How to write mathematics

Norman E. Steenrod
Paul R. Halmos
Menahem M. Schiffer
Jean A. Dieudonné

How to write mathematics

Norman E. Steenrod
Paul R. Halmos
Menahem M. Schiffer
Jean A. Dieudonné

AMERICAN MATHEMATICAL SOCIETY

Library of Congress Cataloging in Publication Data
Main entry under title:

How to write mathematics.

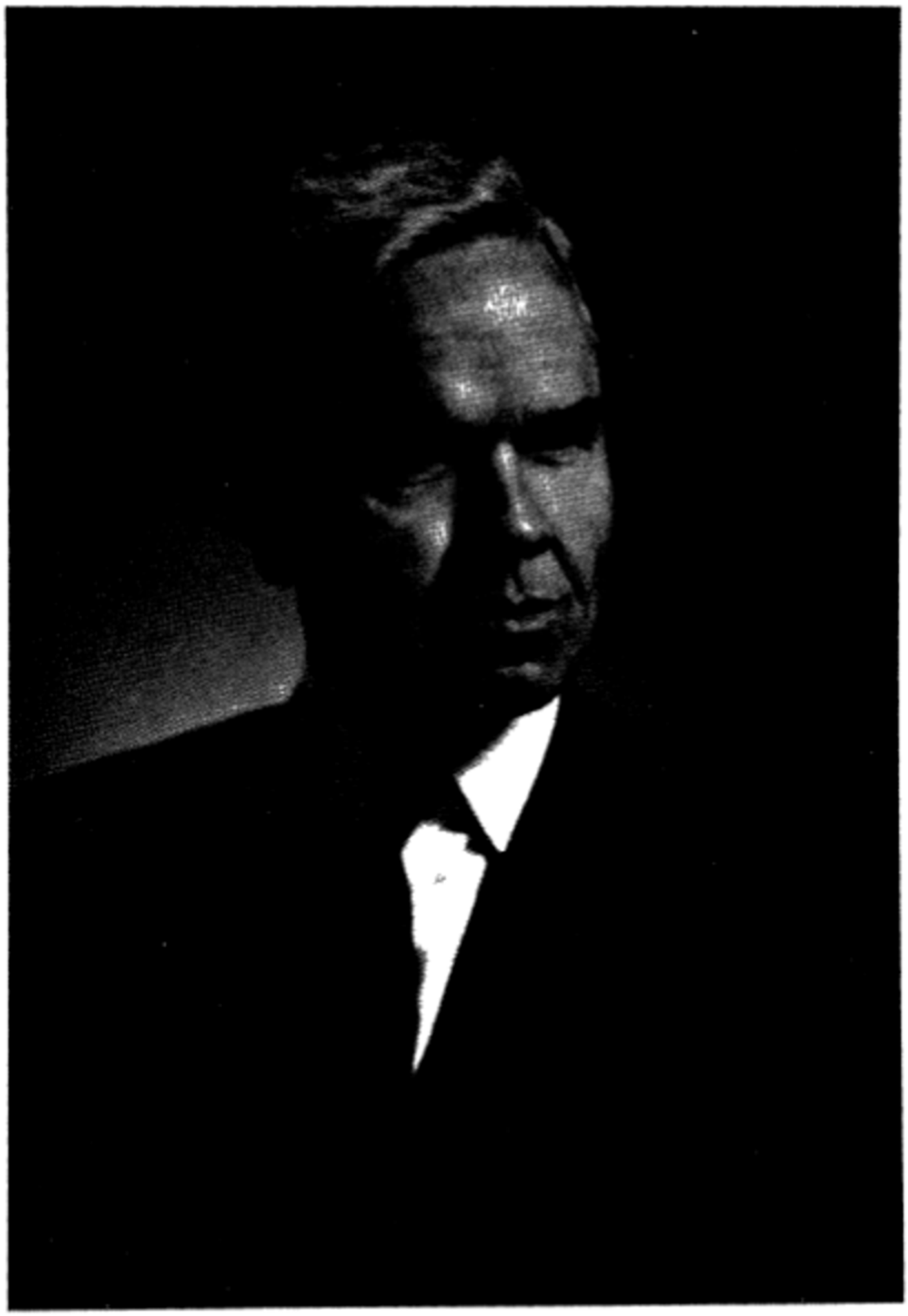
1. Mathematics--Authorship. I. Steenrod, Norman Earl, 1910–1971.

QA41.H6 808'.066'51021 72–13840

ISBN 0-8218-0055-8

CIP

Copyright © 1973 by the American Mathematical Society
Reprinted 1975 in the United States of America



Photograph by Orren Jack Turner

Norman E. Steenrod (1910-1971)

The Council and Board of Trustees of the American Mathematical Society dedicate this book to Norman E. Steenrod (1910-1971)

Report of the Committee on Expository Writing

The committee was authorized by the Council of the American Mathematical Society in August 1968; the last appointment to it was made by Oscar Zariski, then president, in March 1969. The charge was to prepare "a pamphlet on expository writing of books and papers at the research level and at the level of graduate texts".

In May 1969, two months after the committee was completed, one of its members resigned. He said he thought the project was too interesting to leave to a committee, which would never get it done properly, and he said he wanted to be free to write and publish his version independently. Norman Steenrod (the chairman) declined to accept the resignation, preferring to allow the member the freedom he sought. This left the exact membership of the committee up in the air.

The work of the committee proceeded mainly on Steenrod's steam; he wrote to the other members (in triplicate), and occasionally they would write an answer (to him alone). The committee met only once (for an hour, at the Eugene meeting in August 1969, with three present). The result of the correspondence and the meeting was the decision to present to the Council, as the product of the committee, four separate essays, one by each of the four members, with the recommendation that the Society publish them, together, as this book.

A year later (in August 1970) Steenrod had at hand only one essay. A year and six months later (in March 1971) that essay was published. (L'Enseignement Mathématique, 16 (1970), 123-152.) Even so, Steenrod was still hoping; he set August 30, 1971 as a target date for the receipt of all the essays. The solution he proposed for the problem created by the already published essay was to reprint it as is, as part of the AMS publication, provided the editors and publishers of L'Enseignement Mathématique agreed. They did.

Steenrod died in October 1971, before quite completing his own essay. Before he died he asked, through his wife, that his nearly finished work be prepared for submission to the council and presented together with the others. That was done.

Respectfully submitted,

J. A. Dieudonné
P. R. Halmos
M. M. Schiffer

This is the report of the committee to the Council, edited to serve as an introduction to the volume subsequently authorized by the Council and the Board of Trustees of the Society.

Norman E. Steenrod [1]

Introduction

Nearly all my comments will be aimed at the problems of exposition nvolved in the writing of a book, either a research monograph or a ext suitable for graduate study. Most of these comments apply also o expository articles at the research level because such articles are requently research monographs with all difficult proofs deleted.

A major objection to laying down criteria for the excellence of an xposition is that the effectiveness of an expository effort depends o heavily on the knowledge and experience of the reader. A clean nd exquisitely precise demonstration to one reader is a bore to another who has seen the like elsewhere. The same reader can find one part tediously clear and another part mystifying even though the author believed he gave both parts equally detailed treatment. Faced with these well-known facts, one tends to abandon the effort of seeking criteria, leaving it to the personal preferences of authors and readers to determine the outcome. In contrast to this attitude of hopelessness are the facts that many writers seem to agree on a number of aspects of style, and that a few writers have achieved a degree of general acclaim for their expository skill. Surely it must be possible to formulate several general principles to explain and justify these facts.

