found widespread acceptance, so it may help if one's work habits respect these phases:

It is usual to distinguish four phases in creation: preparation, incubation, illumination and verification, or working out.... Preparation is largely conscious, and anyhow *directed* by the conscious. The essential problem has to be stripped of accidentals and brought clearly into view; all relevant knowledge surveyed; possible analogues pondered. It should be kept constantly before the mind during intervals of other work....

Incubation is the work of the subconscious during the waiting time, which may be several years. Illumination, which can happen in a fraction of a second, is the emergence of the creative idea into the conscious. This almost always occurs when the mind is in a state of relaxation and engaged lightly with ordinary matters.... Illumination implies some mysterious rapport between the subconscious and the conscious, otherwise emergence could not happen. What rings the bell at the right moment?

Littlewood (1986)

Pólya's *How to Solve It* proposes a less convincing four-stage "recipe" for the problem-solving process ("understand, plan, act, reflect"), which has nevertheless been widely used at school level. Hadamard's four phases provide a useful framework for thinking and communicating about the creative process; they also separate the relatively routine aspects (which one may be able to influence more easily) from the more elusive ones. The "conscious *preparation*" phase is perhaps the most mundane stage, requiring a combination of method and discipline. Littlewood again offers sound advice. He recognizes that his advice may not suit all tastes, but he insists that we would all benefit from trying different patterns of working in order to identify and cultivate habits that are as effective as possible.

Most people need half an hour or so before being able to concentrate fully.... The natural impulse towards the end of a day's work is to finish the immediate job: this is of course right if stopping would mean doing work all over again. But try to end in the middle of something; in a job of writing out, stop in the middle of a sentence. The usual recipe for warming up is to run over the latter part of the previous day's work; this dodge is a further improvement.... When I am working really hard I wake around 5.30 a.m. ready and eager to start; if I am slack I sleep till I am called.

Littlewood (1986)

At some ill-defined stage, this preparation achieves a sufficiently clear understanding of the immediate problem, together with a level of saturation in relevant background information, to enable the mind to begin trying different approaches and combinations of ideas. We have reached the *incubation* phase.

We cannot know all the facts, since they are practically infinite in number.... Method is precisely the selection of facts.

Poincaré (1908)

I've often observed too that once the first hurdle of preparation has been surmounted, one runs up against a wall. The main error to be avoided is trying to attack the problem head-on. During the incubation phase you have to proceed indirectly, obliquely.... Thought needs to be liberated in such a way that subconscious work can take place.

Alain Connes, in Changeux and Connes (1995)

Temperament, general character, and "hormonal" factors must play a very important role in what is considered to be purely "mental" activity.... A "subconscious brewing" (or pondering) sometimes produces better results than forced, systematic thinking.... [W]hat we call originality... might to some extent consist of a methodical way of exploring all avenues—an almost automatic sorting of attempts....

When I remember a mathematical proof, it seems to me that I remember only salient points, markers, as it were, of pleasure or difficulty. What is easy is easily passed over because it can be reconstructed logically with ease. If, on the other hand, I want to do something new or original, then it is no longer a question of syllogism chains. When I was a boy I felt that the role of rhyme in poetry was to compel one to find the unobvious because of the necessity of finding a word which rhymes. This forces novel associations and almost guarantees deviations from routine chains or trains of thought. It becomes paradoxically a sort of automatic mechanism of originality.... What people think of as inspiration or illumination is really the result of much subconscious work and association through channels in the brain of which one is not aware at all.

Ulam (1991)

It takes two to invent anything. The one makes up combinations; the other one chooses, recognizes what he wishes and what is important to him in the mass of the things which the former has imparted to him. What we call genius is much less the work of the first one than the readiness of the second one to grasp the value of what has been laid before him and to choose it.

Paul Valéry, quoted in Hadamard (1945)

We have reached a double conclusion: that invention is choice [and] that this choice is imperatively governed by the sense of scientific beauty.

Hadamard (1945)

Part of the pleasure (and pain), the magic (and masochism) of mathematics stems from the fact that the next step—from incubation to illumination—remains so mysteriously elusive. *Illumination* can occur at any time. In most cases—especially where the realization is of something relatively straightforward—this occurs

during periods of "official work." However, this need not be so, especially when the corner to be illuminated is especially dark or unfamiliar, or if the leap of imagination required is large. In such cases it seems that, after the hard graft of the preparation and incubation phases, the mind often needs to "step back" in order to see the way forward more clearly. That is, hard work needs to be combined with relaxation, as Connes implies when he warns against "trying to attack the problem head-on." In one oft-quoted example, Poincaré recalls how he realized the profound connection between Fuchsian functions and hyperbolic geometry as he stepped aboard a bus while on a day out! The first three extracts below show that the mind may achieve this in-between state as a result of *sleeplessness*, or *in the very act of waking*. The fourth extract concerns strenuous *hill walking*. What is common to them all is that the moment of enlightenment does not occur while the beneficiary is officially working!

It was his custom to tell his friends that if others would meditate as long and as deeply as he did on mathematical truths, they would be able to make his discoveries. He said that often he meditated for days on a piece of research without finding a solution, which finally became clear to him after a sleepless night.

