THE BOOK OF *Numbers*

John H. Conway • Richard K. Guy

COPERNICUS
AN IMPRINT OF SPRINGER-VERLAG

Published in the United States by Copernicus, an imprint of Springer-Verlag New York, Inc.

Copernicus
Springer-Verlag New York, Inc.
175 Fifth Avenue
New York, NY 10010

Library of Congress Cataloging in Publication Data

Conway, John Horton.
The book of numbers / John Horton Conway, Richard K. Guy.
p. cm.
Includes bibliographical references and index.
ISBN 0-387-97993-X (hardcover : alk. paper)
1. Number theory—Popular works. I. Guy, Richard K. II. Title.
QA241.C6897 1995
512′.7—dc20 95-32588

Manufactured in the United States of America.
Printed on acid-free paper.

9 8 7 6 5 4 3 2 1

ISBN 0-387-97993-X SPIN 10490867

Preface

The Book of Numbers seems an obvious choice for our title, since its undoubted success can be followed by *Deuteronomy*, *Joshua*, and so on; indeed the only risk is that there may be a demand for the earlier books in the series. More seriously, our aim is to bring to the inquisitive reader without particular mathematical background an explanation of the multitudinous ways in which the word "number" is used. We have done this in a way which is free from the formality of textbooks and syllabuses, so that the professional mathematician can also glean important information in areas outside her own speciality, and correspondingly enrich her teaching.

The uses of the word number are diverse, but we can identify at least three separate strands. The development of number (usually written in the singular), continually adapting and generalizing itself to meet the needs of both mathematics and its increasing variety of applications: the counting numbers, zero, fractions, negative numbers, quadratic surds, algebraic numbers, transcendental numbers, infinitesimal and transfinite numbers, surreal numbers, complex numbers, quaternions, octonions.

Then there is the special study of the integers, Gauss's higher arithmetic, or the theory of numbers (usually written in the plural), which overlaps the more recent area of enumerative combinatorics. Special sets or sequences of numbers: the prime numbers, Mersenne and Fermat numbers, perfect numbers, Fibonacci and Catalan numbers, Euler and Eulerian numbers, Bernoulli numbers.

Finally there is a host of special numbers: Ludolph's π, Napier's ϵ, Euler's γ, Feigenbaum's constant, algebraic numbers which arise in specific contexts ranging from the diagonal of a square or of a regular pentagon to the example of degree 71 which arises from the apparently simple "look and say" sequences.

Although mathematics is traditionally arranged in logical sequences, that is not the way the human brain seems to work. While it is often useful to know parts of earlier chapters in the book when reading some of the later ones, the reader can browse at will, picking up nuggets of information, with no obligation to read steadily from cover to cover.

Chapter 1 describes number words and symbols and Chapter 2 shows how many elementary but important facts can be discovered "without using any mathematics." Chapters 3 and 4 exhibit several sets of whole numbers which can manifest themselves in quite different contexts, while Chapter 5 is devoted to the "multiplicative building blocks," the prime numbers. Fractions, or "algebraic numbers of degree one," are dealt with in Chapter 6 and those of higher degree in Chapter 7. So-called "complex" and "transcendental" numbers are explained in Chapters 8 and 9. The final chapter is devoted to the infinite and the infinitesimal, to surreal numbers which constitute an extremely large, yet infinitely small, subclass of the most recent development of number, the values of combinatorial games.

The color graphics in Chapter 2 are the work of Kenny and Andy Guy, using Brian Wyvill's Graphicsland software at the Graphics Laboratory in the Computer Science Department at The University of Calgary. Three of the figures in Chapter 4 are reproduced from Przemyslaw Prusinkiewicz and Aristid Lindenmayer, *The Algorithmic Beauty of Plants*, Springer-Verlag, New York 1990, by kind permission of the authors and publishers. The first figure is due to D.R. Fowler, the second to D.R.F. and P. Prusinkiewicz and the third to D.R.F. and A. Snider.

We thank Andrew Odlyzko and Peter Renz for reading a much earlier draft of the book, and for making many useful suggestions. We

are also grateful for the polite and patient persistence of Jerry Lyons, Liesl Gibson and Henry Krell, and for the general competence of the Springer staff. We are proud to be among the early contributors to what is yet another successful Springer series.

Princeton and Calgary John H. Conway and Richard K. Guy

Contents

1

The Romance of Numbers

Number Words

Throughout history, *number* and *numbers* have had a tremendous influence on our culture and on our language. Thousands of words are obviously associated with numbers, for example,

a monologue	is a speech by 1 person;
a bicycle	has 2 wheels;
a tripod	is a stool with 3 legs;
a quadruped	is an animal with 4 legs;
a pentathlon	consists of 5 athletic events;
a sextet	is a piece for 6 musicians;
a heptagon	is a 7-cornered figure;
an octopus	has 8 "feet";
a nonagenarian	has lived for 9 decades;
and decimal	means counting in 10s;

but in many other cases the connections, once just as vivid, have been obscured by the passage of time and changes in meaning.

Have you ever realized just how many words are associated with "number" itself? This comes from an Indo-European root meaning "share" or "portion" and seems to have been originally associated with the division of land. "Nimble" refers to one who is quick to take his share; your "nemesis" was originally your portion of Fate; and "numb" means "seized" or "taken." A "nomad" is one who wanders about in search of pasture land. There are many technical "nom" words: "binomial" (two-numbered), "astronomy" (numbering or allotting the stars), "economy," "autonomy," and so on.

In German we still have the word *nehmen,* to take, with past participle "nimm." English once had much the same verb, but by Shakespeare's day the word "nim" usually referred to illegal taking or pilfering. In *Henry V*, Corporal Nym is a doubtful character that one might well suspect of nimming. In our century "nim" has become the name of a take-away game.

Latin and Greek *nomisma* was a coin, and this has given us "numismatics," the love of coins, and "nummulite," a coin-shaped fossil. "Supernumerary" now often means "redundant," but originally supernumerary officials were those in addition to the number laid down by the regulations.

HICKORY, DICKORY, DOCK— EENY, MEENY, MINY, MO

There are several systems of words for numbers that survived for special purposes in out-of-the-way places. The particular sequence of words varies both with the part of the country and the purpose for which they were used. One such system is

wan, twan, tethera, methera, pimp,
sethera, lethera, hovera, dovera, dick,
wanadick, twanadick, tetheradick, metheradick, pimpdick,
setheradick, letheradick, hoveradick, doveradick, bumfit,
wanabumfit, ...

Such rustic sequences appear in many countries. They are usually highly corrupted versions of the standard number systems of ancient languages.

Hickory, dickory, dock.
The mouse ran up the clock.
The clock struck one;
The mouse ran down.
Hickory, dickory, dock.

Probably "hickory," "dickory," and "dock" are the words for "eight," "nine," and "ten" in one of these systems (compare "hovera, dovera, dick"), while "eeny, meeny, miny, mo" mean "one, two, three, four" in another.

Most languages have interesting features about their number words, which repay study. We content ourselves with exhibiting those in Welsh. The alternative forms are for masculine and feminine, while the parenthetical forms are elisions in front of certain letters. There are separate ways of saying 50, 70, and 90.