In this endeavor, I shall need to distinguish sharply two parts of a mathematical presentation: the *formal* or *logical* structure consisting of definitions, theorems, and proofs, and the complementary *informal* or *introductory* material consisting of motivations, analogies, examples, and metamathematical explanations. This division of the material should be conspicuously maintained in any mathematical presentation, because the nature of the subject requires above all else that the logical structure be clear. A reader who has become

Copyright © 1973 American Mathematical Society

[1] The last seven paragraphs of this essay (indicated by (*)) were not completed by the author when he died; they are taken from a preliminary outline he prepared. The ideas and their order are his throughout; the only changes are of a minor editorial kind affecting only a small number of words.

aware of a misunderstanding must be able to locate readily the precise step where he has not followed the author's reasoning.

Although the primary purpose of a book is to present the formal structure, a secondary purpose, almost as important, is to offer the reader a method whereby he can fit the new structure into what he already knows, and retain it as part of his working equipment. It is here that authors exhibit greatest variations in skill and art. An author needs to be aware of how he fits the structure into his own pattern of knowledge, and how others do so or might do so. What are the basic questions that will be answered? What are the crucial examples that motivate the development? What are the vaguely formulated principles from which the entire theory seems to unroll effortlessly? In supplying answers to these questions an author's taste and philosophy play a dominant role.

When we write *about* mathematics instead of *doing* it, we face an ever-present danger of saying something nonsensical or even fatuous. The fear of this tends to inhibit many authors, and some are so fearful that they hide behind the formal structure. Moreover the reactions of readers to the informal aspects of an exposition vary greatly. There is the reader whose attitude is the completely antiseptic "Show me your mathematics, I'll supply my own philosophy"; and there is his opposite who, when presented with a formal and dry mathematical system, promptly falls asleep. How can an author write so as to appeal to such diverse readers?

I contend that it is possible if he maintains the distinction between formal and informal material. He must strive throughout to describe his own attitudes towards the various parts of the subject, and also such other views as he regards valid, but all such material must be labelled as distinct from the formal structure so that a reader can omit or skim such parts as are not to his taste.

Since the formal structure does not depend on the informal, the author can write up the former in complete detail before adding any of the latter. This procedure is advantageous in reducing the amount of wasted effort caused by revisions of the formal structure. Many authors of mathematical books complain about the large amount of rewriting and re-rewriting that seems necessary to bring a book to final form. It is my experience that most of this is caused by the author becoming aware of defects or mistakes in his projected formal structure, and then discovering improvements that enforce re-

organization. By postponing the writing of informal material, one saves the writing of explanations of why things are done in certain ways when in fact they are ultimately not done that way. A difficulty with such postponement is that inspiration for the writing of informal parts comes frequently during the writing of the formal structure, and the pain of writing being what it is, inspiration should be given full sway. The answer of course is to make notes of ideas about the informal material while writing the formal structure.

EXPOSITORY PROBLEMS OF THE FORMAL STRUCTURE

In this section I shall discuss problems that an author faces in writing out the formal structure of the mathematics. The purpose of this somewhat disjointed set of comments is to sharpen awareness of the problems that an author will encounter, and to suggest approaches for dealing with them.

The main problem is the choice of the global organization of the mathematics. The number of possibilities for a given body of propositions is large. One can start from a known area, and build a new structure in an entirely constructive way. Different constructions can lead to isomorphic systems. One can start with a system of axioms, deduce a theory, and, at the end, prove consistency by a construction. Quite different axiomatizations can be equivalent.

Homology theory provides a good example of the variety of possible approaches. Forty years ago a course on the subject would begin with the homology of finite complexes based on incidence numbers, and, by the end of the semester, the groups would be proved to be topologically invariant. Twenty years ago the singular homology groups of a space were defined at the start of the course (these are obviously topologically invariant), the axioms for homology theory were then verified, and, by the end of the semester, one deduced from the axioms how to compute the groups of a finite complex. Ten years later homotopy notions reigned supreme, and one could define the nth cohomology groups of a space as the group of homotopy classes of maps of the space into the nth Eilenberg-Mac Lane space.

In deciding which organization is best, one can apply any of the following criteria: (1) length (the less work the better), (2) the quickness with which one obtains major or interesting results, (3) the simplicity of the start, and the gradualness of the approach to

difficulties, (4) the quickness with which examples and intuitive material can be developed, and (5) aesthetic satisfaction (the ease with which the development is motivated by vaguely-formulated principles).

By the time an author decides to write a book, he will already have chosen a global organization, in rough outline at least. His main objective in writing the book is to convince himself and the mathematical world that he has found a good way of doing things. To the extent that this is the case, there is little point to my recommending that he consider the virtues of other organizations. However, I do urge that he consider modifying his organization so that he can discuss other approaches, make comparisons, and establish their validity. To clarify this recommendation, let me take an example in algebraic topology. Suppose the basic approach to homology theory is a classical one. At some stage Hopf's theorem on the homotopy classification of mappings of an n-complex into the n-sphere should be proved. Once this is done, one can prove readily that the nth cohomology group of a complex is isomorphic to the group of homotopy classes of maps of the complex into the nth Eilenberg-Mac Lane space. This should be followed by the remark that this result shows that a psychologically different approach to homology theory can be based on homotopy theory. Subsequently, in exercises, the reader can be asked to derive familiar properties of cohomology groups from this characterization.

Another important problem involving the global organization is to decide the degree of generality to be sustained. Shall the results be proved for continuous functions only or for functions in L^2? Shall we restrict ourselves to separable locally-compact spaces, or to paracompact spaces? Quite often, the restriction of the basic category to a smaller one makes proofs become substantially shorter, and this may entail no loss of important applications. There is no general solution of this problem, the author must weigh the gain in generality against the cost of longer and less transparent proofs. I suggest, however, that a compromise procedure be considered, namely, give the less general results and their proofs in the text, announce the general results, and outline their proofs in exercises.

Once the global structure has been decided so that we have a collection of propositions partially ordered by implication, then comes the problem of reducing this partial order to a compatible

linear order as required for presentation in a book. The author must choose one of numerous possibilities. The five criteria suggested above for use in deciding the best global organization may also be used here. My tendency is to give priority to broad results over specialized ones, and easily proved results over the more difficult. When reading mathematics I have often been annoyed by an author's failure to present a most revealing observation or proposition at the point where it would have done me the most good; given earlier instead of later, it would have helped me plow through the intervening obscurities. In case the revealing propositions cannot be proved until the later stage, mention could be made of it at the earlier one.

It happens all too frequently in the writing process that a projected linear order does not pan out. When the linear order proves incompatible, major revision is called for. A more frequent occurrence is to discover, while trying to write the proof of a proposition, that a stronger form of an earlier proposition is needed, and could have been proved at that point. The amount of this kind of rewriting can be reduced by using what I call the *backward* method of writing. One begins by constructing an outline of the book, section-by-section, in sufficient detail that the definitions, theorems, and lemmas of each section are spelled out. Then one writes out all proofs starting with the last section, and working forward. Whenever an earlier proposition is needed in a proof, one checks that it is adequate as stated in the outline; if not, the outline is revised to provide an adequate one.

An objection to the backward method is that the numbering of forward references must be changed as the earlier sections are revised. A simple solution to this difficulty is to leave blank each forward reference, and signal the blank by a marginal mark followed by a provisional reference number or a note. When the writing is complete, it is a minor task to find the blanks and insert the correct numbers. This suggestion applies also to the forward method of writing.

A minor problem is to decide how many global symbols to use (i.e. symbols whose meaning is fixed throughout the book). These must include of course the standard notations of mathematics, and the commonly accepted notations of the subject of the book. How many more should one have? It is best to be conservative here because the reader can tolerate much less of this than the author finds convenient. The main advantages of an extensive global notation are that it saves writing, reduces the length of a book, and allows for compact

formulas and diagrams without elaborate accompanying explanations. Its disadvantages are that it burdens the memories of readers (a grasshopper reader may run into a block because he missed the first use of a symbol), and typographical errors involving symbols are very hard to spot, and, when uncorrected, produce serious confusion. In my opinion, the disadvantages to the reader far outweigh the advantages, so I urge that the nonstandard global symbols be held to a minimum, say, five. If there are ten or more, an index of notation should be provided. The burden on the memory of the reader can be substantially lightened by strategically placed redundancies of the form "its adjoint T^*," "the C^1-norm $\|f\|$," or "the cohomotopy group $\pi^5(X)$"; these are especially helpful when the notation has not been used for many pages. Also it is clearly more important for the statement of a theorem to be free of dependence on notation than for its proof to be so.

