From L'Enseignement mathématique, vol. XVI, no. 2, pp. 123–152, 1970.

# HOW TO WRITE MATHEMATICS

P. R. HALMOS

## 0. PREFACE

This is a subjective essay, and its title is misleading; a more honest title might be HOW I WRITE MATHEMATICS. It started with a committee of the American Mathematical Society, on which I served for a brief time, but it quickly became a private project that ran away with me. In an effort to bring it under control I asked a few friends to read it and criticize it. The criticisms were excellent; they were sharp, honest, and constructive; and they were contradictory. "Not enough concrete examples" said one; "don't agree that more concrete examples are needed" said another. "Too long" said one; "maybe more is needed" said another. "There are traditional (and effective) methods of minimizing the tediousness of long proofs, such as breaking them up in a series of lemmas" said one. "One of the things that irritates me greatly is the custom (especially of beginners) to present a proof as a long series of elaborately stated, utterly boring lemmas" said another.

There was one thing that most of my advisors agreed on; the writing of such an essay is bound to be a thankless task. Advisor 1: "By the time a mathematician has written his second paper, he is convinced he knows how to write papers, and would react to advice with impatience." Advisor 2: "All of us, I think, feel secretly that if we but bothered we could be really first rate expositors. People who are quite modest about their mathematics will get their dander up if their ability to write well is questioned." Advisor 3 used the strongest language; he warned me that since I cannot possibly display great intellectual depth in a discussion of matters of technique, I should not be surprised at "the scorn you may reap from some of our more supercilious colleagues".

My advisors are established and well known mathematicians. A credit line from me here wouldn't add a thing to their stature, but my possible misunderstanding, misplacing, and misapplying their advice might cause them annoyance and embarrassment. That is why I decided on the unscholarly procedure of nameless quotations and the expression of nameless thanks. I am not the less grateful for that, and not the less eager to acknowledge that without their help this essay would have been worse.

"Hier stehe ich; ich kann nicht anders."

## 1. There is no recipe and what it is

I think I can tell someone how to write, but I can't think who would want to listen. The ability to communicate effectively, the power to be intelligible, is congenital, I believe, or, in any event, it is so early acquired that by the time someone reads my wisdom on the subject he is likely to be invariant under it. To understand a syllogism is not something you can learn; you are either born with the ability or you are not. In the same way, effective exposition is not a teachable art; some can do it and some cannot. There is no usable recipe for good writing.

Then why go on? A small reason is the hope that what I said isn't quite right; and anyway, I'd like a chance to try to do what perhaps cannot be done. A more practical reason is that in the other arts that require innate talent, even the gifted ones who are born with it are not usually born with full knowledge of all the tricks of the trade. A few essays such as this may serve to "remind" (in the sense of Plato) the ones who want to be and are destined to be the expositors of the future of the techniques found useful by the expositors of the past.

The basic problem in writing mathematics is the same thing as in writing biology, writing a novel, or writing directions for assembling a harpsichord: the problem is to communicate and idea. To do so, and to do it clearly, you must have something to say, and you must have someone to say it to, you must organize what you want to say, and you must arrange it in the order you want it said in, you must write it, rewrite it, and re-rewrite it several times, and you must be willing to think hard about and work hard on mechanical details such as diction, notation, and punctuation. That's all there is to it.

## 2. Say something

It might seem unnecessary to insist that in order to say something well you must have something to say, but it's no joke. Much bad writing, mathematical and otherwise, is caused by a violation of that first principle. Just as there are two ways for a sequence not to have a limit (no cluster points or too many), there are two ways for a piece of writing not to have a subject (no ideas or too many).

The first disease is the harder one to catch. It is hard to write many words about nothing, especially in mathematics, but it can be done, and the result is bound to be hard to read. There is a classic crank book by Carl Theodore Heisel [5] that serves as an example. It is full of correctly spelled words strung together in grammatical sentences, but after three decades of looking at it every now and then I still cannot read two consecutive pages and make a one-paragraph abstract of what they say; the reason is, I think, that they don't say anything.

The second disease is very common: there are many books that violate the principle of having something to say by trying to say too many things. Teachers of elementary mathematics in the U.S.A. frequently complain that all calculus books are bad. That is a case in point. Calculus books are bad because there is no such subject as calculus; it is not a subject because it is many subjects. What we call calculus nowadays is the union of a dab of logic and set theory, some axiomatic theory of complete ordered fields, analytic geometry and topology, the latter in both the "general" sense (limits and continuous functions) and the algebraic sense (orientation), real-variable theory properly so called (differentiation), the combinatoric symbol manipulation called formal integration, the first steps of low-dimensional measure theory, some differential geometry, the first steps of the classical analysis of the trigonometric, exponential, and logarithmic functions, and, depending on the space available and the personal inclinations of the author, some cook-book differential equations, elementary mechanics, and a small assortment of applied mathematics. Any one of these is hard to write a good book on; the mixture is impossible.

Nelson's little gem of a proof that a bounded harmonic function is a constant [7] and Dunford and Schwartz's monumental treatise on functional analysis [3] are examples of mathematical writings that have something to say. Nelson's work is not quite half a page and Dunford-Schwartz is more than four thousand times as long, but it is plain in each case that the authors had an unambiguous idea of what they wanted to say. The subject is clearly delineated; it is a subject; it hangs together; it is something to say.

To have something to say is by far the most important ingredient of good exposition—so much so that if the idea is important enough, the work has a chance to be immortal even if it is confusingly misorganized and awkwardly expressed. Birkhoff's proof of the ergodic theorem [1] is almost maximally confusing, and Vanzetti's "last letter" [9] is halting and awkward, but surely anyone who reads them is glad that they were written. To get by on the first principle alone is, however, only rarely possible and never desirable.

## 3. Speak to someone

The second principle of good writing is to write for someone. When you decide to write something, ask yourself who it is that you want to reach. Are you writing a diary note to be read by yourself only, a letter to a friend, a research announcement for specialists, or a textbook for undergraduates? The problems are much the same in any case; what varies is the amount of motivation you need to put in, the extent of informality you may allow yourself, the fussiness of the detail that is necessary, and the number of times things have to be repeated. All writing is influenced by the audience, but, given the audience, an author's problem is to communicate with it as best he can.

Publishers know that 25 years is a respectable old age for most mathematical books; for research papers five years (at a guess) is the average age of obsolescence. (Of course there can be 50-year old papers that remain alive and books that die in five.) Mathematical writing is ephemeral, to be sure, but if you want to reach your audience now, you must write as if for the ages.

I like to specify my audience not only in some vague, large sense (e.g., professional topologists, or second year graduate students), but also in a very specific, personal sense. It helps me to think of a person, perhaps someone I discussed the subject with two years ago, or perhaps a deliberately obtuse, friendly colleague, and then to keep him in mind as I write. In this essay, for instance, I am hoping to reach mathematics students who are near the beginning of their thesis work, but, at the same time, I am keeping my mental eye on a colleague whose ways can stand mending. Of course I hope that (a) he'll be converted to my ways, but (b) he won't take offence if and when he realizes that I am writing for him.

There are advantages and disadvantages to addressing a very sharply specified audience. A great advantage is that it makes it easier the mind reading that is necessary; a disadvantage is that it becomes tempting to indulge in snide polemic comments and heavy-handed "in" jokes. It is surely obvious what I mean by the disadvantage, and it is obviously bad; avoid it. The advantage deserves further emphasis.

The writer must anticipate and avoid the reader's difficulties. As he writes, he must keep trying to imagine what in the words being written may tend to mislead the reader, and what will set him right. I'll give examples of one or two things of this kind later; for now I emphasize that keeping a specific reader in mind is not only helpful in this aspect of the writer's work, it is essential.