1	un
2	dau, dwy
3	toi, tair
4	pedwar, pedair
5	pump (pum)
6	chwech (chwe)
7	saith
8	wyth
9	naw
10	deg
11	un ar ddeg
12	deuddeg
13	toi, tair ar ddeg
14	pedwar, pedair ar ddeg
15	pymtheg
16	un ar bymtheg
17	dau, dwy ar bymtheg
18	deunaw
19	pedwar, pedair ar bymtheg
20	ugain
30	deg ar hugain
40	dugain
50	deg ar deugain; hanner cant
60	trigain
70	trigain a deg; deg ar thrigain
80	pedwar ugain
90	pedwar ugain a deg; deg a phedwar ugain
100	cant

Note that it is less duodecimal than many languages, which have separate words for eleven and twelve. It thus lends itself easily to modern decimalization. There are signs of bases 5 and 20, as well as 10. Curiosities are $18 = 2 \times 9$ and $50 = \frac{1}{2}(100)$.

1 Now we come to words associated with particular numbers. When we say someone's "alone," we're really saying he's "all one." The word "only" means "one-like," and mixing these ideas we get "lonely." Originally you would "atone" for your misdeeds by becoming "at one" with the person you'd affronted. When Mother Hubbard's "poor dog had none," he had "not one." Once upon a

time the word "once" was used just for "one time"; a "nonce word" is a word used "for the nonce," that is to say, just for this one time.

The "United States" are many states that have become one. A trade "union" "unifies" its members into one body. Its members are "unanimous" if they speak with one spirit. A principle is "universal" if it holds throughout the "universe," the world turned into one object. Faculty and students form a "university," again by being turned into one body; students at a school wear a "uniform" if their clothing is of one form. "Unicorns" are mythical beasts with one horn.

"Onion" comes from the same Latin word as "union." Some rustic Romans referred to an onion as "one large pearl."

The Latin *uncia* was a unit that was 1/12 of a larger unit, the *as*. It has given us "inch" for 1/12 of a foot, and "ounce," originally 1/12 of a pound. There are still 12 ounces to the pound in Troy weight. The "uncial" letters of ancient manuscripts are large, inch-high letters.

"Monopoly" still has its original Greek meaning of "selling by (only) one person." A "monk" (Greek *monakos*) is one who is alone or solitary. A "monolithic" monument is made from one (large) stone, and your "monogram" is a way to write your name with one drawing, but perhaps we should leave "one" before we get too "monotonous."

2 The Old English word for this number had two forms: the masculine, which has given us "twain," and the feminine, or neuter, which gave us "two." Many English words that contain "tw" are related to these. "Between," and "betwixt" refer to the position of an object in relation to two others. "Twilight" is the "between light" that separates night from day. "Twine" is formed by a "twist" of two strands ("twiced") into one. "Twill" is a two-threaded fabric, as is "dimity." "Tweed" was a trade name that originated from an accidental misreading of "tweel," the Scottish form of "twill." A "twig" results from one branch splitting into two, and a "twin" from the splitting of one egg into two. Samuel Langhorne Clemens took his *nom de plume* from the cry "Mark twain" (two fathoms) that he must often have heard during his days as a Mississippi river pilot.

"Didymium," a rare metal discovered in 1841, was so named because of its twin-brotherhood with Lanthanum (the Greek for "twin" is *didymos*, a word that mentions "two" twice). The name was specially apposite since, in 1885, didymium was itself found to be a mixture of two elements, Praseodymium (the "leek-green twin") and Neodymium ("new twin").

A concept exhibits "duality" when it has two aspects or has a "dual" purpose; a "duel" is fought by two persons, while a "duet" is played by them. The word "double" has given us "doublet," originally a garment with two layers, "doubloon," originally double the value of a pistole, and "doubt" and "dubious," being in two minds. "Double-tongued" and "bilingual" have literally the same meaning, but their usage is quite different! A "dilemma" is an argument that catches you on either of two opposite assumptions.

A "biscuit" has been "twice cooked," the "biceps" are a "two-headed" muscle, "binoculars" are intended for two eyes, and to "bifurcate" is to split in two at a fork. The "bis" prefix, meaning "twice," derives from the Latin *duis*, which has also given us *dis*, or *di*, which means "split," "apart," or "away." Thus a "digression" is literally a "walking away."

Since splitting is often acrimonious, "dis" words often carry a pejorative, or even negative, sense. For instance, a "devil" (Greek *dia-bolos*, throw "between"—itself a "two" word) sows "discord" by separating two hearts. The variant "bis" is sometimes pejorative (it has given us "embezzlement") and sometimes not. Compare "bisect," or "cut into two (probably equal) parts," with "dissect," or "cut apart." Your dictionary contains lots of "dis" words: "dismember," "dismantle," "disrupt," "distort," "disturb," "discredit," "discourage."

We "distrust" someone who is "duplicitous." The words "duplex," "duplicate," and "duplicity" all have the same etymology, "two-fold," although now a "duplicate" is just a copy, and a "duplicitous" person is one who is two-faced.

A "diploma" is a folded (literally "doubled") piece of paper entitling you to certain privileges. A "diplomat" is one who carries a diploma, entitling him to speak for his country's government.

Some of the most surprising words connected with "two" come from the Latin *diregere* (to lead or guide), which combines the "dis" prefix with *regere* (to rule). "Dirigible," for a directable balloon, comes "directly" from this word, as does "dirge," from the Catholic Office for the Dead:

Dirige, Domine, . . . , viam meam.

So, of course, does "direct" itself, also "director," "directory," and "directoire" (a style of furniture popular in the period of the Directoire). Less obvious is French *dresser* (from late Latin *directiare*, to direct), which has given us "dressage" (the direction and adornment of horses), whence "dress," also "address," and "redress." "Adroit" and "maladroit" come from the French *droit* ("right," hence "right-handed," or dexterous), itself from the late Latin *directus*. When you ask someone to "direct" you, you're really asking which of *two* ways is *right*.

"To dine" originally meant "to breakfast" and so may be cognate with French *dejeuner*, which comes from the late Latin *disjejunare*, to break one's fast. The word *jejunare* has also given us "jejune" (originally, "fasting," later "insipid," "unsatisfying").

"Deuced bad luck, sir!" The "deuce," or two-spot, is the lowest valued card, and "snake's eyes" (a pair of ones) the lowest throw of a pair of dice. These bad luck aspects take us back to the "devil" again, whereas in tennis, "deuce" is the score when you need two points to win the game, or two games to win the set.

We've just used the word "pair." This comes from the Latin *par*, which is related to the Greek *para*, "beside." Of course, the commercial use, "at par," is immediately derived from this. Akin is the Greek *pornē*, a harlot, from "pairing person," which has given us "pornography," writing about harlots. Many mathematical and scientific words involve "par": "parallel," "parabola," "parity" (evenness or oddness, literally just "evenness").

The prefixes "ambi" and "amphi" come from Latin *ambo* and Greek *ampho* (both, either, double). An "amphitheatre" is a double theatre, and an "ambidextrous" person has "both right hands"—he is equally dexterous with each. The pejorative quality of two is again illustrated by "ambivalent," which, less than a century ago, just meant "two-valued."

3 The English county of Yorkshire is traditionally divided into North, East, and West "Ridings" ("thirdings"), a word now used more generally for "constituencies." In a similar way, the country near Rome was divided into the parts inhabited by the Roman, Sabine, and Alban tribes. It has been suggested that "tribe" (Latin *tribus*) means "one third part of." A "tribute" is a tax "contributed" by a tribe. A "tributary" is "one who brings tribute": the geographical sense dates only from 1836, and "distributary," in this sense, from 1886.

A "tribune" was the head of a tribe before whom "tribunals" were held; he might well "distribute" "retribution," having first heard "testamentary" evidence, from *ter-stis*, to be a third party, standing between accuser and accused. "Testimonial," "contest," "protest," and "Protestant" are related words.