An author of a research monograph that is first in its area has the opportunity and the obligation to replace poor by good terminology. If his book is a good one, and is much used, most of the terminology will come to be accepted as standard. The name that a research worker attaches to a new concept is usually chosen before the scope and thrust of the concept is fully understood, so his choice may be an unhappy one. Especially unhappy are notational names such as K-theory, $K(\pi, n)$-spaces, and the J-homomorphism. To avoid the guilt of establishing or perpetuating bad names, the author should make a list of them, consult dictionaries and thesauruses, make a list of alternatives, and then obtain the reactions of a few experts in the area. It is my opinion that a change of name will be accepted if the experts approve or are neutral; otherwise not.

While engaged in writing, an author is frequently required to decide which of several statements shall be called definitions and which theorems. To make the point clear, suppose that a new set of objects is to be introduced that can be expressed in several different ways as an intersection, say, of sets already at hand. One of these expressions must be chosen as definition, and each equivalence of it to another expression becomes a theorem. My tendency is to prefer the simplest expression: an easily verified condition makes a good definition, a subtle property should be a theorem.

In writing out proofs an author must always bear in mind the extent of the knowledge and mathematical maturity of the readers he wants

to attract and serve. Of course he should describe in the preface or the introductory chapter the background material he assumes as known. Because this description is necessarily rough, there will be numerous instances when he will wonder how much detail should be given. My tendency is to play safe by always giving a bit more detail than seems strictly necessary, and also by giving precise references to some of the less familiar background facts. It is especially during the final stages of the writing, when making local revisions, that I find myself adding sentences or paragraphs to ease transitions and clarify arguments. If the addition of a few more words makes the book accessible to many more readers, it is foolish to skimp.

Some authors have tried to solve this problem by inserting a preliminary chapter in which the necessary background material is outlined; a reader who can wade through it is prepared to go on. I am opposed to this scheme for several reasons. I suspect that few potential readers would take the test. Those who are well prepared would find it a bore. Only the ill-prepared or marginally-prepared would find it useful. But wouldn't it be more useful for these persons to try to read the first few chapters of the new material? My final objection is that such a summary is nearly impossible to write because its purpose is so ill-defined.

The proper place to remind the reader about a concept or proposition of the background material is at the point of the text where it is used. If a concept appears first in the statement of a theorem or definition, it is natural to write a preliminary paragraph in which the definition of the concept and some of its properties are recalled in an informal way.

Part of the task of writing the formal structure is the numbering of the statements to which reference must be made. Some editors with non-mathematical backgrounds have insisted that the number follow the leading Definition, Lemma, or Theorem. Some authors have carried this to the logical conclusion of having separate numberings for definitions, lemmas, and theorems; thus a reference of the form 5.3 is inadequate because there is a Lemma 5.3 and a Theorem 5.3.

When deciding questions of exposition an author usually considers only those readers who have read everything up to the point of the question. Let's call a reader who adheres strictly to the order of the presentation a *normal* reader. There is another type, the *grasshopper* reader, who consults the book to fill a gap in his knowledge. I contend that grasshoppers deserve nearly equal consideration with normal

readers because they form a substantial part of the users of any book. To see this, one has only to recall his own reading habits, how often he has been a normal reader, and how often a grasshopper. Once a mathematician becomes fully involved in research, he rarely has the time and patience to be a normal reader. It is also a familiar fact that a normal reader who finds himself stuck at some point behaves for a while like a grasshopper.

The needs of the grasshoppers are served by a good map of the territory, an adequate directory, numerous sign posts, and an index of locations. By a *map* I mean an outline of the results of the book, such as can be given in an introductory chapter (see the next section where this is discussed). By a *directory* I mean the table of contents; to be adequate, it should contain section headings as well as chapter titles. (By a *section* I mean a unit whose average length is three to five pages.) The *sign posts* are the chapter titles, section headings, the paragraph headings such as Definition, Theorem, etc., and the numbers of these and other important statements.

Since there is a bit of pain involved in making up section headings, some authors have been content to give only chapter titles. As a confirmed grasshopper, I deplore this. Some authors have been browbeaten by non-mathematical editors into placing the number attached to a definition, lemma, or theorem to the right of this heading. This makes it difficult to locate a desired reference number by scanning pages, especially so when the author numbers lemmas separately from theorems. Actually it would be most convenient to have the numbers appear in the left margins, but this requires exceptionally expensive setting of type. The next best procedure is to use boldface numbers close to the left margin. Authors must deal firmly with editors who complain of the ugliness of the boldface splotches running down the page.

Finally it is of the utmost importance that there be an index of the first uses of special terms and notations. Even a normal reader uses the index to ease the burden on his memory.

THE INFORMAL STRUCTURE

Now we come to the part of the author's task where restrictions are minimal and guidelines are difficult to discern. A natural procedure to follow is to examine what has been done by various authors, and

then to compare and classify the different parts of the informal structure. There are two books, among the ones I know, whose authors have made extraordinary expository effort, and the results are worthy, in my opinion, of careful study and perhaps emulation; they are *Lectures on the Calculus of Variations* by L. C. Young and *Dimension Theory*, by Hurewicz and Wallman.

I propose the following list of the kinds of informal material authors have used. First there are the introductory parts:

(1) brief reviews of background material to set the stage,

(2) presentation of motivations or leading questions,

(3) consideration of examples to derive conjectures,

(4) rough descriptions of the results to be obtained and methods to be used, and

(5) an outline of the book by chapters.

It is to be understood that items (1) through (4) include the introductory material to chapters and to sections as well as to the book itself. My list is concluded by the items that ordinarily follow the formal material to which they relate:

(6) connections with other subjects,

(7) discussions of alternative treatments, and

(8) historical comments.

I shall begin by discussing the last three items since I have but little to say about them. Item (7) refers to informal discussions of alternative treatments. In the preceding section, we mentioned formal presentations of alternative organizations; each such is based on an equivalence theorem. Obviously there is neither the time nor space for an author to present all reasonable alternatives in a formal way; in most cases he must be content with a brief description of the idea of the alternative.

The historical development of the subject also presents a set of alternatives. In many cases these differ radically from the formal structure of the book. It is my belief that students need to be impressed with and reminded frequently of the fact that the formal presentation they are following is far from unique, and bears little resemblance to the historical development. If the author is to apportion due credit to the research workers involved then he needs to sketch this development in rough outline at least. This is surely a most difficult task because the printed record may be lengthy, confused, and incomplete, and because the author's fellow researchers are sensitive to the

assignment of credits. A way of shirking this task is to substitute for historical discussions brief bibliographical references of the form "see [72, p. 332]"; few readers would pursue such a reference, and fewer still would learn much from it. I do not believe the task should be shirked; students need to be reminded that research work is a human activity, and the reputations of research workers are based on a number of such evaluations.

Hermann Weyl in his book *The Classical Groups* has done an outstanding job in providing historical notes and bibliographical references. Most of these are gathered together as notes at the end of the book, and are printed in smaller type. Another exceptional performance is found in the volumes on linear operators by Dunford and Schwartz. Here the authors present alternative treatments and historical comments as lengthy notes at the ends of chapters. Some of these notes are so detailed that they should be regarded as part of the formal structure. I prefer this placing of notes at the ends of chapters rather than at the end of the book; it saves a number of reference marks and some grasshopper-like activity. Best of all are the notes that are inserted at the appropriate place in the text; they are immediately relevant, and they provide relief from the rigors of the formal presentation.