Perhaps it needn't be said, but it won't hurt to say, that the audience actually reached may differ greatly from the intended one. There is nothing that guarantees that a writer's aim is always perfect. I still say it's better to have a definite aim and hit something else, than to have an aim that is too inclusive or too vaguely specified and have no chance of hitting anything. Get ready, aim, and fire, and hope that you'll hit a target: the target you were aiming at, for choice, but some target in preference to none.

## 4. ORGANIZE FIRST

The main contribution that an expository writer can make is to organize and arrange the material so as to minimize the resistance and maximize the insight of the reader and keep him on the track with no unintended distractions. What, after all, are the advantages of a book over a stack of reprints? Answer: efficient and pleasant arrangement, emphasis where emphasis is needed, the indication of interconnections, and the description of the examples and counterexamples on which the theory is based; in one word, organization.

The discoverer of an idea, who may of course be the same as its expositor, stumbled on it helter-skelter, inefficiently, almost at random. If there were no way to trim, to consolidate, and to rearrange the discovery, every student would have to recapitulate it, there would be no advantage to be gained from standing "on the shoulders of giants", and there would never be time to learn something new that the previous generation did not know.

Once you know what you want to say, and to whom you want to say it, the next step is to make an outline. In my experience that is usually impossible. The ideal is to make an outline in which every preliminary heuristic discussion, every lemma, every theorem, every corollary, every remark, and every proof are mentioned, and in which all these pieces occur in an order that is both logically correct and psychologically digestible. In the ideal organization there is a place for everything and everything is in its place. The reader's attention is held because he was told early what to expect, and, at the same time and in apparent contradiction, pleasant surprises keep happening that could not have been predicted from the bare bones of the definitions. The parts fit, and they fit snugly. The lemmas are there when they are needed, and the interconnections of the theorems are visible; and the outline tells you where all this belongs.

I make a small distinction, perhaps and unnecessary one, between organization and arrangement. To organize a subject means to decide what the main headings and subheadings are, what goes under each, and what are the connections among them. A diagram of the organization is a graph, very likely a tree, but almost certainly not a chain. There are many ways to organize most subjects, and usually there are many ways to arrange the results of each method of organization in a linear order. The organization is more important than the arrangement, but the latter frequently has psychological value.

One of the most appreciated compliments I paid an author came from a fiasco; I botched a course of lectures based on his book. The way it started was that there was a section of the book that I didn't like, and I skipped it. Three sections later I needed a small fragment from the end of the omitted section, but it was easy to give a different proof. The same sort of thing happened a couple of times more, but each time a little ingenuity and an ad hoc concept or two patched the leak. In the next chapter, however, something else arose in which what was needed was not a part of the omitted section but the fact that the results of that section were applicable to two apparently very different situations. That was almost impossible to patch up, and after that chaos rapidly set in. The organization of the book was tight; things were there because they were needed; the presentation had the kind of coherence which makes for ease in reading and understanding. At the same time the wires that were holding it all together were not obtrusive; they became visible only when a part of the structure was tampered with.

Even the least organized authors make a coarse and perhaps unwritten outline; the subject itself is, after all, a one-concept outline of the book. If you know that you are writing about measure theory, then you have a two-word outline, and that's something. A tentative chapter outline is something better. It might go like this: I'll tell them about sets, and then measures, and then functions, and then integrals. At this stage you'll want to make some decisions, which, however, may have to be rescinded later; you may for instance decide to leave probability out, but put Haar measure in.

There is a sense in which the preparation of an outline can take years, or, at the very least, many weeks. For me there is usually a long time between the first joyful moment when I conceive the idea of writing a book and the first painful moment when I sit down and begin to do so. In the interim, while I continue with my daily bread and butter work, I daydream about the new project, and, as ideas occur to me about it, I jot them down on loose slips of paper and put them helter-skelter in a folder. An "idea" in this sense may be a field of mathematics I feel should be included, or it may be an item of notation; it may be a proof, it may be an aptly descriptive word, or it may be a witticism that, I hope, will not fall flat but will enliven, emphasize, and exemplify what I want to say. When the painful moment finally arrives, I have the folder at least; playing solitaire with slips of paper can be a big help in preparing the outline.

In the organization of a piece of writing, the question of what to put in is hardly more important than what to leave out; too much detail can be as discouraging as none. The last dotting of the last i, in the manner of the old-fashioned Cours d'Analyse in general and Bourbaki in particular, gives satisfaction to the author who understands it anyway and to the helplessly weak student who never will; for most serious-minded readers it is worse than useless. The heart of mathematics consists of concrete examples and concrete problems. Big general theories are usually afterthoughts based on small but profound insights; the insights themselves come from concrete special cases. The moral is that it's best to organize your work around the central, crucial examples and counterexamples. The observation that a proof proves something a little more general than it was invented for can frequently be left to the reader. Where the reader needs experienced guidance is in the discovery of the things the proof does not prove; what are the appropriate counterexamples and where do we go from here?

## 5. Think about the alphabet

Once you have some kind of plan of organization, an outline, which may not be a fine one but is the best you can do, you are almost ready to start writing. The only other thing I would recommend that you do first is to invest an hour or two of thought in the alphabet; you'll find it saves many headaches later.

The letters that are used to denote the concepts you'll discuss are worthy of thought and careful design. A good, consistent notation can be a tremendous help, and I urge (to the writers of articles too, but especially to the writers of books) that it be designed at the beginning. I make huge tables with many alphabets, with many fonts, for both upper and lower case, and I try to anticipate all the spaces, groups, vectors, functions, points, surfaces, measures, and whatever that will sooner or later need to be baptized. Bad notation can make good exposition bad and bad exposition worse; ad hoc decisions about notation, made mid-sentence in the heat of composition, are almost certain to result in bad notation.

Good notation has a kind of alphabetical harmony and avoids dissonance. Example: either $ax+by$ or $a_1x_1+a_2x_2$ is preferable to $ax_1+bx_2$. Or: if you must use $\Sigma$ for an index set, make sure you don't run into $\sum_{\sigma\in\Sigma} a_\sigma$. Along the same lines: perhaps most readers wouldn't notice that you used $|z| < \varepsilon$ at the top of the page and $z \,\varepsilon\, U$ at the bottom, but that's the sort of near dissonance that causes a vague non-localized feeling of malaise. The remedy is easy and is getting more and more nearly universally accepted: $\in$ is reserved for membership and $\varepsilon$ for ad hoc use.

Mathematics has access to a potentially infinite alphabet (e.g., $x, x', x''$, $x''', \ldots$), but, in practice, only small finite fragment of it is usable. One reason is that a human being's ability to distinguish between symbols is very much more limited than his ability to conceive of new ones; another reason is the bad habit of freezing letters. Some old-fashioned analysts would speak of "$xyz$-space", meaning, I think, 3-dimensional Euclidean space, plus the convention that a point of that space shall always be denoted by "$(x,y,z)$". This is bad: it "freezes" $x$, and $y$, and $z$, i.e., prohibits their use in another context, and, at the same time, it makes it impossible (or, in any case, inconsistent) to use, say, "$(a,b,c)$" when "$(x,y,z)$" has been temporarily exhausted. Modern versions of the custom exist, and are no better. Example: matrices with "property L"—a frozen and unsuggestive designation.

There are other awkward and unhelpful ways to use letters: "CW complexes" and "CCR groups" are examples. A related curiosity that is probably the upper bound of using letters in an unusable way occurs in Lefschetz [6]. There $x_i^p$ is a chain of dimension $p$ (the subscript is just an index), whereas $x_p^i$ is a co-chain of dimension $p$ (and the superscript is an index). Question: what is $x_3^2$?