However, the "tribulations" one might suffer from the deliberations of a tribunal are *not* associated with the number "three," but rather with rubbing and beating: compare "tribology," the study of friction, and "tribadism."

A "triumphal" procession is one after the style of the old Bacchic processions, accompanied by music played in triple-time. A "trump" card is one that helps you to "triumph."

The "trivium" was where three ways meet: where folks met to discuss "trivial" subjects. Later it was used for the three liberal arts: grammar, rhetoric, logic. In the standard medieval curriculum, these were considered "trivial" compared to those in the "quadrivium": arithmetic, music, geometry, and astronomy. At universities one studies for "trimesters" or terms, and at Cambridge University one takes the "tripos" examination: at one time the questions were asked by "the bachelor of the tripos," who sat on a three-legged stool. He might well have asked about the logical notion (*tertium non datur*) of the law of the excluded middle, that there is no *tertium quid* (third alternative to truth and falsehood).

Neptune's "trident" has three teeth, a "trammel" was originally a three-meshed net, "drill" is a three-threaded fabric (cf. "twill"), geology has both "tertiary" and "triassic" strata, while "tierce" has many meanings: (the office said at) the third hour of the canonical day; an old wine measure equal to one-third of a pipe; the third of the eight parries in sword-play; a run of three cards in piquet; the vertex of a pointed arch; a heraldic charge of three triangles; or the musical interval of a "third."

4 Workers in a "quarry" cut "square" stones. The word "square" is derived from the Latin *ex quadrem*, which has also given us "squadron," now often "squad," and which refers to troops drawn up in a battle square. Of course, the Romans used *quadrem* because it has four sides. "Quadrille" can be a "squadron," or a square dance for four couples, or a game for four players with forty cards, or a pattern of small squares.

In geometry, a "quadrilateral" is any four-sided figure, and you might call it a "quadrangle" if you counted its corners instead. Our college "quads" are quadrangles, and the slang "quod" for "prison" is probably the same word.

The "trapeze" artist takes his name from the Greek *trapezion*, a four-sided figure, or more literally, a little table (having four feet). The root words here are of great antiquity and are closely related in many languages. Thus the Hindi word *charpoy* for bed is much the same as "trapeze," as are the scientific terms "quadruped" for a generic four-legged animal and "tetrapod" for a division of butterflies. Of course, butterflies belong to the class Hexapoda (six legs), but the anterior pair of legs of Tetrapoda are unfit for walking.

The German *Karo* and French *carreau* for the diamond suit in cards describe their four-sided shape, and *tessera* and *tessella*, four-sided tile, has given us "tessellation."

In mathematics, "quadratic" equations are those that contain the "square" (4 sides) of the variable. Unfortunately, these equations have degree 2, so that in other mathematical contexts "quadratic" has come to mean "of second degree," and this has entailed the use of "biquadratic" (now more sensibly "quartic") for fourth-degree equations. "Cater-cornered" is really "quater-cornered."

"Quarter" is a fourth part, e.g., a "quartern" (loaf); and a "quart" is a quarter of a gallon; the English "firkin" and "farthing" (fourth-ing) are respectively a quarter of a barrel and a quarter of a penny. A "quire" of paper was originally 4 sheets, though later 24.

5 Your "fingers" are the five things that make up your "fist," and "fives" is the name of a game played with the hand. The early

alchemists recognized four basic elements (earth, air, fire, and water) from which all ordinary things were made, but the "quintessence" of ancient and medieval philosophy was the "fifth essence" of which the heavenly bodies were made.

Masefield's "quinquereme of Nineveh" had five rows of oars, as opposed to the "trireme." The word "quincunx" for the pattern ⁙ on a die involves both five and one; in the Roman system for fractions it was the symbol for five *uncia* (units, ounces, or twelfths).

A "pentagon" is named for its five corners, while a "pentagram" is drawn in five parts. This has also been called a "pentacle," referring to its five sharp points. "Punch" is a drink with five ingredients, from the Hindi word, *punch*, for five, which also gives us "Punjab," the land of the five rivers. The "Pentateuch" comprises the first five books of the Bible.

6 The "Sistine" Chapel was constructed by Pope Sixtus IV (1414–84), whose papal style, "Sixtus," was taken from that of Sixtus I (Pope ca. 116–125), who took the name to record that he was the sixth successor to St. Peter.

A "sextain" is the last six lines of a sonnet, while a "sestina" is a poem with six stanzas of six lines each; a Latin poem written in iambic "hexameters" had lines that were six measures long.

The position of a ship at sea was traditionally found by a "sextant," a navigational instrument in the shape of the sixth part of a circle. You take your "siesta" at midday, the sixth hour of the old 12-hour day. University "semesters" are from Latin *semestris* (six months); and Tennyson's lady,

clothed in white samite, mystic, wonderful,

was wearing a garment made from a six-threaded fabric (cf. "dimity").

7–10 September, October, November, and December take their names from the fact that they correspond to the seventh, eighth, ninth, and tenth months of the old Roman year which started in March. From 154 B.C. two new proconsuls were appointed each January 1, and this was gradually adopted for the year's beginning. In 44 B.C., Quinctilis, the old fifth month, was renamed after Julius Caesar, who had recently reformed the calendar and whose birthday was July 3. Sextilis was renamed Augustus, for Augustus Caesar, previously called Octavius, in 8 B.C. The names Septimus (Septima) and Octavius (Octavia) were customary for the seventh and eighth children in a family.

The Sabbath is the seventh day. "Hebdomadal" (weekly) is derived from "hebdomad," a variant of "heptad," from the Greek for a set of seven. The French and Italian words *semaine* and *settimana* are also obvious "seven" words (Latin *septimana*).

The first and last of eight consecutive white keys on a piano produce notes whose frequencies differ by a factor of two. This interval is called an "octave," although it consists of seven tones.

An "enneagon" is a nine-sided figure. "The Enneads" are the work of the philosopher Plotinus in Porphyry's edition (he collected them into six books of nine sections each).

The word "noon" has an interesting history. It comes from the Latin *nona*. The "nones" were prayers originally said at the ninth hour (3 p.m.) of the old 12-hour day, but which gradually drifted to midday. A "novena" is a devotional period of nine days. In the Roman calendar, the "nones" of a month was the ninth day before the "ides" of that month. The word "nonillion" was originally used for the ninth power of a million, but later it became the tenth power of a thousand.

The "Decalogue" consists of ten commandments, while Boccaccio's "Decameron" contains 100 tales that were supposed to be related in ten days. A "dicker" was a bartering unit of ten hides: Since traders would often argue down to the last "dime" (tenth of a dollar), it has come to mean "haggle." A "tithe," of course, is a tax of one tenth part.

A religious or scholarly "dean" takes his name from the Latin *decanus,* a leader of ten, which has also given us "doyen," the most senior person in a group. However, the apparently similar word "deacon" comes from the Greek *diakonos*, a servant (literally "one who passes through, raising the dust").

11–20

The English word "eleven" is "one left" (over ten) and "twelve" is "two left." "Dozen" is a corruption of the Latin *duodecim*. Your "duodenum" is so called because it is twelve inches long.

Hotel floors are often numbered 1, 2, . . . , 12,14,15, ... Hotel owners obviously expect that some of their customers will suffer from "triskaidekaphobia" (Greek for 3&10-fear, a morbid fear of the number 13).

The English words "thirteen" to "nineteen" derive from "three" to "nine" with "ten" in an obvious way, and there are analogs in most other languages. In England a "fortnight" is used for "two weeks" (= 14 nights), while in France "quinzaine" refers to the same period (of 15 days). A "quatorzaine" is like a sonnet, having 14 lines.