Let me turn now to a discussion of the introductory parts. As remarked above, these include the introductory parts of chapters and sections. Of course, a particular section may be so well motivated by preceding sections that no introduction is needed; the formal structure can continue without break. In case a section needs an introduction, the part (4), descriptions of results and methods is not needed, since the results and methods are immediately at hand. However, all of the first three parts may be appropriate: the stage setting, the problem, and examples.

In the case of an introductory section to a chapter the reader needs to be reminded of the overall purpose or plan of the book, as set forth in the introduction to the book, and to be told where this chapter fits into that plan. This is a part of the setting of the stage. The remainder of the section is an enlargement and elaboration of the parts of the introductory chapter having to do with the chapter at hand.

When we come to considering introductory chapters, we find much less consistency among authors than in the case of the local intro-

ductory material. Some authors omit them completely. For example, Dunford and Schwartz make no attempt to entice readers to study their book, they do not say at the start what linear operators are about nor why they are important. The reasons for this omission are undoubtedly that their book is a reference work and text for a well-known standard field, and every mathematical education already includes much about linear operators; a sales-promotion job is unnecessary. One consequence of this omission is that they give no overall picture of the results they obtain; I would like to have had such a review for study, and I suspect that some students of their book would have found it useful.

It is notable that many textbooks of the calculus and other elementary courses make no attempt to sell their subject to students. Is it good to depend only on the fact that these courses are required for other subjects? Are they purely technical subjects having no intrinsic interest?

The arguments against having an introductory chapter are: (1) they are difficult to write, (2) it is a waste of effort to say imprecisely what is said precisely later on, (3) a reader who completes the book will forget that there was an introduction, and (4) a sales-promotion job should be beneath the dignity of a mathematician. I have no sympathy for reason (4); the direction of a young mathematician's career is largely determined by interests that have been aroused. It is absurd to suppose that a graduate student will learn just enough of all areas to be able to make a logical choice of his research topic. I am in sympathy with reason (1), but a purpose of this essay is to ease the task.

Reasons (2) and (3) go together. The fact that a reader forgets the introduction is no objection if the introduction helps him grasp the formal structure more quickly. At stake here is the question of how a student learns best. The first of two contending procedures is to ask him to examine first the lumber, bricks, and small structural members out of which the building is to be made, then to make subassemblies, and finally to erect the building from these. The second procedure is first to describe the building roughly but globally and provide a framework for viewing it, and then examine the construction of the building in detail. The first procedure would appeal to a student with a leisurely attitude who enjoys successive revelations. The second procedure, which I espouse, has the advantage that motivation

is present at every stage; the student knows where each item belongs when he examines it.

The second procedure can be elaborated by inserting between the first rough scan and the final detailed examination a series of scannings revealing successively finer details. Max Eastman in his book *The Enjoyment of Laughter* advocates and exemplifies this procedure in an amusing and convincing fashion. An argument favoring these successive approximations goes as follows. It has been observed that one learns a subject best not when first exposed to it but later when using the material in another study, or else when required to teach the subject. This can be paraphrased by saying that the nth scan is fixed in the memory by making the $(n+1)$st. Stated otherwise, when a reader has finished a book, he will retain in his memory only a more or less rough picture of the formal structure. This being so, why shouldn't the author assist the reader in formulating this rough picture? Surely the author's condensed version of the overall picture will be better balanced and more nearly accurate than one formed by an average reader.

Successive refinements

Let me illustrate the method of successive refinements with an example. Suppose we are to write a research monograph on the subject of elementary complex analysis. Our assumption is that it has just recently been discovered that the real field can be extended to the complex number field, that only a few experts are aware of the basic theorems about complex analytic functions, and that these theorems have been published only in ten to twenty scattered papers using a variety of definitions, approaches, and notations. The purpose of the monograph is to provide an organized account so that graduate students and mathematicians in other areas can penetrate more quickly to the heart of the subject. In short we might suppose that the present year is 1840, and we are Gauss or Cauchy. Such a supposition is not necessary to the validity of my illustration, and this is fortunate since I lack the detailed knowledge it would require.

The first approximation to the subject of a book is its title. The title *Complex Analysis* is too short to be meaningful to anyone other than an expert; our intended readers will not have heard of complex numbers, and the word analysis is a bit cryptic. The title *Calculus*

of Functions of a Complex Variable gives a reader something to hold on to. However, he is likely to be mystified by the word Complex, so I would replace it by Planar; then every word of the title is meaningful to the intended reader. (Observe that, in keeping with the advice given above, I do not perpetuate poor terminology, such as the adjectives "real", "imaginary", and "complex" for numbers.)

I have often felt that title pages and covers of books convey too little information to the reader about the contents; these have plenty of space to spare that could be used effectively. Publishers tend now to fill such space with designs that are somewhat pleasant but have little relevance. One way of using a part of this space effectively is to amplify the title. But titles need to be short for ease of reference. The dilemma is resolved by using a subtitle; bibliographers can omit it. In the case of our *Calculus of Functions of a Planar Variable*, I would adjoin the subtitle: *The two-variable calculus can be done by one-variable methods if, first, we enlarge our number system to a two-dimensional system called planar numbers.* This subtitle is our second approximation.

Our third approximation appears as the first half of the preface, and might go as follows:

> An important discovery of the last twenty years is that the concepts of the calculus of functions of one variable are meaningful in a context quite different from the usual one, and that most of the theorems of the calculus remain true in this new context. This is achieved by extending the ordinary number system **R**, thought of as making up a line, to a larger number system consisting of the points of a plane **C**. The new numbers of **C** are called *planar* in contrast to *linear* for the numbers of **R**. The variable z of a function $f(z)$, such as $z^2/(1+z)$, can then be regarded as a variable point of the plane **C**, and the corresponding values $w=f(z)$ likewise vary in **C**. In this way one can study functions $f: D \to \mathbf{C}$, where D is a domain of **C**, by the methods of the one-variable calculus.
>
> By the introduction of cartesian coordinates in **C** so that z becomes a pair (x, y) of ordinary (linear) variables, a function $f(z)$ obtains a representation by a pair $w=(u, v)$ of ordinary functions of the linear variables x and y. When $f(z)$ is differentiable with respect to the planar variable z, it turns out that u and v

must be quite special functions of x and y, in particular, they must satisfy Laplace's differential equation: $\Delta^2 u = 0$ and $\Delta^2 v = 0$. Nevertheless there are enough differentiable functions $f(z)$ to provide a flexible and adequate theory. For example we can use this theory to solve Dirichlet's problem for a large class of plane domains, and we can provide effective means of computing solutions in numerous specific situations.

The remainder of the preface would say what level of knowledge is expected of the reader, namely, ordinary one-variable calculus, and would then conclude with the customary statement about the circumstances of the writing of the book with credits to the sponsors and other helpful individuals.

I have often experienced feelings of envy while reading expository articles in areas of science other than mathematics. Their authors exercise a freedom of expression that seems to be denied to mathematicians; terms and phrases do not need to be defined with utmost precision, and statements need only be roughly true. Mathematicians suffer from the conviction that terms without precise definitions are meaningless, and statements that are not true are false or, at best, undecidable. These are essential restrictions to the presentation of a formal structure, but need not apply to the accompanying informal material. As authors and readers we must accustom ourselves to shifting gears when making a transition. Notice that the above sample preface covers the major features of complex analysis with a high degree of imprecision. None of the statements are sufficiently well-defined for even an expert to label them true or false. Yet together they enable the reader to construct a rough picture, and prepare his mind for the next approximation. That approximation, the introductory chapter of the book, might go like this.