As history progresses, more and more symbols get frozen. The standard examples are $e$, $i$, and $\pi$, and, of course, $0, 1, 2, 3, \ldots$. (Who would dare write "Let 6 be a group."?) A few other letters are almost frozen: many readers would feel offended if "$n$" were used for a complex number, "$\varepsilon$" for a positive integer, and "$z$" for a topological space. (A mathematician's nightmare is a sequence $n_\varepsilon$ that tends to zero as $\varepsilon$ becomes infinite.)

Moral: do not increase the rigid frigidity. Think about the alphabet. It's a nuisance, but it's worth it. To save time and trouble later, think about the alphabet for an hour now; then start writing.

## 6. Write in spirals

The best way to start writing, perhaps the only way, is to write on the spiral plan. According to the spiral plan the chapters get written and re-written in the order 1, 2, 1, 2, 3, 1, 2, 3, 4, etc. You think you know how to write Chapter 1, but after you've done it and gone on to Chapter 2, you'll realize that you could have done a better job on Chapter 2 if you had done Chapter 1 differently. There is no help for it but to go back, do Chapter 1 differently, do a better job on Chapter 2, and then dive into Chapter 3. And, of course, you know what will happen: Chapter 3 will show up the weaknesses of Chapters 1 and 2, and there is no help for it … etc., etc., etc. It's an obvious idea, and frequently an unavoidable one, but it may help a future author to know in advance what he'll run into, and it may help him to know that the same phenomenon will occur not only for chapters, but for sections, for paragraphs, for sentences, and even for words.

The first step in the process of writing, rewriting, and re-rewriting, is writing. Given the subject, the audience, and the outline (and, don't forget, the alphabet), start writing, and let nothing stop you. There is no better incentive for writing a good book than a bad book. Once you have a first draft in hand, spiral-written, based on a subject, aimed at an audience, and backed by as detailed an outline as you could scrape together, then your book is more than half done.

The spiral plan accounts for most of the rewriting and re-rewriting that a book involves (most, but not all). In the first draft of each chapter I recommend that you spill your heart, write quickly, violate all rules, write with hate or with pride, be snide, be confused, be "funny" if you must, be unclear, be ungrammatical—just keep on writing. When you come to rewrite, however, and however often that may be necessary, do not edit, but rewrite. It is tempting to use a red pencil to indicate insertions, deletions, and permutations, but in my experience it leads to catastrophic blunders. Against human impatience, against the all too human partiality everyone feels toward his own words, a red pencil it much too feeble a weapon. You are faced with a first draft that any reader except yourself would find all but unbearable; you must be merciless about changes of all kinds, and, especially, about wholesale omissions. Rewrite means write again—every word.

I do not literally mean that, in a 10-chapter book, Chapter 1 should be written ten times, but I mean something like three or four. The chances are that Chapter 1 should be re-written, literally, as soon as Chapter 2 is finished, and, very likely, at least once again, somewhere after Chapter 4. With luck you'll have to write Chapter 9 only once.

The description of my own practice might indicate the total amount of rewriting that I am talking about. After a spiral-written first draft I usually rewrite the whole book, and then add the mechanical but indispensable reader's aids (such as a list of prerequisites, preface, index, and table of contents). Next I rewrite again, this time on the typewriter, or, in any event, so neatly and beautifully that a mathematically untrained typist can use this version (the third in some sense) to prepare the "final" typescript with no trouble. The rewriting in this third version is minimal; it is usually confined to changes that affect one word only, or in the worst case, one sentence. The third version is the first that others see. I ask friends to read it, my wife reads it, my students may read parts of it, and, best of all, an expert junior-grade, respectably paid to do a good job, reads it and is encouraged not to be polite in his criticisms. The changes that become necessary in the third version, can, with good luck, be effected with a red pencil; with bad luck they will cause one third of the pages to be retyped. The "final" typescript is based on the edited third version, and, once it exists, it is read, reread, proofread, and reproofread. Approximately two years after it was started (two working years, which may be much more than two calendar years) the book is sent to the publisher. Then beings another kind of labor pain, but that is another story.

Archimedes taught us that a small quantity added to itself often enough becomes a large quantity (on, in proverbial terms, every little bit helps). When it comes to accomplishing the bulk of the world's work, and, in particular, when it comes to writing a book, I believe that the converse of Archimedes' teaching is also true: the only way to write a large book is to keep writing a small bit of it, steadily every day, with no exception, with no holiday. A good technique, to help the steadiness of your rate of production, is to stop each day by priming the pump for the next day. What will you begin with tomorrow? What is the content of the next section to be; what is its title? (I recommend that you find a possible short title for each section, before or after it's written, even if you don't plan to print section titles. The purpose is to test how well the section is planned: if you cannot find a title, the reason may be that the section doesn't have a single unified subject.) Sometimes I write tomorrow's first sentence today; some authors begin today by revising and rewriting the last page or so of yesterday's work. In any case, end each work session on an up-beat; give your subconscious something solid to feed on between sessions. It's surprising how well you can fool yourself that way; the pump-priming technique is enough to overcome the natural human inertia against creative work.

## 7. Organize always

Even if your original plan of organization was detailed and good (and especially if it was not), the all-important job of organizing the material does not stop when the writing starts; it goes on all the way through the writing and even after.

The spiral plan of writing goes hand in hand with the spiral plan of organization, a plan that is frequently (perhaps always) applicable to mathematical writing. It goes like this. Begin with whatever you have chosen as your basic concept—vector spaces, say—and do right by it: motivate it, define it, give examples, and give counterexamples. That's Section 1. In Section 2 introduce the first related concept that you propose to study—linear dependence, say—and do right by it: motivate it, define it, give examples, and give counterexamples, and then, this is the important point, review Section 1, as nearly completely as possible, from the point of view of Section 2. For instance: what examples of linearly dependent and independent sets are easily accessible within the very examples of vector spaces that Section 1 introduced? (Here, by the way, is another clear reason why the spiral writing plan is necessary: you may think, in Section 2, of examples of linearly dependent and independent sets in vector spaces that you forgot to give as examples in Section 1.) In Section 3 introduce your next concept (of course just what that should be needs careful planning, and, more often, a fundamental change of mind that once again makes spiral writing the right procedure), and, after clearing it up in the customary manner, review Sections 1 and 2 from the point of view of the new concept. It works, it works like charm. It is easy to do, it is fun to do, it is easy to read, and the reader is helped by the firm organizational scaffolding, even if he doesn't bother to examine it and see where the joins come and how they support one another.

The historical novelist's plots and subplots and the detective story writer's hints and clues all have their mathematical analogues. To make the point by way of an example: much of the theory of metric spaces could be developed as a "subplot" in a book on general topology, in unpretentious comments, parenthetical asides, and illustrative exercises. Such an organization would give the reader more firmly founded motivation and more insight than can be obtained by inexorable generality, and with no visible extra effort. As for clues: a single word, first mentioned several chapters earlier than its definition, and then re-mentioned, with more and more detail each time as the official treatment comes closer and closer, can serve as an inconspicuous, subliminal preparation for its full-dress introduction. Such a procedure can greatly help the reader, and, at the same time, make the author's formal work much easier, at the expense, to be sure, of greatly increasing the thought and preparation that goes into his informal prose writing. It's worth it. If you work eight hours to save five minutes of the reader's time, you have saved over 80 man-hours for each 1000 readers, and your name will be deservedly blessed down the corridors of many mathematics buildings. But remember: for an effective use of subplots and clues, something very like the spiral plan of organization is indispensable.