Carl Friedrich Gauss (1777–1855) showed that the regular "heptadecagon" (a 17-sided polygon) could be constructed with ruler and compasses.

"Score" is related to "share" and comes from the Old Norse "skor" meaning a "notch" or "tally" on a stick used for counting. The total count became the "score", still used in games and contests. Often people counted in 20s; every 20th notch was larger, and so "score" also came to mean 20. Carpenters still make score lines on wood they're about to saw, but they may not realize that the biblical use of "three score and ten" for man's allotted span of 70 years contains essentially the same word. In the sense of "cut off," the word "score" is cognate with both "shirt" and "skirt."

21–99

The card game of "pontoon" was brought home by English soldiers from World War I, where they had been speaking "Paddykelly patois." Americans call it "blackjack"—of course, "pontoon" is a corruption of the French *vingt-et-un*, 21.

"Quarantine" is an isolation period, originally of forty days (French *quarante*). "Pentecost" comes from the Greek name for the Jewish festival observed on the fiftieth day of Omer.

"Quadragesima" is the forty-day period of Lent, and "Quinquagesima" is the fifty-day period from Quinquagesima Sunday to Easter Sunday. So Quinquagesima Sunday is the seventh Sunday before Easter. "Sexagesima" and "Septuagesima" name the two previous Sundays, apparently as the result of a misunderstanding.

The "Septuagint" is the Greek version of the Old Testament, named after the "seventy" scholars who were commissioned by Ptolemy Philadelphus (284–247 B.C.) to translate it. (According to tradition, there were actually 72 translators, and their work took 72 days.)

The English words "twenty" to "ninety" arise in the obvious way, but in many languages there is a break in the system after "sixty." This has sometimes been interpreted as the survival of an old number system in which 60 was the base; compare the Babylonian system. Thus, in French,

70 is *soixante-dix*	$(60 + 10)$,
80 is *quatre-vingt*	(4×20),
90 is *quatre-vingt-dix*	$(4 \times 20 + 10)$,

although the regular formations *septante, huitante, nonante* have also been used.

English once treated these numbers in much the same way. Queen Elizabeth would certainly have said "four score" (cf. *quatre-vingt*) rather than "eighty," and of course our biblical lifespan is "three score and ten."

There have also been special words for "sixty" that survived in particular places. The English word "shock" for a collection of sixty things is still used for sixty sheaves of corn (a variant is "shook") and was once used for sixty pieces of haberdashery, or a German unit of account consisting of 60 Groschen. It has also been used for collections of other large sizes and might be the origin of the phrase "a shock of hair."

100 . . .

A Roman "centurion" was in charge of one hundred soldiers. "Centigrade" is the traditional name for the thermometric system devised by Celsius, with one hundred degrees between its two fixed points. A "cent" is a hundredth part of a dollar, a "century" contains one hundred years and a "millennium" one thousand. A "mile" consists of one thousand double paces (Latin *mille passus*).

Despite their names, "centipedes" have between 30 and 42 legs, while only a few "millipedes" have as many as 200. One might expect the proper scientific terms to be more cautious, but in fact the centipedes and millipedes together form the group *Myriapoda* ("ten thousand legs"). The centipedes alone form the order *Chiliapoda*, meaning "lip-legged": it is no more than a learned pun that this could also mean "thousand-legged."

The Greek word *hekaton* has given us "hectare" for one hundred units of area, "hecatomb" (Greek *bous*, ox) for the sacrifice of one hundred oxen, and "hectograph" for a machine that will make hundreds of copies of a drawing. The name of the Greek goddess Hecate is unrelated, but the hundredth asteroid to be discovered (in 1868) was named "Hecate," both after the goddess and as a pun on *hekaton*.

A British Member of Parliament cannot resign. Instead, the member applies for the Stewardship of the Chiltern Hundreds, which (theoretically) is an office of profit under the Crown, a disqualification from membership. The word "hundred" here is part of a county, usually supposed to consist of one hundred hides of land. This word was also used in the British American colonies and still survives in Delaware.

Etymologically, "hundred" is just an amplification of the Old English word *hund*, and the root here is common to all the Indo-European languages (Latin *centum*, Greek *hekaton*, Gothic *hund* from a supposed Indo-European root, *kentom*). The last letters in "thousand" have the same meaning: A "thousand" is literally a "strong hundred."

The numbers indicated by the various names have varied with the years. Today "hundred" always refers to the "short hundred" of "five score," but originally it meant the "long" or "great" hundred of "six score" or "two shocks" (120). In a similar way, a "hundredweight" has varied between 100 and 120 pounds, before stabilizing at 112 pounds in England or 100 pounds in the United States.

MILLIONS, BILLIONS, AND OTHER ZILLIONS

The Italians added an augmenting suffix to *mille* (Latin, "thousand") to obtain *millione* ("great thousand"), now milione, from which we get our "million." Around 1484, N. Chuquet coined the words *billion*, *trillion*, . . . , *nonillion*, which also appeared in print in a 1520 book by Emile de la Roche. These arithmeticians used "illion" after the prefixes

b, tr, quadr, quint, sext, sept, oct and non

to denote the

2nd, 3rd, 4th, 5th, 6th, 7th, 8th and 9th powers of a million. But around the middle of the 17th century, some other French arithmeticians used them instead for the

3rd, 4th, 5th, 6th 7th, 8th, 9th and 10th

powers of a thousand.

Although condemned by the greatest lexicographers as "erroneous" (Littré) and "an entire perversion of the original nomenclature of Chuquet and de la Roche" (Murray), the newer usage is now standard in the U.S., although the older one survives in Britian and is still standard in the continental countries (but the French spelling is nowadays "llon" rather than "llion").

These huge numbers were once just arithmeticians' playthings, but the march of science has forced people to find even more names for them. Here are the number-prefixes recommended by the Conference Générale des Poids et Mesures in 1991 (the number names are in the American system):

unit × N		unit / N		the number N		
deca	(da)	deci	(d)	10	=	ten
hecto	(h)	centi	(c)	100	=	hundred
kilo	(k)	milli	(m)	1000	=	thousand
mega	(M)	micro	(μ)	10^6	=	million
giga	(G)	nano	(n)	10^9	=	billion
tera	(T)	pico	(p)	10^{12}	=	trillion
peta	(P)	femto	(f)	10^{15}	=	quadrillion
exa	(E)	atto	(a)	10^{18}	=	quintillion
zetta	(Z)	zepto	(z)	10^{21}	=	sextillion
yotta	(Y)	yocto	(y)	10^{24}	=	septillion

How do the "illion" words continue? We shall use "zillion" for the typical one of them, so that the N'th zillion is 10^{3N+3} (American) or 10^{6N} (British). The first 9 "zillion" names are Chuquet's "million", "billion", "trillion", "quadrillion", "quintillion", "sextillion", "septillion", "octillion", "nonillion", and "centillion" is already a well-established word for the 100th one. You can find a name for any zillion from the 10th to the 999th by combining parts from the appropriate columns of the following table, and then replacing the final vowel by "illion":

	units	tens	hundreds
1	un	^{n} deci	^{nx} centi
2	duo	^{ms} viginti	^{n} ducenti
3	tre (*)	^{ns} triginta	^{ns} trecenti
4	quattuor	^{ns} quadraginta	^{ns} quadringenti
5	quinqua	^{ns} quinquaginta	^{ns} quingenti
6	se (*)	^{n} sexaginta	^{n} sescenti
7	septe (*)	^{n} septuaginta	^{n} septingenti
8	octo	^{mx} octoginta	^{mx} octingenti
9	nove (*)	nonaginta	nongenti

*Note: when it is immediately before a component marked with ^{s} or ^{x}, "tre" increases to "tres" and "se" to "ses" or "sex" as appropriate. Similarly "septe" and "nove" increase to "septem" and "novem" or "septen" and "noven" immediately before components marked with ^{m} or ^{n}.