SAMPLE INTRODUCTORY CHAPTER

Introduction

The key step in the developments recorded in this book is the enlargement of the ordinary number system **R** to the planar number system **C**. This must be done so that all the laws of arithmetic, holding in **R**, continue to hold in **C**. If we picture **R** as a coordinate line (the x-axis) in a plane **C** with cartesian co-

ordinates (x, y), then arithmetic operations and their properties can be visualized as geometric operations and their properties. Addition in **R** can be pictured as the superposition of intervals, and this extends very nicely to the ordinary vector addition. That is to say, we regard a planar number z as a vector whose initial point is the 0 point of **R**, and whose end is the point z; and then we define the addition of planar numbers to be ordinary vector addition (based on the parallelogram law). In this way, the function $f(z) = z + \beta$, where β is a fixed vector, is just the translation of the plane by the vector β.

The main difficulty in constructing the number system **C** is to choose a good definition of the product αz of planar numbers α, z. When $\alpha \in \mathbf{R}$, the product αz is taken to be the usual product of the vector z by the scalar α; then the function $f(z) = \alpha z$, for variable z and fixed α in **R**, is a radial expansion $(\alpha > 1)$ or contraction $(0 < \alpha < 1)$ of the plane centered at the origin. For a general fixed α in **C**, we define the mapping $f(z) = \alpha z$ to be the unique similarity transformation of **C** that leaves the origin fixed, carries the point 1 of **R** to the point α, and preserves the orientation of **C**. This rule for forming the product of α and z is most readily expressed in terms of the polar coordinates of α and z based on 0 as origin and the positive **R**-ray as initial direction. The rule reads: the polar angle of αz is the sum of the angles of α and z, and the polar radius of αz is the product of the radii of α and z.

Having defined addition and multiplication, we must now show that the laws of arithmetic hold in **C**; for example, multiplication is commutative and associative, and inverses exist for all numbers other than zero. Once this has been done, we observe that the numbers of **C** are exactly the same as the complex numbers that algebraists have introduced to provide enough roots of polynomials; these are the numbers of the form $x + iy$ where x, y are in **R**, and i is the "imaginary" number $\sqrt{-1}$. The unit point $(0, 1)$ on the y-axis is not at all imaginary, and its square in **C** is easily seen to be the point -1 in **R**. If we set $i = (0, 1)$, then $x + iy$ is defined in **C**, and is the point with coordinates (x, y). In this way our construction of the planar number system provides a logical and complete justification of the mystical conjurations of algebraists.

Once the number system **C** has been constructed then all the elementary functions of one variable make sense in this new context. For example, if α ($\neq 1$) and β are fixed in **C**, and z is variable, then the function $f(z) = \alpha z + \beta$ is a similarity transformation with $\beta/(1-\alpha)$ as its fixed point, and every similarity has such a form. The function $f(z) = 1/z$ is the composition of reflection in the line **R** followed by reflection (involution) in the circle of radius 1 with center at 0. The function $f(z) = z^2$ doubles the polar angle of a point and squares its polar radius, hence it maps each ray from 0 into another such, and maps each circle, centered at 0, twice around another such.

To discuss the derivative of such a function, we need the notion of the absolute value of a planar number; this is defined to be its polar radius (i.e. its distance from 0). Then the two basic laws for the absolute values of numbers in **R** continue to hold for planar numbers, namely, the triangle inequality $|z_1 + z_2| \leqq |z_1| + |z_2|$, and the product condition $|z_1 \cdot z_2| = |z_1| \cdot |z_2|$. It follows now that the usual definition of $\lim_{z \to a} f(z)$ is meaningful for functions of a planar variable, and the notion of limit has the same properties as for functions of a linear variable. Then also the definition of derivative is meaningful, and most of the standard theorems about derivatives continue to hold. In particular, the rational functions mentioned above are differentiable, and their derivatives are computed by the customary rules. For example, the derivative of z^n is nz^{n-1}.

The inverse function theorem for a function of a planar variable holds in the following form: if $f(z)$ is differentiable, and z_0 is a point where $f'(z_0) \neq 0$, then, in some neighborhood of $w_0 = f(z_0)$, the equation $w = f(z)$ can be solved for $f = g(w)$ and $g(w)$ is differentiable. For example, we can take square roots, and, more generally, any of the standard algebraically defined functions can be defined for a planar variable, and their derivatives can be found by the usual rules.

(*) Transcendental functions are extended by using power series; in particular, $\sin z$, $\cos z$, and e^z are defined by their Maclaurin series. The reason for defining them this way is to ensure that the rule for extending a function of a linear variable to a function of a planar variable shall commute with taking limits of functions. It follows that these extended functions satisfy the same algebraic identities and have the expected derivatives. But

now additional identities appear, for example $e^z = \cos z + i \sin z$, (which, incidentally, implies de Moivre's theorem). The principle illustrated by this last phenomenon is general: the extended theory illuminates and often completely explains the obscurities and puzzles of the old theory.

(*) We turn next to the definite integral. The interval $[a, b]$ of integration in the linear case must be replaced, in the planar case, by a rectifiable curve $\mathscr{C}$ from α to β. Then the Riemann sums corresponding to partitions of $\mathscr{C}$ are well defined, and the definite integral is their limit. The fundamental theorem of the integral calculus is valid in this form: if $f(z)$ is differentiable in a domain D, and if $\mathscr{C}$ lies in D, then $\int_{\mathscr{C}} f'(z)dz = f(\beta) - f(\alpha)$.

(*) In the representation of $f(z)$ by a pair (u, v) of functions of two linear variables it will be shown that $f(z)$ is differentiable as a planar function if and only if u and v satisfy the Cauchy-Riemann equations: $u_x = v_y$ and $u_y = -v_x$. This implies that, at a point where the derivative is not zero, the level curves of u and v are orthogonal and the mapping f is conformal.

(*) As an additional application, it is observed that the curves $u =$ constant are streamlines of a 2-dimensional flow of an incompressible fluid, and the curves $v =$ constant are equipotential lines; it follows that the theory may be used to solve fluid flow problems with prescribed boundary conditions.

(*) Now comes a striking difference between the old theory and the new. In the planar case we have Cauchy's integral formula, expressing a given differentiable function as an integral. Its proof is deep and subtle. It implies that any once differentiable function is, in fact, analytic. The Cauchy formula for the nth derivative is then obtained by formal differentiation under the sign of the integral.

(*) In elaboration of the theory of the coordinate parts u and v of a planar function f: if u and v are solutions of the Cauchy-Riemann equations, then they are analytic in the sense of the theory of functions of linear variables; it then follows that they are harmonic functions.

(*) In elaboration of the applications: Poisson's formula is given for the solution of Dirichlet's problem for a circle, the Riemann mapping theorem is stated, and it is indicated how it may be used to convert a solution of Dirichlet's problem or a fluid flow problem for one domain into a solution for another.

Paul 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

Reprinted with the kind permission of L'Enseignement Mathématique from Volume 16 (1970), 123-152.

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 as in writing biology, writing a novel, or writing directions for assembling a harpsichord: the problem is to communicate an 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 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 an 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 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 Σ 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 ε for ad hoc use.

Mathematics has access to a potentially infinite alphabet (e.g., x, x', x'', x''', ...), but, in practice, only a 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 π, and, of course, 0, 1, 2, 3, ... (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, "ε" for a positive integer, and "z" for a topological space. (A mathematician's nightmare is a sequence n_ε that tends to 0 as ε 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, and against the all too human partiality everyone

feels toward his own words, a red pencil is 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 do 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 begins 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 (or, 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 plan of writing 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 a 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, with, 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 known 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 non-

implication 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 the 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 insufficiently 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 dilatation 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 theorems: 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 pay 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 col-

lected 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 aspect 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 first 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 same 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 over 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 ε 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 it 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 references, 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 that 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 alphabetic 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 $0 \leqq \lim_n \alpha_n^{1/n} = \rho \leqq 1$, then $\lim_n \alpha_n = 0$." What does "ρ" 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 "ρ" 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 "ρ" 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

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

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's logical and alpha-

betical 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 also can 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 where 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 3×3) 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, collects 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 ε, let δ 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 at the end comes out to be less than ε, 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 ε, 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 it is probably impossible; nonetheless it's a better aim than no aim at all. How are we 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 "$\leqq$". A sentence such as "Whenever a positive number is $\leqq 3$, its square is $\leqq 9$" 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^* also is invertible", but "For invertible X, the adjoint X^* also is 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 "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 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 a lecture audience cannot do that. The reader can try to prove your theorems for himself, and use your 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 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.

a period or a 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 that 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 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.