The last, least, but still very important aspect of organization that deserves mention here is the correct arrangement of the mathematics from the purely logical point of view. There is not much that one mathematician can teach another about that, except to warn that as the size of the job increases, its complexity increases in frightening proportion. At one stage of writing a 300-page book, I had 1000 sheets of paper, each with a mathematical statement on it, a theorem, a lemma, or even a minor comment, complete with proof. The sheets were numbered, any which way. My job was to indicate on each sheet the numbers of the sheets whose statement must logically come before, and then to arrange the sheets in linear order so that no sheet comes after one on which it's mentioned. That problem had, apparently, uncountably many solutions; the difficulty was to pick one that was as efficient and pleasant as possible.

## 8. Write good English

Everything I've said so far has to do with writing in the large, global sense; it is time to turn to the local aspects of the subject.

Why shouldn't an author spell "continuous" as "continous"? There is no chance at all that it will be misunderstood, and it is one letter shorter, so why not? The answer that probably everyone would agree on, even the most libertarian among modern linguists, is that whenever the "reform" is introduced it is bound to cause distraction, and therefore a waste of time, and the "saving" is not worth it. A random example such as this one is probably not convincing; more people would agree that an entire book written in reformed spelling, for instance, "izi" for "easy" is not likely to be an effective teaching instrument for mathematics. Whatever the merits of spelling reform may be, words that are misspelled according to currently accepted dictionary standards detract from the good a book can do: they delay and distract the reader, and possibly confuse or anger him.

The reason for mentioning spelling is not that it is a common danger or a serious one for most authors, but that it serves to illustrate and emphasize a much more important point. I should like to argue that it is important that mathematical books (and papers, and letters, and lectures) be written in good English style, where good means "correct" according to currently and commonly accepted public standards. (French, Japanese, or Russian authors please substitute "French", "Japanese", or "Russian" for "English".) I do not mean that the style is to be pedantic, or heavy-handed, or formal, or bureaucratic, or flowery, or academic jargon. I do mean that it should be completely unobtrusive, like good background music for a movie, so that the reader may proceed with no conscious or unconscious blocks caused by the instrument of communication and not its content.

Good English style implies correct grammar, correct choice of words, correct punctuation, and, perhaps above all, common sense. There is a difference between "that" and "which", and "less" and "fewer" are not the same, and a good mathematical author must know such things. The reader may not be able to define the difference, but a hundred pages of colloquial misusage, or worse, has a cumulative abrasive effect that the author surely does not want to produce. Fowler [4], Roget [8], and Webster [10] are next to Dunford-Schwartz on my desk; they belong in a similar position on every author's desk. It is unlikely that a single missing comma will convert a correct proof into a wrong one, but consistent mistreatment of such small things has large effects.

The English language can be a beautiful and powerful instrument for interesting, clear, and completely precise information, and I have faith that the same is true for French or Japanese or Russian. It is just as important for an expositor to familiarize himself with that instrument as for a surgeon to know his tools. Euclid can be explained in bad grammar and bad diction, and a vermiform appendix can be removed with a rusty pocket knife, but the victim, even if he is unconscious of the reason for his discomfort, would surely prefer better treatment than that.

All mathematicians, even very young students very near the beginning of their mathematical learning, know that mathematics has a language of its own (in fact it is one), and an author must have thorough mastery of the grammar and vocabulary of that language as well as of the vernacular. There is no Berlitz course for the language of mathematics; apparently the only way to learn it is to live with it for years. What follows is not, it cannot be, a mathematical analogue of Fowler, Roget, and Webster, but it may perhaps serve to indicate a dozen or two of the thousands of items that those analogues would contain.

## 9. Honesty is the best policy

The purpose of using good mathematical language is, of course, to make the understanding of the subject easy for the reader, and perhaps even pleasant. The style should be good not in the sense of flashy brilliance, but good in the sense of perfect unobtrusiveness. The purpose is to smooth the reader's way, to anticipate his difficulties and to forestall them. Clarity is what's wanted, not pedantry; understanding, not fuss.

The emphasis in the preceding paragraph, while perhaps necessary, might seem to point in an undesirable direction, and I hasten to correct a possible misinterpretation. While avoiding pedantry and fuss, I do not want to avoid rigor and precision; I believe that these aims are reconcilable. I do not mean to advise a young author to be ever so slightly but very very cleverly dishonest and to gloss over difficulties. Sometimes, for instance, there may be no better way to get a result than a cumbersome computation. In that case it is the author's duty to carry it out, in public; the best he can do to alleviate it is to extend his sympathy to the reader by some phrase such as "unfortunately the only know proof is the following cumbersome computation".

Here is the sort of thing I mean by less than complete honesty. At a certain point, having proudly proved a proposition $p$, you feel moved to say: "Note, however, that $p$ does not imply $q$", and then, thinking that you've done a good expository job, go happily on to other things. Your motives may be perfectly pure, but the reader may feel cheated just the same. If he knew all about the subject, he wouldn't be reading you; for him the nonimplication is, quite likely, unsupported. Is it obvious? (Say so.) Will a counterexample be supplied later? (Promise it now.) Is it a standard but for present purposes irrelevant part of the literature? (Give a reference.) Or, horribile dictu, do you merely mean that you have tried to derive $q$ from $p$, you failed, and you don't in fact know whether $p$ implies $q$? (Confess immediately!) In any event: take the reader into your confidence.

There is nothing wrong with the often derided "obvious" and "easy to see", but there are certain minimal rules to their use. Surely when you wrote that something was obvious, you thought it was. When, a month, or two months, or six months later, you picked up the manuscript and re-read it, did you still think that that something was obvious? (A few months' ripening always improves manuscripts.) When you explained it to a friend, or to a seminar, was that something at issue accepted as obvious? (Or did someone question it and subside, muttering, when you reassured him? Did your assurance consist of demonstration or intimidation?) The obvious answers to these rhetorical questions are among the rules that should control the use of "obvious". There is another rule, the major one, and everybody knows it, the one whose violation is the most frequent source of mathematical error: make sure that the "obvious" is true.

It should go without saying that you are not setting out to hide facts from the reader; you are writing to uncover them. What I am saying now is that you should not hide the status of your statements and your attitude toward them either. Whenever you tell him something, tell him where it stands: this has been proved, that hasn't, this will be proved, that won't. Emphasize the important and minimize the trivial. There are many good reasons for making obvious statements every now and then; the reason for saying that they are obvious is to put them in proper perspective for the uninitiate. Even if your saying so makes an occasional reader angry at you, a good purpose is served by your telling him how you view the matter. But, of course, you must obey the rules. Don't let the reader down; he wants to believe in you. Pretentiousness, bluff, and concealment may not get caught out immediately, but most readers will soon sense that there is something wrong, and they will blame neither the facts nor themselves, but, quite properly, the author. Complete honesty makes for greatest clarity.

## 10. Down with the irrelevant and the trivial

Sometimes a proposition can be so obvious that it needn't even be called obvious and still the sentence that announces it is bad exposition, bad because it makes for confusion, misdirection, delay. I mean something like this: "If $R$ is a commutative semisimple ring with unit and if $x$ and $y$ are in $R$, then $x^2 - y^2 = (x-y)(x+y)$." The alert reader will ask himself what semisimplicity and a unit have to do with what he had always thought was obvious. Irrelevant assumptions wantonly dragged in, incorrect emphasis, or even just the absence of correct emphasis can wreak havoc.