With Allan Wechsler we propose to extend this system indefinitely ([1]) by combining these according to the convention that "XilliYilliZillion" (say) denotes the (1000000X + 1000Y + Z)th zillion, using "nillion" for the zeroth "zillion" when this is needed as a placeholder. So for example the million-and-third zillion is a "millinillitrillion."

You can now use the usual rules for combining this complete system of zillion words (which first appears in the present *Book of Numbers*) so as to obtain correct 'English names', like

"four millinillitrillion (and) fifteen,"

meaning

$$4.10^{3000012} + 15 \text{ (American) or } 4.10^{6000018} + 15 \text{ (British),}$$

for all of the integers!

[1]Our proposal does have one sad effect. Donald Knuth's number "eight billion, eighteen million, eighteen thousand eight hundred fifty-seven" is probably no longer the alphabetically earliest prime number in the American system.

We close with a table of English words for powers of ten.

ten or decad	10^1
hundred or hecatontad	10^2
thousand or chiliad	10^3
myriad	10^4
lac or lakh	10^5
million	10^6
crore	10^7
myriamyriad	10^8
milliard	10^9
billion	10^9 (formerly 10^{12})
trillion	10^{12} (formerly 10^{18})
quadrillion	10^{15} (formerly 10^{24})
decillion	10^{33} (formerly 10^{60})
vigintillion	10^{63} (formerly 10^{120})
centillion	10^{303} (formerly 10^{600})
googol	10^{100}
googolplex	10^{googol}
*N*plex	10^N (Rudy Rucker)
*N*minex	10^{-N} (Tadashi Tokieda)

We recommend "Nplex" and "Nminex" both as the nouns for those numbers and as adjectives indicating multiplication by them. So "hundredplex" is another name for a googol, the mass of the Sun is roughly two 33plex grams, and an Angstrom unit is a 10minex meter.

Some other cultures have more interesting names for these numbers. Japanese has individual names for various powers of ten between 10^{-32} and 10^{88}, many of Indian origin. Some of these are

88plex	= muryoutaisuu	"large amount of nothing"
80plex	= fukashigi	"don't even think about it"
56plex	= kougasha	"sands of the Ganges"
16minex	= shunsoku	"breathing instant"
17minex	= danshi	"finger snapping"
23minex	= jou	"clean"

HOW NUMBERS ARE WRITTEN

There have been many different number names, notations, and symbols in the world's history. We'll only describe a few that have had significant influence on Western civilization.

Perhaps the earliest recorded occurrences of numerals are on some Sumerian clay tablets dating from the first half of the third millennium B.C. The Sumerian system was later taken over by the Babylonians. The surviving Babylonian documents are mostly from two periods: Old Babylonian, from around the time of Hamurabi ≈ 1500 B.C. and the Seleucid era, from 300 B.C.. The Babylonian cuneiform (Latin *cuneus* wedge) script was formed by impressing wedge-shaped marks in clay tablets. Their notation for numbers used the base 60, with individual symbols, 𒁹 and 𒌋, for 1 and 10; there is a famous example in Chapter 7. Another is Plimpton 322, a remarkable table of "Pythagorean triples (see Chapter 6). This system is very easy to read and is an early instance of place notation. For instance, the symbols 𒌋𒐊 and 𒎙𒐈 mean 15 and 23, while 𒌋𒐊 𒎙𒐈 would read $(15 \times 60) + 23 = 923$.

It is because the ancients made astronomical calculations in base 60 that we still use this system for measuring time, dividing an hour into 60 minutes, and a minute into 60 seconds. In its path through the heavens, the sun takes roughly 360 days (actually 365.242199) to describe a complete circle, so it seems that the Babylonians divided a complete circle into 360 degrees (°). In the Babylonian manner, each degree is divided into 60 minutes (′), each minute into 60 seconds (″), each second into 60 thirds (‴), and so on.

Again, it is because mathematical and astronomical tables written in this notation were so hard to recompute that a modern astronomer writes an angle $25°32'14''$ just as his Babylonian forebears wrote

𒎙𒐊 𒌍𒐖 𒌋𒐉

and similarly, that you say that the time is 9:45 (9 hours, 45 minutes) rather than 9.75 hours.

We see that the Babylonians used place value, the symbol 𒁹 meaning 1 or 60 or 60^2 or . . . , according to its position. This could be confusing, because they had no zero. Occasionally they left a space, and later the Seleucids introduced the symbol 𒑱. Unfortunately, place value was not used in the clumsier Greek and Roman systems and only reappeared in our own system, using the Hindu-Arabic notation. The Egyptians used a different system, which we'll discuss when we describe the Roman system, to which it is very similar.

THE GREEK SYSTEM

The Greeks, from about the fifth century B.C., used the notation illustrated in Figure 1.1.

Since their alphabet had only 24 letters and 27 were needed, they resurrected three letters of Semitic origin, namely digamma (F) or vau (ϛ), qoph or koppa (ϙ, Ϟ) and san or sampi (ϡ) to represent 6, 90, and 900. There were various systems for numbers larger than a myriad (10,000). Diophantus, around the third century A.D., used a dot to indicate that the preceding numbers should be multiplied by 10,000. He gave the following example.

,α τ λ α . ,ε σ ι δ
1 3 3 1 5 2 1 4

α	β	γ	δ	ε	F	ζ	η	θ
1	2	3	4	5	6	7	8	9
ι	κ	λ	μ	ν	ξ	ο	π	ϙ
10	20	30	40	50	60	70	80	90
ρ	σ	τ	υ	φ	χ	ψ	ω	ϡ
100	200	300	400	500	600	700	800	900
,α	,β	,γ	,δ	,ε	,F	,ζ	,η	,θ
1000	2000	3000	4000	5000	6000	7000	8000	9000

FIGURE 1.1 *The Greek number system.*

THE ROMAN SYSTEM

The fact that we are still familiar with the Roman numerals (sometimes used on clock faces, monuments, and some documents) is because they were the only ones in use in Europe for well over a thousand years. The Roman number system was based on a similar one used by the Etruscans, the letters I, V, X, L, C, D, M being based on the Etruscan originals. The Romans assimilated the earlier C and ↀ for 100 and 1,000 to the Latin letters C and M, which happened to be the initial letters of *centum* and *mille*. Half of ↀ was simplified to IↃ or D for 500. The symbols

	ↁ	ↂ	ↇ	ↈ
for	5000	10000	50000	100000

are fortunately no longer used, although twenty-two of the original 32 copies of ↈ can still be seen on the *Columna Rostrata*, describing the capture of 3,200,000 prisoners.

The symbol ∞, also a corruption of ↀ, was occasionally used as an alternative for 1,000. In 1655 John Wallis (1616–1703) proposed that this symbol be used for "infinity," and mathematicians have used it ever since.

The subtractive principle, whereby 9 and 40 are written as IX and XL, only became popular in medieval times. It was hardly used by the Romans.

Roman fractions were based on twelfths ("uncia" from "unus," i.e., one of the lesser units); we discuss them later.

THE EGYPTIAN SYSTEM

The Egyptians had earlier used a similar system to the Roman one, but with fewer symbols: Figure 1.2 compares the two systems.