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 to 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 of diction, your misspellings—everything that you can do wrong, short of the 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

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 all 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

Menahem M. Schiffer

1.

When I put down some ideas on expository writing in mathematics, I write more as a reader of many articles, textbooks and monographs than as an author. Indeed, the reader feels the difficulties and problematics of the exposition much more than the author, who in general likes his own style and wishes that everyone would write in a similar way. However, having written several expository papers and books, I should be able to tell something about the problems of the writer and to suggest some ways to meet them.

It should be stated at the beginning that it is impossible to give a universal prescription for writing in a clear, informative and attractive manner. Every exposition is a communication between the author and his reader and depends on the temperament, taste and scientific background of both. The following suggestions are therefore largely subjective and should only be considered by writers who feel a general affinity for my preferences and taste.

2.

In planning expository writing, the author should first of all decide whom he is addressing and what amount and type of information he wishes to transmit. Let us subdivide the various expositions into four different types: research paper, monograph, survey and textbook. It is evident that the style and the presupposed knowledge of the reader will have to be very different in these four types of exposition. It seems superfluous to stress this fact, but unfortunately many authors do not observe this obvious rule and may write a textbook in the style of a research paper with devastating consequences. Let us therefore briefly discuss the four types of exposition.

Copyright © 1973 American Mathematical Society

3. The research paper

Here the writer has the greatest freedom and needs indeed the least advice. He addresses himself to colleagues and coworkers whose knowledge of the subject and interest in his contribution can be taken for granted. He may be as brief and concise as he wishes and omit history, background and motivation for his work. However, even here it might be worthwhile to consider that by adding a little background information one might widen the audience from the close circle of specialists on the subject to a much more extended group of interested mathematicians. After all, the best achievements on research are made if methods and facts of two different groups of ideas can be combined. But even if one speaks only to experts in the field, one must avoid the danger of assuming that the reader knows every fact and trick of the subject under consideration and sees everything as clearly as the author who has devoted weeks of intensive thought to his particular investigation. I recommend here generous quotations of sources, clear stating of facts used, precise definitions and complete proofs, if proofs are given at all. I think it permissible, and often even unavoidable, to quote theorems without proof if the reader is given proper reference. It is surely not admissible to quote a theorem in such a way that it can only be understood if another book or periodical lies next to the reader. While writing the paper, the author should envisage the reader who has taken the paper to a place without a library and who is willing to believe a few facts on the say-so of the author, but also wishes to understand what he means.

It is very important to write a good introduction to the research paper. One should not expect the reader to work through many pages to find out eventually that the paper is of no interest to him. The introduction should allow him to orient himself in the field, the main results and the methods of the paper. If possible, the paper should be structured so that the most important results and definitions stand out and are clearly displayed. This enables the reader to skip details on first reading and to take a rapid look over the paper. Then he may decide to follow the argument in detail, but if he is an expert in the subject, he might prefer to provide his own proofs and arguments and so enjoy the paper even more.

These are the remarks of a person who likes to follow the current literature in his field, but is often frustrated to find how many papers

he cannot understand without devoting a disproportionate amount of labor. However, the writer of research papers needs advice least and will, in any case, follow his own taste.

4. The monograph

The monograph needs much more planning and attention than the research paper. In the present situation of fast developing theories and enormous output of research papers, there is a particular need for an exposition of larger fields of mathematical research. Such an exposition or monograph should allow professional mathematicians to inform themselves about progress and development in fields which are wider than their own speciality. The monograph should allow us to extend our knowledge faster and easier than is possible by reading and sifting numerous research papers; it should enable us to know and appreciate what is going on in nearby fields. The research paper may be written for the man who works on boundary value problems for quasi-linear partial differential equations in two variables; the monograph should aim at all people who work on partial differential equations.

In the long range, the monograph is more important and more widely read than the research paper. It should be very carefully organized and planned. The monograph should provide background and motivation for basic concepts, the growth of ideas and methods should be described and explained and more detailed proofs should be provided. An extensive bibliography is a natural must.

Every good mathematician hates to become too narrow a specialist and tries to widen his field and look for new applications of his old results. He shops through monographs to get new ideas and to find new problems. Hence, the monograph should be attractive and enticing. It is repulsive if a monograph stocks many introductory pages with definitions and trivial lemmas and forces the reader to work through this material without knowing what it is good for. If the reader skips this boring beginning and proceeds to the interesting parts, he is again forced to refer to the introductory pages for the notations, definitions and sometimes even the letters for certain quantities. There should be a way to develop a theory logically but also attractively and lead the reader to the main body of the subject

in an interesting way. Could one not define a concept when it is needed and prove a lemma close to the theorem for which it is used? Surely, the interest of the reader would be much greater if he knew the context of the definition or the lemma.

A good introductory chapter should whet the appetite of the reader. A historic and genetic approach may yield a good general guideline for organizing the monograph. An important special case might be discussed at the beginning, without too much apparatus, to show the beauty and significance of the theory, and as the theory develops through the book, the same special case might be discussed from a progressively deepening point of view. I remember a classical exposition of the calculus of variations in which one and the same problem was subjected to the various conditions and criteria of extremality; I enjoyed the increase of insight with each progression of the theory. Once the reader is convinced that the subject matter is of interest and significance and to his taste, he is quite willing to make greater efforts to penetrate deeper and to master the subject.

There are some warnings for writer of monographs: Do not use the jargon and notations which are common in seminars with closest collaborators in the field and suppose that everybody knows them. Assume always that the reader knows less than you. The monograph is not written to show how erudite or skillful you are, but in order to teach the reader some new material. Hence, do not always use the shortest argument if it is not the most natural one—better say a little more than too little. Do not heap too much recent material into the text only to be up-to-date—judge material by its significance rather than by its novelty. An easy and clear exposition can be made a valuable guide to the whole field if the bibliographical references are put in the most appropriate places. One can provide considerable help to the reader by a clear and detailed table of contents as this allows a quick orientation and overview at the beginning.

I find it always stimulating if the author adds some remarks on the future trend in his subject and on open problems in the field at the end of the monograph. The beginner is, in general, overwhelmed by the wealth of methods and results and gets the impression that the subject is exhausted. Hence a list of unsolved problems and research desiderata will stimulate him to deeper study and will direct his attention to the right questions.

5. Surveys

The survey is a report to the mathematical community at large and many excellent models can be found in the traditional hour lectures given at the meetings of the American Mathematical Society and published in the *Bulletin* of the AMS. In the present state of our science it is nearly impossible for all mathematicians to benefit equally from such a survey. The author should try at least to give to all an understanding of the problems discussed and the general progress made. In a more specific way, the survey should be directed toward a large subgroup of the mathematical community, say to all analysts, algebraists, or topologists. The survey should enable the listener or reader to grasp the general ideas, methods and main results of a sufficiently wide field of research. It is not necessary to give proofs for all facts described but sometimes a typical proof might exemplify a characteristic method of research. The survey should serve to cross-fertilize with distant fields of mathematics, and I stress again that often mathematical progress results from conjunction of ideas and methods from separate disciplines.

A survey is an invitation to a field of research and not an introduction, as is the monograph. Therefore beware of special details, of definitions whose role in the theory is not quite clear. Motivation, background in the general wide field of research, history and problems should be displayed. An educated reader should be able to follow the survey without the need to look up additional literature if he is willing to trust the author that the theorems and facts given are correct. If he is then really interested in the topic surveyed and knows roughly what it is all about, the bibliography of the survey should enable him to find his way to a detailed study of the subject. The bibliography is also a good indicator of the significance of the field surveyed. If many authors over a considerable period of time are quoted, one may suppose that an important and permanent field of research has been discussed; if only the author and a few other authorities are cited, or if the whole literature on the subject is dated within a very short period of time, the survey will probably be too narrow.