Just as distracting as an irrelevant assumption and the cause of just as much wasted time is an author's failure to gain the reader's confidence by explicitly mentioning trivial cases and excluding them if need be. Every complex number is the product of a non-negative number and a number of modulus 1. That is true, but the reader will feel cheated and insecure if soon after first being told that fact (or being reminded of it on some other occasion, perhaps preparatory to a generalization being sprung on him) he is not told that there is something fishy about 0 (the trivial case). The point is not that failure to treat the trivial cases separately may sometimes be a mathematical error; I am not just saying "do not make mistakes". The point is that insistence on legalistically correct but insufficient explicit explanations ("The statement is correct as it stands—what else do you want?") is misleading, bad exposition, bad psychology. It may also be almost bad mathematics. If, for instance, the author is preparing to discuss the theorem that, under suitable hypotheses, every linear transformation is the product of a dilation and a rotation, then his ignoring of 0 in the 1-dimensional case leads to the reader's misunderstanding of the behavior of singular linear transformations in the general case.

This may be the right place to say a few words about the statements of theorem: there, more than anywhere else, irrelevancies must be avoided.

The first question is where the theorem should be stated, and my answer is: first. Don't ramble on in a leisurely way, not telling the reader where you are going, and then suddenly announce "Thus we have proved that . . . ". The reader can play closer attention to the proof if he knows what you are proving, and he can see better where the hypotheses are used if he knows in advance what they are. (The rambling approach frequently leads to the "hanging" theorem, which I think is ugly. I mean something like: "Thus we have proved

THEOREM 2 . . . ".

The indentation, which is after all a sort of invisible punctuation mark, makes a jarring separation in the sentence, and, after the reader has collected his wits and caught on to the trick that was played on him, it makes an undesirable separation between the statement of the theorem and its official label.)

This is not to say that the theorem is to appear with no introductory comments, preliminary definitions, and helpful motivations. All that comes first; the statement comes next; and the proof comes last. The statement of the theorem should consist of one sentence whenever possible: a simple implication, or, assuming that some universal hypotheses were stated before and are still in force, a simple declaration. Leave the chit-chat out: "Without loss of generality we may assume ..." and "Moreover it follows from Theorem 1 that ..." do not belong in the statement of a theorem.

Ideally the statement of a theorem is not only one sentence, but a short one at that. Theorems whose statement fills almost a whole page (or more!) are hard to absorb, harder than they should be; they indicate that the author did not think the material through and did not organize it as he should have done. A list of eight hypotheses (even if carefully so labelled) and a list of six conclusions do not a theorem make; they are a badly expounded theory. Are all the hypotheses needed for each conclusion? If the answer is no, the badness of the statement is evident. If the answer is yes, then the hypotheses probably describe a general concept that deserves to be isolated, named, and studied.

## 11. Do and do not repeat

One important rule of good mathematical style calls for repetition and another calls for its avoidance.

By repetition in the first sense I do not mean the saying of the same thing several times in different words. What I do mean, in the exposition of a precise subject such as mathematics, is the word-for-word repetition of a phrase, or even many phrases, with the purpose of emphasizing a slight change in a neighboring phrase. If you have defined something, or stated something, or proved something in Chapter 1, and if in Chapter 2 you want to treat a parallel theory or a more general one, it is a big help to the reader if you use the same words in the same order for as long as possible, and then, with a proper roll of drums, emphasize the difference. The roll of drums is important. It is not enough to list six adjectives in one definition, and re-list five of them, with a diminished sixth, in the second. That's the thing to do, but what helps is to say, in addition: "Note that the first five conditions in the definitions of $p$ and $q$ are the same; what makes them different is the weakening of the sixth."

Often in order to be able to make such an emphasis in Chapter 2 you'll have to go back to Chapter 1 and rewrite what you thought you had already written well enough, but this time so that its parallelism with the relevant part of Chapter 2 is brought out by the repetition device. This is another illustration of why the spiral plan of writing is unavoidable, and it is another aspects of what I call the organization of the material.

The preceding paragraphs describe an important kind of mathematical repetition, the good kind; there are two other kinds, which are bad.

One sense in which repetition is frequently regarded as a device of good teaching is that the oftener you say the same thing, in exactly the same words, or else with slight differences each time, the more likely you are to drive the point home. I disagree. The second time you say something, even the vaguest reader will dimly recall that there was a first time, and he'll wonder if what he is now learning is exactly the same as what he should have learned before, or just similar but different. (If you tell him "I am now saying *exactly* what I said on p. 3", that helps.) Even the dimmest such wonder is bad. Anything is bad that unnecessarily frightens, irrelevantly amuses, or in any other way distracts. (Unintended double meanings are the woe of many an author's life.) Besides, good organization, and, in particular, the spiral plan of organization discussed before is a substitute for repetition, a substitute that works much better.

Another sense in which repetition is bad is summed up in the short and only partially inaccurate precept: never repeat a proof. If several steps in the proof of Theorem 2 bear a very close resemblance to parts of the proof of Theorem 1, that's a signal that something may be less than completely understood. Other symptoms of the disease are: "by the same technique (or method, or device, or trick) as in the proof of Theorem 1 ...", or, brutally, "see the proof of Theorem 1". When that happens the chances are very good that there is a lemma that is worth finding, formulating, and proving, a lemma from which both Theorem 1 and Theorem 2 are more easily and more clearly deduced.

## 12. THE EDITORIAL WE IS NOT ALL BAD

One aspect of expository style that frequently bothers beginning authors is the use of the editorial "we", as opposed to the singular "I", or the neutral "one". It is in matters like this that common sense is most important. For what it's worth, I present here my recommendation.

Since the best expository style is the least obtrusive one, I tend nowadays to prefer the neutral approach. That does *not* mean using "one" often, or ever; sentences like "one has thus proved that ..." are awful. It does mean the complete avoidance of first person pronouns in either singular or plural. "Since $p$, it follows that $q$." "This implies $p$." "An application of $p$ to $q$ yields $r$." Most (all ?) mathematical writing is (should be ?) factual; simple declarative sentences are the best for communicating facts.

A frequently effective and time-saving device is the use of the imperative. "To find $p$, multiply $q$ by $r$." "Given $p$, put $q$ equal to $r$." (Two digressions about "given". (1) Do not use it when it means nothing. Example: "For any given $p$ there is a $q$." (2) Remember that it comes from an active verb and resist the temptation to leave it dangling. Example: Not "Given $p$, there is a $q$", but "Given $p$, find $q$".)

There is nothing wrong with the editorial "we", but if you like it, do not misuse it. Let "we" mean "the author and the reader" (or "the lecturer and the audience"). Thus, it is fine to say "Using Lemma 2 we can generalize Theorem 1", or "Lemma 3 gives us a technique for proving Theorem 4". It is not good to say "Our work on this result was done in 1969" (unless the voice is that of two authors, or more, speaking in unison), and "We thank our wife for her help with the typing" is always bad.

The use of "I", and especially its overuse, sometimes has a repellent effect, as arrogance or ex-cathedra preaching, and, for that reason, I like to avoid it whenever possible. In short notes, obviously in personal historical remarks, and, perhaps, in essays such as this, it has its place.

## 13. Use words correctly

The next smallest units of communication, after the whole concept, the major chapters, the paragraphs, and the sentences are the words. The preceding section about pronouns was about words, in a sense, although, in a more legitimate sense, it was about global stylistic policy. What I am now going to say is not just "use words correctly"; that should go without saying. What I do mean to emphasize is the need to think about and use with care the small words of common sense and intuitive logic, and the specifically mathematical words (technical terms) that can have a profound effect on mathematical meaning.

The general rule is to use the words of logic and mathematics correctly.

The emphasis, as in the case of sentence-writing, is not encouraging pedantry; I am not suggesting a proliferation of technical terms with hairline distinctions among them. Just the opposite; the emphasis is on craftsmanship so meticulous that it is not only correct, but unobtrusively so.