The Romans would have had to write

ↈ ↈ ↈ ↈ ↈ ↈ ↈ ↈ ↈ ↈ

for a million. The Egyptians had a much nicer symbol: a man with his arms uplifted:

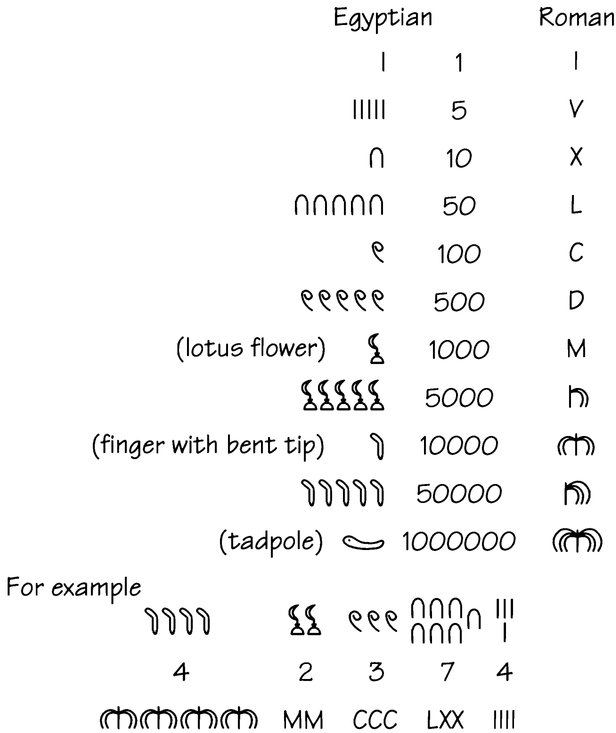


FIGURE 1.2 *The Egyptian and Roman number systems.*

HINDU-ARABIC NUMERALS

The numeration we use now, in which numbers are formed by juxtaposing the ten digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 is often referred to as Arabic notation. In fact it is of Hindu origin and was transmitted to Europe, like a lot of the world's knowledge, by Arab scholars.

The value of a digit depends on its position in the system (its place value) so that one needs a symbol for zero, to distinguish 904 from 94, for example. It was in this way that the concept of zero forced itself onto the Indian mathematicians. Theoretically, zero is also needed occasionally in the Babylonian system, but as the base is so much larger, the context would usually supply the missing information, and, as we've seen, the Babylonians struggled on without it for over a thousand years.

The system is so familiar to us that we regard it as giving *the* names for numbers, as being what a number really is! When a date is given in Roman numerals, for example, most of us painfully translate it into the Arabic system. Originally it was the other way around! In medieval times the merchants of Europe were thoroughly familiar with the Roman system and found the new Arabic system very confusing, and early examples often contain mistakes or mix the two systems.

M50iv = 1504

Lowercase numerals were also introduced in medieval times, and it was not uncommon to see mcdxxvi = 1426. Nowadays, this system of lowercase numerals only survives in special places such as subsections of a list or in numbering the preliminary pages of a book.

NUMBERS IN OTHER BASES

Of course, a similar system of notation can be defined using *any* whole number, N, as a base. In the base N system, "*abcd*" means

$$aN^3 + bN^2 + cN + d,$$

and the "digits" a, b, c, d customarily range from 0 to $N-1$.

Our Hindu-Arabic system is also called *decimal* (base 10); the Babylonian system was *sexagesimal* (base 60); and most computers internally use the *binary* system (base 2). The octal (base 8) and hexadecimal (base 16) systems are easily convertible to the binary system and so are often used by computer programmers. Numbers written in base 2 are often called "binary numbers." Of course, binary numbers are not a new kind of number, they're just the old numbers under a different name:

$$1,\ 10 = 2,\ 11 = 3,\ 100 = 4,\ 101 = 5,$$
$$110 = 6,\ 111 = 7,\ 1000 = 8, \ldots$$

KINDS OF NUMBERS

We've organized this book around different kinds of numbers: the following brief survey introduces them.

The familiar **whole numbers**, 1, 2, 3, . . . (Chapter 2) have a history that starts even before our own. Certain species of bird can detect that one of the eggs in their nest has been removed. Presumably they have some primitive idea of how many eggs there should be. An experiment described by Tobias Dantzig shows that crows could recognize when up to four men with guns had come and gone. The birds kept a safe distance in the meantime but could be fooled if as many as five came and only four went.

However, many primitive human languages only have names for numbers of particular objects, and not for the *idea* of numbers. The Fiji Islanders use "bolo" for ten boats, but "koro" for ten coconuts and "salora" for one thousand coconuts. The divorce between the abstract concept of number and the objects being counted took a

$\frac{1}{12}$	•	uncia	from *unus*, "*one*" (smaller unit)
$\frac{1}{6} = \frac{2}{12}$	••	sextans	one-sixth (of the larger unit)
$\frac{1}{4} = \frac{3}{12}$	•••	quadrans	one-quarter
$\frac{1}{3} = \frac{4}{12}$	::	triens	one-third
$\frac{5}{12}$	⁙	quincunx	five uncia
$\frac{1}{2} = \frac{6}{12}$	S	semis	one-half
$\frac{7}{12}$	S•	septunx	seven uncia
$\frac{2}{3} = \frac{8}{12}$	S••	bes	dues partes asis
$\frac{3}{4} = \frac{9}{12}$	S•••	dodrans	de quadrans (a quarter short)
$\frac{5}{6} = \frac{10}{12}$	S::	dextans	de sextans
$\frac{11}{12}$	S⁙	deunx	de uncia
$1 = \frac{12}{12}$	I	as	the next unit

FIGURE 1.3 *Roman fractions.*

long time, and evidence is still visible; a couple (of people), a brace (of pheasants), a century (of years, or runs at cricket).

Special families of numbers are discussed in Chapters 3 and 4. A particularly interesting family is the prime numbers, whose intriguing properties demand a chapter of their own, Chapter 5.

Fractions, or **rational numbers** (Chapter 6), are surely a human discovery, whose origins are lost in antiquity. However, most of the early systems named only a few obvious common fractions.

The Rhind papyrus, dated from 1650 B.C., had simple names only for the unit fractions, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$, . . . , and for $\frac{2}{3}$ (). Other fractions, when required, were made by adding these, e.g., $\frac{4}{5} = \frac{1}{2} + \frac{1}{5} + \frac{1}{10}$ or $\frac{3}{4} = \frac{1}{2} + \frac{1}{4}$, although some early papyri have  for $\frac{3}{4}$.

The Roman notation for fractions is not generally very well known. Their unit was customarily divided into twelve smaller units, and they had names and symbols for all the multiples of these (Figure 1.3).

The fact that there are **irrational numbers**, that are not fractions, came as a great surprise to the Greeks and is still probably unfamiliar to most of the world's inhabitants. Traditionally, this discovery is credited to Pythagoras, who found that the diagonal of a square is not a rational multiple of its side; that is, the ratio of diagonal to side cannot be expressed by whole numbers. The Pythagorean brotherhood is supposed to have sacrificed a hecatomb, or a hundred oxen, in honor of this stupendous discovery. The irrational number involved is one of the **algebraic numbers** treated in Chapter 7.

We've already seen how the problems of notation brought the new number **zero**, 0, to the attention of mathematicians. It was the introduction of algebra (solve $2x + 7 = 3$) that forced us to recognize **negative numbers**. Initially, negative numbers were regarded as very mysterious things, although the advances of science and technology have made them very familiar to us.