Dunnington (1955)

One phenomenon is certain and I can vouch for its absolute certainty: the sudden and immediate appearance of a solution at the very moment of sudden awakening. On being very abruptly awakened by an external noise a solution long searched for appeared to me without the slightest instant of reflection on my part... and in a quite different direction from any of those which I had previously tried to follow.

Hadamard (1945)

Most striking at first is this appearance of sudden illumination, a manifest sign of long, conscious prior work.... The role of this unconscious work in mathematical invention appears to me incontestable....

For a fortnight I had been attempting to prove that there could not be any function analogous to what I have since called Fuchsian functions. I was at that time very ignorant. Every day I sat down at my table and spent an hour or two trying a great number of combinations, and I arrived at no result. One night I took some black coffee, contrary to my custom, and was unable to sleep. A host of ideas kept surging in my head; I could feel them jostling one another, until two of them coalesced, so to speak, to form a stable combination. When morning came, I had established the existence of

one class of Fuchsian functions, those that are derived from the hypergeometric series. I had only to verify the results, which only took a few hours. Poincaré (1908)

I had been struggling for two months to prove a result I was pretty sure was true. When ... walking up a Swiss mountain, fully occupied by the effort, a very odd device emerged—so odd that though it worked I could not grasp the resulting proof as a whole.... I had a sense that my subconscious was saying, "Are you *never* going to do it, confound you; try this."

Littlewood (1986)

The resulting sense of satisfaction is familiar even to those whose mathematical experience is limited.

Illumination is not only marked by the pleasure—the exhilaration!—one inevitably experiences at the moment it strikes, but also by the relief one suddenly feels at seeing a fog abruptly lift, and disappear.

Alain Connes in Changeux and Connes (1995)

However, after months of hard work, such intoxication can sometimes be deceptive.

In mathematics one cannot stop at drawing with a big, wide brush; all the details have to be filled in at some time.

Ulam (1991)

The *verification*, or working-out, process often appears mundane; but it is rarely routine, and regularly reveals hidden subtleties that force us to reassess the anticipated approach. Unforeseen difficulties may remain unresolved, and we may be obliged reluctantly to begin the cycle all over again. It is tempting to think of this as "failure." But mathematics is not a mere machine for solving problems; it is a way of life. In their different ways success and failure both send us back to the drawing board—as Gauss observed in a letter to BOLYAI [VI.34] in 1808.

It is not knowledge, but the act of learning, not possession but the act of getting there, which grants the greatest enjoyment. When I have clarified and exhausted a subject, then I turn away from it, in order to go into darkness again; the never-satisfied man is so strange—if he has completed a structure, then it is not in order to dwell in it peacefully, but in order to begin another. I imagine the world conqueror must feel thus, who after one kingdom is scarcely conquered, stretches out his arms for others.

Further Reading

- Barbeau, E. 1989. *Polynomials*. New York: Springer.
- Changeux, J.-P., and A. Connes. 1995. *Conversations on Mind, Matter, and Mathematics*. Princeton, NJ: Princeton University Press.
- Dixon, J. D. 1973. *Problems in Group Theory*. New York: Dover.
- Dunnington, G. W. 1955. *Carl Friedrich Gauss: Titan of Science*. New York: Hafner. (Reprinted with additional material by J. J. Gray, 2004. Washington, DC: The Mathematical Association of America.)
- Eisenstein, G. F. 1975. *Mathematische Werke*. New York: Chelsea. (English translation available at <http://www-ub.massey.ac.nz/~wwilms/research/letters/volume6/>.)
- Engel, A. 1991. *Problem-Solving Strategies*. Problem Books in Mathematics. New York: Springer.
- Gardiner, A. 1987. *Discovering Mathematics: The Art of Investigation*. Oxford: Oxford University Press.
- . 1997. *The Mathematical Olympiad Handbook: An Introduction to Problem Solving*. Oxford: Oxford University Press.
- Hadamard, J. 1945. *The Psychology of Invention in the Mathematical Field*. Princeton, NJ: Princeton University Press. (Reprinted 1996.)
- Hilbert, D. 1902. Mathematical problems. *Bulletin of the American Mathematical Society* 8:437–79.
- Kac, M. 1985. *Enigmas of Chance: An Autobiography*. Berkeley, CA: University of California Press.
- Kac, M., G.-C. Rota, and J. T. Schwartz. 1986. *Discrete Thoughts: Essays on Mathematics, Science, and Philosophy*. Boston, MA: Birkhäuser.
- Koestler, A. 1976. *The Act of Creation*. London: Hutchinson.
- Littlewood, J. E. 1986. *A Mathematician's Miscellany*. Cambridge: Cambridge University Press.
- Lovasz, L. 1979. *Combinatorial Problems and Exercises*. Amsterdam: North-Holland.
- NCTM. 1980. *Problem Solving in School Mathematics*. Reston, VA: National Council of Teachers of Mathematics.
- Newman, D. 1982. *A Problem Seminar*. New York: Springer.
- Poincaré, H. 1904. *La Valeur de la Science*. Paris: E. Flammarion. (In *The Value of Science: Essential Writings of Henri Poincaré* (2001), and translated by G. B. Halsted. New York: The Modern Library.)
- . 1908. *Science et Méthode*. Paris: E. Flammarion. (In *The Value of Science: Essential Writings of Henri Poincaré* (2001), and translated by F. Maitland. New York: The Modern Library.)
- Pólya, G. 1981. *Mathematical Discovery*, two volumes combined. New York: John Wiley.
- . 1990. *Mathematics and Plausible Reasoning*, two volumes. Princeton, NJ: Princeton University Press.
- . 2004. *How to Solve It*. Princeton, NJ: Princeton University Press.