Here is a sample: "Prove that any complex number is the product of a non-negative number and a number of modulus 1." I have had students who would have offered the following proof: "$-4i$ is a complex number, and it is the product of 4, which is non-negative, and $-i$, which has modulus 1; q.e.d." The point is that in everyday English "any" is an ambiguous word; depending on context it may hint at an existential quantifier ("have you any wool?", "if anyone can do it, he can") or a universal one ("any number can play"). Conclusion: never use "any" in mathematical writing. Replace it by "each" or "every", or recast the whole sentence.

One way to recast the sample sentence of the preceding paragraph is to establish the convention that all "individual variables" range of the set of complex numbers and then write something like

$$\forall z \exists p \exists u\,[(p = |p|) \wedge (|u| = 1) \wedge (z = pu)].$$

I recommend against it. The symbolism of formal logic is indispensable in the discussion of the logic of mathematics, but used as a means of transmitting ideas from one mortal to another it becomes a cumbersome code. The author had to code his thoughts in it (I deny that anybody thinks in terms of $\exists$, $\forall$, $\wedge$, and the like), and the reader has to decode what the author wrote; both steps are a waste of time and an obstruction to understanding. Symbolic presentation, in the sense of either the modern logician or the classical epsilontist, is something that machines can write and few but machines can read.

So much for "any". Other offenders, charged with lesser crimes, are "where", and "equivalent", and "if ... then ... if ... then". "Where" is usually a sign of a lazy afterthought that should have been thought through before. "If $n$ is sufficiently large, then $|a_n| < \varepsilon$, where $\varepsilon$ is a preassigned positive number"; both disease and cure are clear. "Equivalent" *for theorems* is logical nonsense. (By "theorem" I mean a mathematical truth, something that has been proved. A meaningful statement can be false, but a theorem cannot; "a false theorem" is self-contradictory). What sense does is make to say that the completeness of $L^2$ is equivalent to the representation theorem for linear functionals on $L^2$? What is meant is that the proofs of both theorems are moderately hard, but once one of them has been proved, either one, the other can be proved with relatively much less work. The logically precise word "equivalent" is not a good word for *that*. As for "if ... then ... if ... then", that is just a frequent stylistic bobble committed by quick writers and rued by slow readers. "If $p$, then if $q$, then $r$." Logically all is well $(p \Rightarrow (q \Rightarrow r))$, but psychologically it is just another pebble to stumble over, unnecessarily. Usually all that is needed to avoid it is to recast the sentence, but no universally good recasting exists; what is best depends on what is important in the case at hand. It could be "If $p$ and $q$, then $r$", or "In the presence of $p$, the hypothesis $q$ implies the conclusion $r$", or many other versions.

## 14. USE TECHNICAL TERMS CORRECTLY

The examples of mathematical diction mentioned so far were really logical matters. To illustrate the possibilities of the unobtrusive use of precise language in the everyday sense of the working mathematician, I briefly mention three examples: function, sequence, and contain.

I belong to the school that believes that functions and their values are sufficiently different that the distinction should be maintained. No fuss is necessary, or at least no visible, public fuss; just refrain from saying things like "the function $z^2 + 1$ is even". It takes a little longer to say "the function $f$ defined by $f(z) = z^2 + 1$ is even", or, what is from many points of view preferable, "the function $z \to z^2 + 1$ is even", but it is a good habit that can sometimes save the reader (and the author) from serious blunder and that always makes for smoother reading.

"Sequence" means "function whose domain is the set of natural numbers". When an author writes "the union of a sequence of measurable sets is measurable" he is guiding the reader's attention to where it doesn't belong. The theorem has nothing to do with the firstness of the first set, the secondness of the second, and so on; the *sequence* is irrelevant. The correct statement is that "the union of a countable set of measurable sets is measurable" (or, if a different emphasis is wanted, "the union of a countably infinite set of measurable sets is measurable"). The theorem that "the limit of a sequence of measurable functions is measurable" is a very different thing; there "sequence" is correctly used. If a reader knows what a sequence is, if he feels the definition in his bones, then the misuse of the word will distract him and slow his reading down, if ever so slightly; if he doesn't really know, then the misuse will seriously postpone his ultimate understanding.

"Contain" and "include" are almost always used as synonyms, often by the same people who carefully coach their students that $\in$ and $\subset$ are not the same thing at all. It is extremely unlikely that the interchangeable use of contain and include will lead to confusion. Still, some years ago I started an experiment, and I am still trying it: I have systematically and always, in spoken word and written, used "contain" for $\in$ and "include" for $\subset$. I don't say that I have proved anything by this, but I can report that (a) it is very easy to get used to, (b) it does no harm whatever, and (c) I don't think that anybody ever noticed it. I suspect, but that is not likely to be provable, that this kind of terminological consistency (with no fuss made about it) might nevertheless contribute to the reader's (and listener's) comfort.

Consistency, by the way, is a major virtue and its opposite is a cardinal sin in exposition. Consistency is important in language, in notation, in reference, in typography—it is important everywhere, and its absence can cause anything from mild irritation to severe misinformation.

My advice about the use of words can be summed up as follows. (1) Avoid technical terms, and especially the creation of new ones, whenever possible. (2) Think hard about the new ones you must create; consult Roget; and make them as appropriate as possible. (3) Use the old ones correctly and consistently, but with a minimum of obtrusive pedantry.

## 15. RESIST SYMBOLS

Everything said about words applies, mutatis mutandis, to the even smaller units of mathematical writing, the mathematical symbols. The best notation is no notation; whenever it is possible to avoid the use of a complicated alphabetical apparatus, avoid it. A good attitude to the preparation of written mathematical exposition is to pretend that it is spoken. Pretend that you are explaining the subject to a friend on a long walk in the woods, with no paper available; fall back on symbolism only when it is really necessary.

A corollary to the principle that the less there is of notation the better it is, and in analogy with the principle of omitting irrelevant assumptions, avoid the use of irrelevant symbols. Example: "On a compact space every real-valued continuous function $f$ is bounded." What does the symbol "$f$" contribute to the clarity of that statement? Another example: "If $0 \le \lim_n \alpha_n^{1/n} = \rho < 1$, then $\lim_n \alpha_n = 0$." What does "$\rho$" contribute here? The answer is the same in both cases (nothing), but the reasons for the presence of the irrelevant symbols may be different. In the first case "$f$" may be just a nervous habit; in the second case "$\rho$" is probably a preparation for the proof. The nervous habit is easy to break. The other is harder, because it involves more work for the author. Without the "$\rho$" in the statement, the proof will take a half line longer; it will have to begin with something like "Write $\rho = \lim_n \alpha_n^{1/n}$." The repetition (of "$\lim_n \alpha_n^{1/n}$") is worth the trouble; both statement and proof read more easily and more naturally.

A showy way to say "use no superfluous letters" is to say "use no letter only once". What I am referring to here is what logicians would express by saying "leave no variable free". In the example above, the one about continuous functions, "$f$" was a free variable. The best way to eliminate that particular "$f$" is to omit it; an occasionally preferable alternative is to convert it from free to bound. Most mathematicians would do that by saying "If $f$ is a real-valued continuous function on a compact space, then $f$ is bounded." Some logicians would insist on pointing out that "$f$" is still free in the new sentence (twice), and technically they would be right. To make it bound, it would be necessary to insert "for all $f$" at some grammatically appropriate point, but the customary way mathematicians handle the problem is to refer (tacitly) to the (tacit) convention that every sentence is preceded by all the universal quantifiers that are needed to convert all its variables into bound ones.