Algebra also led to the introduction of **complex numbers**, which are even more puzzling. We remove some of the mystery in Chapter 8.

On the other hand, algebraic equations do not suffice to define all the numbers of interest to us, and we are forced to recognize **transcendental numbers**, whose existence was first shown by

τ =	1.6180339887	4989484820	4586834365	6381177203	0917980576	2862135448-	
	6227052604	6281890244	9707207204	1893911374	8475408807	5386891752	. . .
$\sqrt{2}$ =	1.4142135623	7309504880	1688724209	6980785696	7187537694	8073176679-	
	7379907324	7846210703	8850387534	3276415727	3501384623	0912297024	. . .
$\sqrt{3}$ =	1.7320508075	6887729352	7446341505	8723669428	0525381038	0628055806-	
	9794519330	1690880003	7081146186	7572485756	7562614141	5406703029	. . .
π =	3.1415926535	8979323846	2643383279	5028841971	6939937510	5820974944-	
	5923078164	0628620899	8628034825	3421170679	8214808651	3282306647	. . .
e =	2.7182818284	5904523536	0287471352	6624977572	4709369995	9574966967-	
	6277240766	3035354759	4571382178	5251664274	2746639193	2003059921	. . .
γ =	0.5772156649	0153286060	6512090082	4024310421	5933593992	3598805767-	
	2348848677	2677766467	0936947063	2917467495	1463144724	9807082480	. . .
ln 10 =	2.3025850929	9404568401	7991454684	3642076011	0148862877	2976033327-	
	9009675726	0967735248	0235997205	0895982983	4196778404	2286248633	. . .
ln 2 =	0.6931471805	5994530941	7232121458	1765680755	0013436025	5254120680-	
	0094933936	2196969471	5605863326	9964186875	4200148102	0570685733	. . .
$\log_{10} 2$ =	0.3010299956	6398119521	3738894724	4930267681	8988146210	8541310427-	
	4611271081	8927442450	9486927252	1181861720	4068447719	1430995379	. . .
$\log_2 3$ =	1.5849625007	2115618145	3738943947	8165087598	1440769248	1060455752-	
	6545410982	2779435856	2522280474	9180882420	9098066247	5059167343	. . .
F_1 =	4.6692016091	0299067185	3203820466	2016172581	8557747576	8632745651	. . .
F_2	2.5029078750	9589282228	3902873218	2157863812	7137672714	9977336192	. . .

FIGURE 1.4 *Some of our favorite numbers.*

Liouville in 1840. It had long been suspected that the notorious number π was transcendental, although this was only finally established by Ferdinand Lindemann in 1882. We'll read about π and other transcendental numbers in Chapter 9.

Children sometimes ask "How many numbers are there?" A coherent theory of **infinite numbers** was first developed by Georg Cantor from 1872 to 1884. The number of whole numbers is Cantor's $\aleph_0$, as is the number of algebraic numbers, while the number of transcendental numbers is $2^{\aleph_0}$. In fact, Cantor introduced two distinct systems of infinite numbers, **cardinal numbers** and **ordinal numbers**, and we explain them in Chapter 10. These are generalizations of the ordinary whole numbers, 0, 1, 2, 3, ... The appropriate generalization for fractional numbers and other real numbers was discovered only recently, by one of us. The resulting **surreal numbers** are also described in Chapter 10.

SOME SPECIAL FRIENDS

As we explore the different kinds of numbers, we can't help becoming familiar with interesting individual numbers and families of numbers. On the way we hope that you'll make friends with many particular numbers.

Two is celebrated as the only even prime, which in some sense makes it the oddest prime of all. We'll show you how to construct a regular **seventeen**-sided polygon with ruler and compass, and we'll tell you what's special about **one hundred and sixty-three**.

In addition to **Pythagoras's number** $\sqrt{2}$ and Ludolph's number π, geometry has given us the **golden number** τ, while analysis produces **Napier's number** e and the **Euler-Mascheroni** number γ. Special numbers like these continue to appear as our knowledge expands; two new ones are **Feigenbaum's numbers** F_1 and F_2 (see Chapter 7). Some of our favorite numbers are displayed in Figure 1.4.

The last two items were taken from Simon Plouffe's Table of Constants, under construction in connexion with the Inverse Symbolic Calculator mentioned below. What are your favorite numbers? What do you do if you come across a number and don't know what it is?

We quote from the Preface of the Borweins' book:

> How do we decide that such a number is actually a special value of a familiar function without the tools Gauss had at his disposal, which were, presumably, phenomenal insight and a prodigious memory?
>
> *Jonathan Borwein & Peter Borwein*

The Borweins are currently building an online, internet-based, number checker. It is like a spell checker. You put your number in and it confirms hits. They call it the "Inverse Symbolic Calculator" and it can be accessed at http://www.cecm.sfu.ca/projects/ISC.html

REFERENCES

Asger Aaboe. *Episodes from the Early History of Mathematics*, New Math. Library 13, Math. Assoc. of America, 1964.

J. Borwein & P. Borwein. *A Dictionary of Real Numbers*, Wadsworth, 1990.

Lucas N. H. Bunt, Phillip S. Jones, and Jack D. Bedient. *The Historical Roots of Elementary Mathematics.* Prentice-Hall, Englewood Cliffs, NJ, 1976.

A. B. Chace, and others. *The Rhind Mathematical Papyrus.* Math. Assoc. of America, Oberlin, OH, 1927–1929.

John H. Conway, *On Numbers and Games*, Academic Press, London, 1976.

Tobias Dantzig. *Number, the Language of Science*, Doubleday Anchor, 1956.

H.-D. Ebbinghaus, H. Hermes, F. Hirzebruch, M. Koecher, K. Mainzer, J. Neukirch, A. Prestel & R. Remmert. *Numbers*, Springer-Verlag, Graduate Texts in Mathematics, 123, 1991.

Richard J. Gillings. *Mathematics in the Time of the Pharaohs.* M.I.T. Press, Cambridge, MA, 1972.

Donald E. Knuth, *Surreal Numbers*, Addison-Wesley, Reading MA, 1974.

Karl Menninger. *Number Words and Number Symbols. A Cultural History of Numbers,* trans. Paul Broneer from rev. German ed., M.I.T. Press, Cambridge, MA, 1969, 1977.

Otto Neugebauer. *The Exact Sciences in Antiquity*, Brown Univ. Press, 1957; reprinted Dover, 1969.

Eric Partridge. *Origins, A Short Etymological Dictionary of Modern English.*

Rudy Rucker. *Mindtools*, Houghton Mifflin Company, Boston, 1987.

David Eugene Smith & Jekuthiel Ginsberg. *Numbers and Numerals,* and see James R. Newman, *The World of Mathematics*, George Allen & Unwin, London, 1960, pp. 442–464.

2

Figures from Figures Doing Arithmetic and Algebra by Geometry

Patterns Provide Pretty Proofs

We can learn a good deal of arithmetic just by writing the numbers in rows of 1, 2, 3, . . . , as in Figure 2.1. The left-hand column in each section is the list of **multiples** of the number of entries in each row.

When we write just two numbers in each row, the left column contains the **even numbers** (Figure 2.2) and the right column contains the **odd numbers** (Figure 2.3).