6. Textbooks

The present essays should be most helpful to writers who intend to prepare a textbook. I shall confine myself to the discussion of more advanced texts, say on the senior or graduate level, since more elementary texts need pedagogical rather than mathematico-logical considerations. The need and importance of advanced textbooks in mathematics can hardly be overrated. Indeed, the number of advanced courses which a student can take at the university is rather limited, and a large fraction of the knowledge of the future mathematician has to come from independent reading in good textbooks. By the way, a conscientious teacher giving a course in a more advanced subject will be aware that his exposition and arrangement of material is one of many possible ones and, to avoid one-sidedness, will recommend considerable collateral reading from textbooks. Thus the future of our science depends to a considerable extent on the production of excellent texts.

The purpose of a textbook is to take a student with a specified amount of preparation and introduce him to a new field of mathematical endeavor. It is most essential that the presupposed knowledge of the reader be precisely realized and that the treatment in the textbook take this carefully into consideration. In contradistinction to the preceding types of exposition, the author of a textbook has also to consider a psychological problem besides the purely logical one. He has to attract the student to the subject and convince him that he is learning a significant, beautiful and worthwhile piece of knowledge. Many text writers fail to realize the difference between a monograph and a textbook. The monograph reader is already motivated for his study, but the student has still to be convinced of the importance of the field. On the other hand, of course, some monographs may make excellent textbooks and a good textbook may serve also as a monograph.

The textbook should rise from the known to the unknown in easy steps. It should start from the special and the intuitive and proceed hence to the general and the abstract. The old logical rule holds that when you gain in extension, you loose in intension. Thus, the special case allows many insights which get lost in greater generality; if the student then sees how much of the special argument survives in the general context, he will develop a healthy respect for the method of mathematical abstraction.

Applications and examples should be generously given and repetitions and redundancies need not be avoided. A special case may be proved by a simple argument and when the basic method has been driven home, the general case may be attacked by the same but more involved argument. While mathematical rigor and precision must be observed, the author might well begin his discussion with some convincing intuitive reasoning. In my opinion, the general theory is no more than the sum of all special cases and very general theorems without concrete applications often fail to impress the student. E. Schmidt once said that the value of a new mathematical theory should be judged by the problems in previous mathematics which it could help to solve and not by the internal results between the concepts created by the new theory. Hence it is always very gratifying if some applications of the theory can be given which show its power and significance. Complex analysis may be applied to number theory, differential equations, algebra, or physics. The textbook should be rich enough to serve many different tastes.

Many textbooks are written with a quite specific course in mind and contain material which can be covered in a semester or a quarter. While such books may fill some local needs for some time, they are not very valuable for the student in general. They are printed lecture notes and are best used as aids for the lecture course. A real textbook should contain more material than can be covered in a course. This allows teachers a certain amount of flexibility when they use the book as their main textbook and makes it possible to recommend it in various courses for collateral reading. A textbook should contain enough material to serve as a good reference book in the subject. It will be of great value for many years to come since a good book which has once served as a study text will remain a very helpful tool to refresh the memory and add information to the future research worker. In particular, it should be remembered that a text, say in applied mathematics or in differential equations, may be a stepping stone in the education of a mathematician, but may mark the highest point in the mathematical education of a scientist or engineer. Such users will need to refer to the book for a long time in their careers.

There is a trend in good textbooks to develop general ideas in the main body of the book and to put many concrete applications and amplifications into a well-organized problem section. This method has advantages and disadvantages. An obvious advantage is that

many arguments which have been presented in the text can be used by the student to derive important new results. He deepens the knowledge of the method and widens the results and information on the topic. The disadvantage lies in the possibility that a wrong perspective of the importance of ideas may be created. The application in the exercise may be the motivation for the general concept in the text and the relative importance of the two may be misunderstood. For example, suppose the concept of compact families of analytic functions is given in the text and the Riemann mapping theorem added as an exercise. This is very feasible and has been done in some texts on analysis. A student who would skip some problems might have learned a general concept without knowing an important result of analysis whose proof has motivated the concept. But the main desideratum would be that the solution to all significant problems in the exercise section should be given, or at least clearly hinted at. The classical book in analysis by Pólya and Szegö [4] gives a convincing example that it is possible to teach advanced mathematical topics by a sequence of graded problems and I recommend the writers of textbooks to study this model of teaching by problems. Observe that in the second half of this book all problems posed are solved, so that the student can check his efforts if he has solved the problem or learn the correct solution if he failed. A refinement of this procedure was developed by A. Ostrowski [2] in his textbook on differential and integral calculus. He has an imposing list of very instructive problems after each section. In the second third of the book hints for solutions are provided, while the last third gives the complete solution. This allows the author to include many tough problems which strain the ability of the student to the utmost but avoid discouragement. While on a lower level than the advanced textbooks which I discuss here the arrangement and organization of Ostrowski's book is recommended as a good example. In the second edition of his classical *How to solve it,* G. Pólya [3] uses the same device.

A textbook should not be too tightly written and too pedantic a notation should be avoided. Often a student wishes to learn a part of the subject matter treated in the text and this will in general be a more advanced part. If a systematic notation is used throughout the text, and the definition of certain letters is kept the same in all chapters, it is very difficult to skim. It might be a good idea to write

each chapter in such a way that it can be understood without too close a study of the preceding chapters. Letters and symbols might be briefly recalled when they first appear in each chapter and references and applications of preceding chapters should be cited carefully. In this way an advanced student may enter the book at the point which is most important to him. A model in this respect for me is *Methods of Mathematical Physics* by Courant and Hilbert [1]; this book is really not repetitious at all, but the interested reader can get his information on a large number of subjects without starting the book from the beginning. Another difficulty for the reader may be avoided if not too many logical symbols are used. One is accustomed to read at a certain speed, and if one proceeds through a forest of logical symbols and disentangles them step by step, one may be quite discouraged. Frankly, I have not yet found any arrangement of $\forall$, $\exists$, $\vee$, $\wedge$, etc., which I could not dispense with by a few well-chosen words.

The author of the textbook should aim for clear and interesting exposition rather than for completeness or novelty. One sometimes sees the inferiority complex of an author looking through the crowded references to recent literature and quotations which do not help the student at all at his level of preparation but tend rather to discourage him. If one wants to bring in new developments and hint of further applications and problems, one may use a very helpful device. To each chapter, a section on bibliography with hints and annotations with complements and additional problems might be added that may be read by the more advanced student but skipped by the beginner.

Let me discuss again at the end of this section the weakest point in textbook writing, that is, undue conciseness. Most mathematicians form their expository style by writing research papers which they wish to publish in scientific journals. The lack of space in these media and the consequent need for extreme brevity affect their writing in general and condition them to a telegraphic style and utmost condensation of argument. This is not even desirable in research publications, where it is, however, unavoidable. By no means should this habit spill over into textbook writing.

On the contrary, the present style in research papers adds a great responsibility to the textbook author. When we read biographies of outstanding mathematicians from the 19th and the beginning of the 20th century, we often run across a statement that they learned less

from their regular university courses than from studying the works of the mathematical classics. We all agree that similar inspiration would be much harder to find in the laconic and parsimonious writing of our present masters. Here the modern textbook has inherited an additional task with respect to the gifted student; it has to a large extent to replace the role which the collected works of the great mathematicians of yesterday have played.

Some teachers say that they expect the textbook to contain the definitions and precise proofs, while they are quite willing to provide the background, motivation and amplification of the subject matter. But observe that a textbook serves in general only for a short time as a tool for specific courses and hence, if it is worthwhile at all, it should stand on its own feet and allow the student to use it for self-study.