The rule of never leaving a free variable in a sentence, like many of the rules I am stating, is sometimes better to break than to obey. The sentence, after all, is an arbitrary unit, and if you want a free "$f$" dangling in one sentence so that you may refer to it in a later sentence in, say, the same paragraph, I don't think you should necessarily be drummed out of the regiment. The rule is essentially sound, just the same, and while it may be bent sometimes, it does not deserve to be shattered into smithereens.

There are other symbolic logical hairs that can lead to obfuscation, or, at best, temporary bewilderment, unless they are carefully split. Suppose, for an example, that somewhere you have displayed the relation

$$\int_0^1 |f(x)|^2\,dx < \infty, \tag{*}$$

as, say, a theorem proved about some particular $f$. If, later you run across another function $g$ with what looks like the same property, you should resist the temptation to say "$g$ also satisfies (*)". That is logical and alphabetical nonsense. Say instead "(*) remains satisfied if $f$ is replaced by $g$", or, better, give (*) a name (in this case it has a customary one) and say "$g$ also belongs to $L^2(0,1)$".

What about "inequality (*)", or "equation (7)", or "formula (iii)"; should all displays be labelled or numbered? My answer is no. Reason: just as you shouldn't mention irrelevant assumptions or name irrelevant concepts, you also shouldn't attach irrelevant labels. Some small part of the reader's attention is attracted to the label, and some small part of his mind will wonder why the label is there. If there is a reason, then the wonder serves a healthy purpose by way of preparation, with no fuss, for a future reference to the same idea; if there is no reason, then the attention and the wonder were wasted.

It's good to be stingy in the use of labels, but parsimony can also be carried to extremes. I do not recommend that you do what Dickson once did [2]. On p. 89 he says: "Then ... we have (1) ..."—but p. 89 is the beginning of a new chapter, and happens to contain no display at all, let alone one bearing the label (1). The display labelled (1) occurs on p. 90, overleaf, and I never thought of looking for it *there*. That trick gave me a helpless and bewildered five minutes. When I finally saw the light, I felt both stupid and cheated, and I have never forgiven Dickson.

One place when cumbersome notation quite often enters is in mathematical induction. Sometimes it is unavoidable. More often, however, I think that indicating the step from 1 to 2 and following it by an airy "and so on" is as rigorously unexceptionable as the detailed computation, and much more understandable and convincing. Similarly, a general statement about $n \times n$ matrices is frequently best proved not by the exhibition of many $a_{ij}$'s, accompanied by triples of dots laid out in rows and columns and diagonals, but by the proof of a typical (say $3 \times 3$) special case.

There is a pattern in all these injunctions about the avoidance of notation. The point is that the rigorous concept of a mathematical proof can be taught to a stupid computing machine in one way only, but to a human being endowed with geometric intuition, with daily increasing experience and with the impatient inability to concentrate on repetitious detail for very long, that way is a bad way. Another illustration of this is a proof that consists of a chain of expressions separated by equal signs. Such a proof is easy to write. The author starts from the first equation, makes a natural substitution to get the second, collect terms, permutes, inserts and immediately cancels an inspired factor, and by steps such as these proceeds till he gets the last equation. This is, once again, coding, and the reader is forced not only to learn as he goes, but, at the same time, to decode as he goes. The double effort is needless. By spending another ten minutes writing a carefully worded paragraph, the author can save each of his readers half an hour and a lot of confusion. The paragraph should be a recipe for action, to replace the unhelpful code that merely reports the results of the act and leaves the reader to guess how they were obtained. The paragraph would say something like this: "For the proof, first substitute $p$ for $q$, then collect terms, permute the factors, and finally, insert and cancel a factor $r$."

A familiar trick of bad teaching is to begin a proof by saying: "Given $\varepsilon$, let $\delta$ be $\left(\frac{\varepsilon}{3M^2+2}\right)^{1/2}$." This is the traditional backward proof-writing of classical analysis. It has the advantage of being easily *verifiable* by a machine (as opposed to *understandable* by a human being), and it has the dubious advantage that something in the end comes out to be less than $\varepsilon$, instead of less than, say, $\left(\frac{(3M^2+7)\varepsilon}{24}\right)^{1/3}$. The way to make the human reader's task less demanding is obvious: write the proof forward. Start as the author always starts, by putting something less than $\varepsilon$, and then do what needs to be done—multiply by $3M^2+7$ at the right time and divide by 24 later, etc., etc.—till you end up with what you end up with. Neither arrangement is elegant, but the forward one is graspable and rememberable.

## 16. Use symbols correctly

There is not much harm that can be done with non-alphabetical symbols, but there too consistency is good and so is the avoidance of individually unnoticed but collectively abrasive abuses. Thus, for instance, it is good to use a symbol so consistently that its verbal translation is always the same. It is good, but is is probably impossible; nonetheless it's a better aim than no aim at all. How are you to read "$\in$": as the verb phrase "is in" or as the preposition "in"? Is it correct to say: "For $x \in A$, we have $x \in B$," or "If $x \in A$, then $x \in B$"? I strongly prefer the latter (always read "$\in$" as "is in") and I doubly deplore the former (both usages occur in the same sentence). It's easy to write and it's easy to read "For $x$ in $A$, we have $x \in B$"; all dissonance and all even momentary ambiguity is avoided. The same is true for "$\subset$" even though the verbal translation is longer, and even more true for "$\leq$". A sentence such as "Whenever a positive number is $\leq 3$, its square is $\leq 9$" is ugly.

Not only paragraphs, sentences, words, letters and mathematical symbols, but even the innocent looking symbols of standard prose can be the source of blemishes and misunderstandings; I refer to punctuation marks. A couple of examples will suffice. First: an equation, or inequality, or inclusion, or any other mathematical clause is, in its informative content, equivalent to a clause in ordinary language, and, therefore, it demands just as much to be separated from its neighbors. In other words: punctuate symbolic sentences just as you would verbal ones. Second: don't overwork a small punctuation mark such as a period or a comma. They are easy for the reader to overlook, and the oversight causes backtracking, confusion, delay. Example: "Assume that $a \in X$. $X$ belongs to the class $C$, ...". The period between the two $X$'s is overworked, and so is this one: "Assume that $X$ vanishes. $X$ belongs to the class $C$, ...". A good general rule is: never start a sentence with a symbol. If you insist on starting the sentence with a mention of the thing the symbol denotes, put the appropriate word in apposition, thus: "The set $X$ belongs to the class $C$, ...".

The overworked period is no worse than the overworked comma. Not "For invertible $X$, $X^*$ is also invertible", but "For invertible $X$, the adjoint $X^*$ is also invertible". Similarly, not "Since $p \neq 0$, $p \in U$", but "Since $p \neq 0$, it follows that $p \in U$". Even the ordinary "If you don't like it, lump it" (or, rather, its mathematical relatives) is harder to digest than the stuffy-sounding "If you don't like it, then lump it"; I recommend the "then" with "if" in all mathematical contexts. The presence of "then" can never confuse; its absence can.

A final technicality that can serve as an expository aid, and should be mentioned here, is in a sense smaller than even the punctuation marks, it is in a sense so small that it is invisible, and yet, in another sense, it's the most conspicuous aspect of the printed page. What I am talking about is the layout, the architecture, the appearance of the page itself, of all the pages. Experience with writing, or perhaps even with fully conscious and critical reading, should give you a feeling for how what you are now writing will look when it's printed. If it looks like solid prose, it will have a forbidding, sermony aspect; if it looks like computational hash, with a page full of symbols, it will have a frightening, complicated aspect. The golden mean is golden. Break it up, but not too small; use prose, but not too much. Intersperse enough displays to give the eye a chance to help the brain; use symbols, but in the middle of enough prose to keep the mind from drowning in a morass of suffixes.