- Pólya, G., and G. Szego. 1972. *Problems and Theorems in Analysis*, two volumes. New York: Springer.
- Reznick, B. 1994. Some thoughts on writing for the Putnam. In *Mathematical Thinking and Problem Solving*, edited by A. H. Schoenfeld. Mahwah, NJ: Lawrence Erlbaum.
- Ringel, G. 1974. *Map Color Theorem*. New York: Springer.
- Roberts, J. 1977. *Elementary Number Theory: A Problem Oriented Approach*. Cambridge, MA: MIT Press.
- Ulam, S. 1991. *Adventures of a Mathematician*. Berkeley, CA: University of California Press.
- Yaglom, A. M., and I. M. Yaglom. 1987. *Challenging Mathematical Problems with Elementary Solutions*, two volumes. New York: Dover.
- Zeitz, P. 1999. *The Art and Craft of Problem Solving*. New York: John Wiley.
- Zinoviev, A. A. 1980. *The Radiant Future*. New York: Random House.

VIII.2 "Why Mathematics?" You Might Ask Michael Harris

It seems to me that they have a poor opinion of our religion if they think it needs the protection of philosophy.

Lorenzo Valla, *Dialogue on Free Will*

1 A Metaphysical Burden

ANDRÉ WEIL [VI.93], speaking at the 1978 International Congress of Mathematicians at Helsinki, concluded his address entitled "History of Mathematics: Why and How?" with these words:

Thus my original question "Why mathematical history?" finally reduces itself to the question "Why mathematics?," which fortunately I do not feel called upon to answer.

Proceedings of the ICM, Helsinki, 1978
(pp. 227-36, quotation on p. 236)

I heard Weil's address, and the applause that followed, and remember imagining circumstances in which that final question could not be so easily evaded. For instance, in 1991 the House Committee on Science, Space, and Technology called upon the American Mathematical Society (AMS) to answer a very similar question: "What are the main goals in the mathematical sciences?" Weil knew his audience, and the committee of twelve mathematicians responding to the government body responsible for research budgets knew theirs:

The most important long-term goals for the mathematical sciences are: provision of fundamental tools for science and technology, improvement of mathematics education, discovery of new mathematics, facilitation of technology transfer, and support of efficient computation.¹

"Meaning is what makes things sell," wrote Roland Barthes (1967), and the AMS adopted the posture of FOURIER [VI.25], who, according to a celebrated comment of JACOBI [VI.35], included in a letter to LEGENDRE [VI.24] of July 2, 1830,

...had the opinion that the principal aim of mathematics was public utility and explanation of natural phenomena; but a philosopher like him should have known that the sole end of science is the honor of the human mind.

It might seem that the AMS has left a place for "honor" in its third goal, but a later elaboration of that goal directs the reader toward "unexpected" applications of pure mathematics.

Few pure mathematicians are as indifferent to practical applications as HARDY [VI.73], who in *A Mathematician's Apology* famously claimed that: "Judged by all practical standards, the value of my mathematical life is nil." But it is fair to assume that, when they are addressing one another rather than government committees, most pure mathematicians (including those who represented the AMS in 1991) would choose a quite different list of "most important long-term goals."

In this they have long been able to count on the protection of philosophy. It has been a commonplace since Plato to grant mathematics intrinsic value on metaphysical grounds.² The topos of mathematics as a source of certain knowledge was already well established by the second century, when Ptolemy wrote

Only mathematics, if one attacks it critically, provides for those who practice it sure and unswerving knowledge, since the demonstration comes about

1. From "Pilot assessment of the mathematical sciences (prepared for the House Committee on Science, Space, and Technology)." *Notices of the American Mathematical Society* 39 (1992):101-10.

2. The present essay is mainly concerned with metaphysical certainty. Descartes wrote in *Principles of Philosophy* (chapter CCVII) of "certainty...founded on the metaphysical ground that, as God is supremely good and the source of all truth, the faculty of distinguishing truth from error which he gave us, cannot be fallacious so long as we use it aright, and distinctly perceive anything by it," and cites "the demonstrations of mathematics" as his first example. Plato (*in Republic*, VII, 522-31) saw mathematics rather as a source of "knowledge of that which exists forever." Certainty and its cognates are some, but only some, of the apparent blessings of mathematics that so impressed certain philosophers as to "infect" the whole of their work, as Ian Hacking (2000) argues.