Therefore do not fear to be accused of verbosity and prolixity. Allow even a certain redundancy in your exposition. It is often desirable to provide a heuristic argument for a theorem which explains the basic idea of the proof without going into the ϵ's and δ's. When the method is clearly understood, the rigorous argument will follow. Take, for example, the existence proof for solutions of ordinary differential equations with given initial data by the method of successive approximations. I would not start with enumerating all assumptions on Lipschitz conditions, boundedness requirements and admissible intervals. Rather show first how all conditions to be fulfilled can be united into one integral equation. Next bring in the concept of a functional transformation and the idea of fixed points under such transformations. Then discuss contraction mappings and their significance. After all these ideas have been explored intuitively, prepare the ground for the final and rigorous proof by making the usual preparatory assumptions and, if possible, explain where each becomes necessary in the general plan of attack. My example deals with a very elementary theorem, since I wish to be understood by all colleagues, but the value of the illustration should not be affected by this fact. Thus, summarizing: Give the important theorems in two stages, the heuristic argument and the rigorous logical chain.

7.

I come now to the most important part of this essay. Namely, instead of discussing what book to write, discussing how to write it.

There is a big difference between the completed opus and the way the manuscript looks during most of the writing. Even the plan of a book changes often, while the writing proceeds and the production of any book is a process of successive approximation. The worst moment in the preparation of a book is surely the first moment when one starts on a blank sheet. In our profession actually the situation is not quite so bad since most of us lecture on the subject of interest, and we may suppose that a lecture manuscript has already been prepared. This already enforces a certain logical order and structure for the book; however, it is possible that you are not quite satisfied with your original arrangement or that you wish to write about a wide field which you have never covered in one lecture course.

In this case my advice is: Start with that chapter which interests you most and in which you think you can be most original and helpful. The best sections of a book are always those which are written enthusiastically in one piece. If you are satisfied that the main pieces of the book are very good, you have an excellent starting point. Ask next, what material is needed to bridge from the presupposed knowledge of the reader to the main sections of the book. It is remarkable that the aim for the advanced chapter brings order and system into the introductory and auxiliary sections. In a similar way, connecting sections between the different highlights of the book have to be prepared.

When this stage of the manuscript is reached, you will find that the auxiliary chapters are dry and uninteresting and that a certain disproportion in the material selected prevails. The auxiliary chapters have to be fleshed out. The important principle is that no chapter shall be entirely auxiliary and be the servant of some other chapter. Each chapter must obtain its own highlight and its moments of achievement and satisfaction. We are dealing with a balanced textbook and personal preference may weigh the choice of material but must not prevent the exposition of standard matter. The rewriting of the added chapters will influence again the exposition of your "pièce de resistance" from which you started. There is no harm in rewriting a few times; once the general idea is clear, the labor of rewriting is not too big and the style and clarity improve in general quite a bit under such repeated goings-over. An important precaution for rewriting and making additions to the manuscript is the right kind of numbering for chapters, sections and formulas. It is advisable

to number the formulas in each section separately so that changes affect at most one section and not the whole book. If new sections are added, one should try to do this at the end of the chapter, so as not to disturb the section numbers which might involve a correction of numerous cross-references.

I should like to mention a different method of writing a textbook which is recommended and followed by one of the most successful authors of mathematical books, my colleague G. Polya. He uses the system which he calls "skeleton writing". Prepare the book according to your plan but in a very sketchy and incomplete way. Then, when the skeleton of the book is ready, flesh it out and bring it to life. One advantage of this method is that you do not mind major changes and reshufflings in the manuscript while you might have great inhibitions to change or drop a carefully written section. This method allows an intensive interaction between earlier and later sections in the book. This method differs from the above, but it may be usefully combined with it in the construction of various chapters.

Once the manuscript has obtained well-proportioned structure and covers the field one has set out to describe, one should compare it with available textbooks, monographs and papers in the field and neighboring subjects. This will prevent omissions and oversights which might otherwise occur. If the new book is really worthwhile, it should be possible to incorporate additional and new material in an original way. Indeed, this is the best test for a good new book, that if you know its content, you are able to read and understand the literature on the subject. This final testing and adding to the text gives in general great pleasure. One has a collector's pride in having nice illustrations, applications and amplifications. One must now beware not to overdo it and transform the book into an encyclopedia on the subject. Finally, add problems, exercises, addenda and bibliography.

When the manuscript is finished, it is advisable to use it as a basis for a course or more to test the clearness of exposition and the logic of the arrangement. Criticism of colleagues and graduate students may be very helpful, for it is remarkable how blind an author can be to his own misconceptions. After all these tests, the manuscript should be ready for publication. The problems of printing which will then arise would make a good topic for another essay, and I shall not discuss them.

A final advice is: Enjoy your writing and relax while doing so. Write in a natural style and leave the officialese and formal style to administrators and government departments.

I am sure that other writers of books have a very different procedure and that the above method will fit only a part of prospective authors. But there may be a number of colleagues who have a tendency similar to mine and they may benefit from my experience.

References

[1] R. Courant and D. Hilbert, *Methods of mathematical physics*. Vols. I, II (Vol. II by R. Courant), Interscience, New York, 1953, 1962. MR **16**, 426; MR**25** # 4216.

[2] A. Ostrowski, *Differential and integral calculus, with problems, hints for solution, and solutions*, Scott, Foresman, Glennview, Illinois, 1968.

[3] G. Pólya, *How to solve it*. Princeton Univ. Press, Princeton, N. J., 1971.

[4] G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*. Band I: *Reihen, Integralrechnung, Funktionen Theorie*, Vierte Auflage, Heidelberger Taschenbücher, Band 73, Springer-Verlag, Berlin and New York, 1970. MR **42** # 6160.

Stanford University

Jean A. Dieudonné

1. Distinction between research monographs and textbooks

I think this has not been sufficiently pointed out. More precisely, the style of writing need not be the same when you address yourself to an expert or to a beginner. In particular, I think it is only an expert who can indulge in the "grasshopper" way of reading which Steenrod emphasizes; a student who knows nothing on the subject would be hopelessly bewildered if he tried to read in that way. For research monographs, I would, therefore, consider as satisfactory the method Steenrod recommends, allowing some looseness in the general organization, the skipping of a lot of proofs or comments which are trivial for experts, etc. On the contrary, when it comes to textbooks aimed at beginners, I am entirely in agreement with Halmos regarding the necessity of a very tight organization, and I would even go beyond him with regard to the "dotting of the i's"; this may well be annoying to the cognoscenti, but sometimes it will prevent the student from entertaining completely false ideas, simply because it has not been pointed out that they are absurd.

This brings me to my second point.

2. How detailed should a proof be?

Here again, in a research monograph a great many things may remain unsaid, since one expects the expert reader to be able to fill in the gaps; one should, however, even in that case, remember Littlewood's advice: you may very often skip a single link in a proof, but never two consecutive ones. For textbooks, on the contrary, I again go beyond Halmos in believing that all the details must be filled in with only the exception of the completely trivial ones. In my opinion, a textbook where a lot of proofs are "left to the reader" or relegated to exercises, is entirely useless for a beginner. Any time a previously proved theorem is used, a reference to it should be given

Copyright © 1973 American Mathematical Society

unless it comes in with such frequency that the most obtuse reader will have memorized it. Similarly, in addition to a thorough Index of notations, any time a notation comes up which has not been used for many pages, a reference to its definition should be given.

3. Introductory material and writing "about" mathematics

I am not convinced by Steenrod's arguments. In a research monograph a long introduction seems quite unnecessary, since the (expert) reader is supposed to have already a good background in the topics treated; the table of contents should, in fact, be enough. For a textbook, an introduction going into many details will simply be ununderstandable to the beginning student, since by assumption he has never heard of the subject. Partial introductions to the various chapters may be more useful, since they may enable the student, *after* he has gone through the chapter, to come back and have a bird's eye view of it, with the main points being properly emphasized.

Université de Nice