## 17. All communication is exposition

I said before, and I'd like for emphasis to say it again, that the differences among books, articles, lectures, and letters (and whatever other means of communication you can think of) are smaller than the similarities.

When you are writing a research paper, the role of the "slips of paper" out of which a book outline can be constructed might be played by the theorems and the proofs that you have discovered; but the game of solitaire that you have to play with them is the same.

A lecture is a little different. In the beginning a lecture is an expository paper; you plan it and write it the same way. The difference is that you must keep the difficulties of oral presentation in mind. The reader of a book can let his attention wander, and later, when he decides to, he can pick up the thread, with nothing lost except his own time; a member of the lecture audience cannot do that. The reader can try to prove your theorems for himself, and use you exposition as a check on his work; the hearer cannot do that. The reader's attention span is short enough; the hearer's is much shorter. If computations are unavoidable, a reader can be subjected to them; a hearer must never be. Half of the art of good writing is the art of omission; in speaking, the art of omission is nine-tenths of the trick. These differences are not large. To be sure, even a good expository paper, read out loud, would make an awful lecture—but not worse than some I have heard.

The appearance of the printed page is replaced, for a lecture, by the appearance of the blackboard, and the author's imagined audience is replaced for the lecturer by live people; these are big differences. As for the blackboard: it provides the opportunity to make something grow and come alive in a way that is not possible with the printed page. (Lecturers who prepare a blackboard, cramming it full before they start speaking, are unwise and unkind to audiences.) As for live people: they provide an immediate feedback that every author dreams about but can never have.

The basic problems of all expository communication are the same; they are the ones I have been describing in this essay. Content, aim and organization, plus the vitally important details of grammar, diction, and notation—they, not showmanship, are the essential ingredients of good lectures, as well as good books.

## 18. Defend your style

Smooth, consistent, effective communication has enemies; they are called editorial assistants or copyreaders.

An *editor* can be a very great help to a writer. Mathematical writers must usually live without this help, because the editor of a mathematical book must be a mathematician, and there are very few mathematical editors. The ideal editor, who must potentially understand every detail of the author's subject, can give the author an inside but nonetheless unbiased view of the work that the author himself cannot have. The ideal editor is the union of the friend, wife, student, and expert junior-grade whose contribution to writing I described earlier. The mathematical editors of book series and journals don't even come near the ideal. Their editorial work is but a small fraction of their life, whereas to be a good editor is a full-time job. The ideal mathematical editor does not exist; the friend-wife-etc. combination is only an almost ideal substitute.

The *editorial assistant* is a full-time worker whose job is to catch your inconsistencies, your grammatical slips, your errors in diction, your misspellings—everything that you can do wrong, short of mathematical content. The trouble is that the editorial assistant does not regard himself as an extension of the author, and he usually degenerates into a mechanical misapplier of mechanical rules. Let me give some examples.

I once studied certain transformations called "measure-preserving". (Note the hyphen: it plays an important role, by making a single word, an adjective, out of two words.) Some transformations pertinent to that study failed to deserve the name; their failure was indicated, of course, by the prefix "non". After a long sequence of misunderstood instructions, the printed version spoke of a "nonmeasure preserving transformation". That is nonsense, of course, amusing nonsense, but, as such, it is distracting and confusing nonsense.

A mathematician friend reports that in the manuscript of a book of his he wrote something like "$p$ or $q$ holds according as $x$ is negative or positive". The editorial assistant changed that to "$p$ or $q$ holds according as $x$ is positive or negative", on the grounds that it sounds better that way. That could be funny if it weren't sad, and, of course, very very wrong.

A common complaint of anyone who has ever discussed quotation marks with the enemy concerns their relation to other punctuation. There appears to be an international typographical decree according to which a period or comma immediately to the right of a quotation is "ugly". (As here: the editorial assistant would have changed that to "ugly." if I had let him.) From the point of view of the logical mathematician (and even more the mathematical logician) the decree makes no sense; the comma or period should come where the logic of the situation forces it to come. Thus,

He said: "The comma is ugly."

Here, clearly, the period belongs inside the quote; the two situations are different and no inelastic rule can apply to both.

Moral: there are books on "style" (which frequently means typographical conventions), but their mechanical application by editorial assistants can be harmful. If you want to be an author, you must be prepared to defend your style; go forearmed into the battle.

## 19. Stop

The battle against copyreaders is the author's last task, but it's not the one most authors regard as the last. The subjectively last step comes just before; it is to finish the book itself—to stop writing. That's hard.

There is always something left undone, always either something more to say, or a better way to say something, or, at the very least, a disturbing vague sense that the perfect addition or improvement is just around the corner, and the dread that its omission would be everlasting cause for regret. Even as I write this, I regret that I did not include a paragraph or two on the relevance of euphony and prosody to mathematical exposition. Or, hold on a minute!, surely I cannot stop without a discourse on the proper naming of concepts (why "commutator" is good and "set of first category" is bad) and the proper way to baptize theorems (why "the closed graph theorem" is good and "the Cauchy-Buniakowski-Schwarz theorem" is bad). And what about that sermonette that I haven't been able to phrase satisfactorily about following a model. Choose someone, I was going to say, whose writing can touch you and teach you, and adapt and modify his style to fit to your personality and your subject—surely I must get that said somehow.

There is no solution to this problem except the obvious one; the only way to stop is to be ruthless about it. You can postpone the agony a bit, and you should do so, by proofreading, by checking the computations, by letting the manuscript ripen, and then by reading the whole thing over in a gulp, but you won't want to stop any more then than before.

When you've written everything you can think of, take a day or two to read over the manuscript quickly and to test it for the obvious major points that would first strike a stranger's eye. Is the mathematics good, is the exposition interesting, is the language clear, is the format pleasant and easy to read? Then proofread and check the computations; that's an obvious piece of advice, and no one needs to be told how to do it. "Ripening" is easy to explain but not always easy to do: it means to put the manuscript out of sight and try to forget it for a few months. When you have done all that, and then re-read the whole work from a rested point of view, you have done all you can. Don't wait and hope for one more result, and don't keep on polishing. Even if you do get that result or do remove that sharp corner, you'll only discover another mirage just ahead.

To sum it up: begin at the beginning, go on till you come to the end, and then, with no further ado, stop.

## 20. The last word

I have come to the end of all the advice on mathematical writing that I can compress into one essay. The recommendations I have been making are based partly on what I do, more on what I regret not having done, and most on what I wish others had done for me. You may criticize what I've said on many grounds, but I ask that a comparison of my present advice with my past action not be one of them. Do, please, as I say, and not as I do, and you'll do better. Then rewrite this essay and tell the next generation how to do better still.

## References

[1] Birkhoff, G. D. Proof of the ergodic theorem, *Proc. N.A.S., U.S.A* 17 (1931) 656-660.

[2] Dickson, L. E., *Modern algebraic theories,* Sanborn, Chicago (1926).

[3] Dunford N. and Schwartz J. T., *Linear operators,* Interscience, New York (1958, 1963).

[4] Fowler H. W., *Modern English usage* (Second edition, revised by Sir Ernest Gowers), Oxford, New York (1965).

[5] Heisel C. T., *The circle squared beyond refutation,* Heisel, Cleveland (1934).

[6] Lefschetz, S. Algebraic Topology, *A.M.S.,* New York (1942).

[7] Nelson E. A proof of Liouville's theorem, *Proc. A.M.S.,* 12 (1961) 995.

[8] *Roget's International Thesaurus,* Crowell, New York (1946).

[9] Thurber J. and Nugent E., *The male animal,* Random House, New York (1940).

[10] *Webster's New International Dictionary* (Second edition, unabridged), Merriam, Springfield (1951).

Indiana University