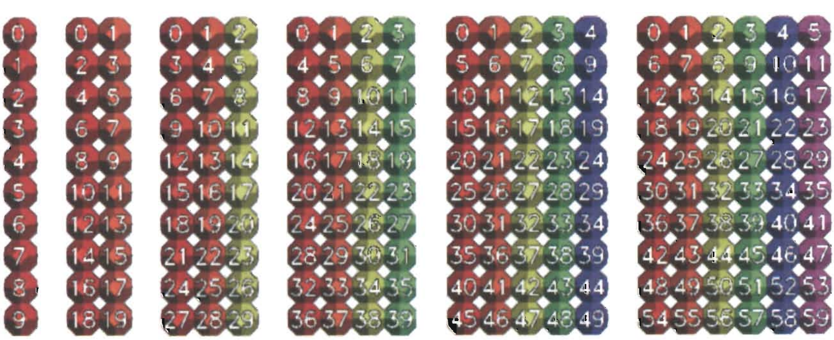


FIGURE 2.1 *Writing numbers in rows reveals the residue classes in columns.*

FIGURE 2.2 *The even numbers; the multiples of two.*

FIGURE 2.3 *The odd numbers.*

The columns of numbers in Figure 2.1 are the **residue classes**, or sets of numbers that leave the same **remainder** (or residue) when divided by the number of numbers in a row. In our figures the residue classes are distinguished by different colors. For example, the middle (orange) column in the third section of Figure 2.1 contains numbers that are one more than a multiple of 3, that is, numbers that are **congruent to 1 modulo 3**. One of the many great contributions of Carl Friedrich Gauss (1777–1855) to number theory was the arithmetic of residue classes. We say that two numbers are **congruent modulo** n when their difference is divisible by n.

For example, in 1938, May 3, May 10, May 17, May 24, May 31 were Tuesdays; the numbers 3, 10, 17, 24, 31 are all congruent modulo 7.

CASTING OUT NINES

A three-line version of the equals sign is used to denote congruence. For example, modulo 9,

$$1 \equiv 10 \equiv 100 \equiv 1000 \equiv 10000 = \dots,$$

and this is the basis of ***preuve par neuf***, or "casting out the nines," a useful arithmetic check. To cast out nines from a number, just add its digits, subtracting 9 whenever you can. To check your additions, subtractions, and multiplications, repeatedly cast out nines: they should remain valid. For example, we get five by casting out nines from each of 239 and 4649, so for their product we should obtain

$5 \times 5 = 25 \equiv 2 + 5 = 7$, agreeing with $239 \times 4649 = 1111111 \equiv 7$. You can test the multiplication $127 \times 9721 = 1234567$ similarly. Now suppose we wish to check that

$$2^{32} + 1 = 4294967297 = 641 \times 6700417.$$

The sums of the digits of the two factors are 11 and 25, and the sums of ***their*** digits are 2 and 7, whose product is 14, the sum of whose digits is 5. The sum of the digits of the big number is 59, which likewise gives 14 and then 5. The agreement doesn't *guarantee* the answer, but we can be sure that there's no copying error of just one digit, except that it is possible that a 0 was swapped with a 9, or vice versa. The test does not detect a mistake in the ***order*** of the digits. You can also check the first equality with five squarings ($32 = 2^5$): $2^2 = 4$, $4^2 = 16 \equiv 7$, $7^2 = 49 \equiv 4$, $4^2 = 16 \equiv 7$, $7^2 \equiv 4$ and $4 + 1 = 5$.

COLORS REVEAL PATTERNS

When we color one section of Figure 2.1 with the colors from another, the residue classes make regular patterns. If the length of

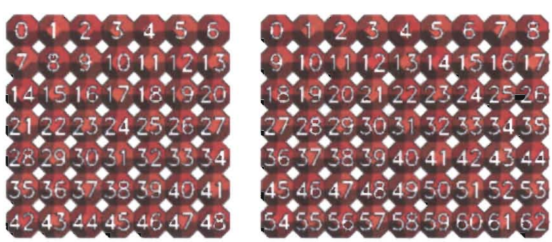


FIGURE 2.4 *The odd (orange) and even (red) numbers form checkerboard patterns.*

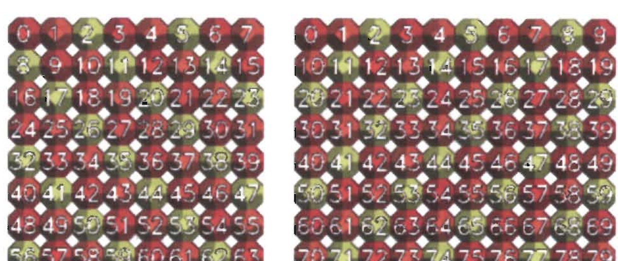


FIGURE 2.5 *The residue classes modulo 3: 0 (red), 1 (orange), 2 (yellow).*

FIGURE 2.6 *The residue classes modulo 4: 0 (red), 1 (orange), 2 (yellow), 3 (green).*

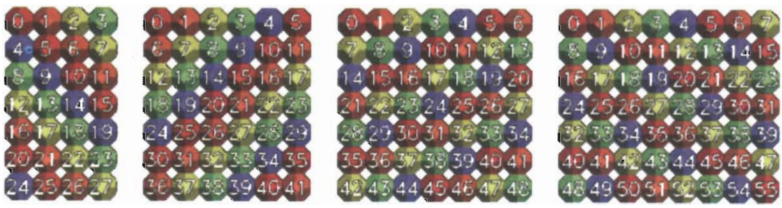


FIGURE 2.7 *The residue classes modulo 5: 0 (red), 1 (orange), 2 (yellow), 3 (green), 4 (blue).*

the row is odd, then the odd (orange) and even (red) numbers form a checkerboard pattern (Figure 2.4).

If the **modulus**, that is, the number of columns, is not a multiple of 3, the (red, orange, yellow) residue classes modulo 3 appear as diagonal stripes (Figure 2.5).

Modulo 4 the (red, orange, yellow, green) residue classes form columns, or diagonals (left and right of Figure 2.6), or (when the row length is a **singly even number**, $4n + 2$, such as 6 or 10; i.e., divisible by 2, but not by 2^2) a design reminiscent of knights' moves in chess (middle of Figure 2.6).

The knights' moves are especially noticeable for the residue classes modulo 5, whenever the row length is congruent to ± 2 modulo 5 (two rightmost sections of Figure 2.7).

SQUARE NUMBERS

If we collect the multiples of 1, 2, 3, . . . , that is, the first columns from each part of Figure 2.1, into a single table, we have the **multiplication table** (Figure 2.8), whose main diagonal gives us the **square numbers** (Figure 2.9).

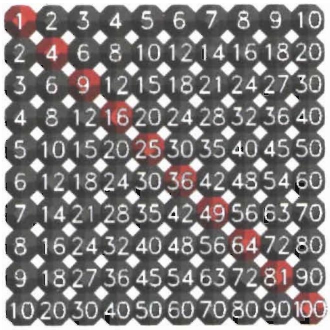
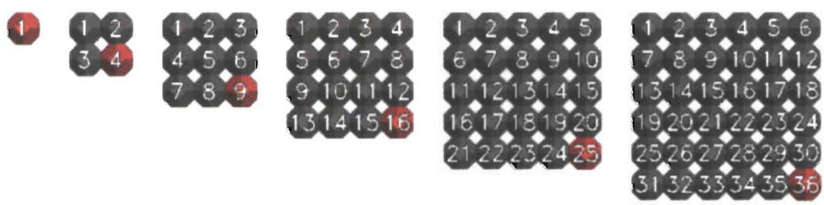


FIGURE 2.8 *The multiplication table has the square numbers as its leading diagonal.*



$1^2=1$ $2^2=4$ $3^2=9$ $4^2=16$ $5^2=25$ $6^2=36$

FIGURE 2.9 *The square numbers.*

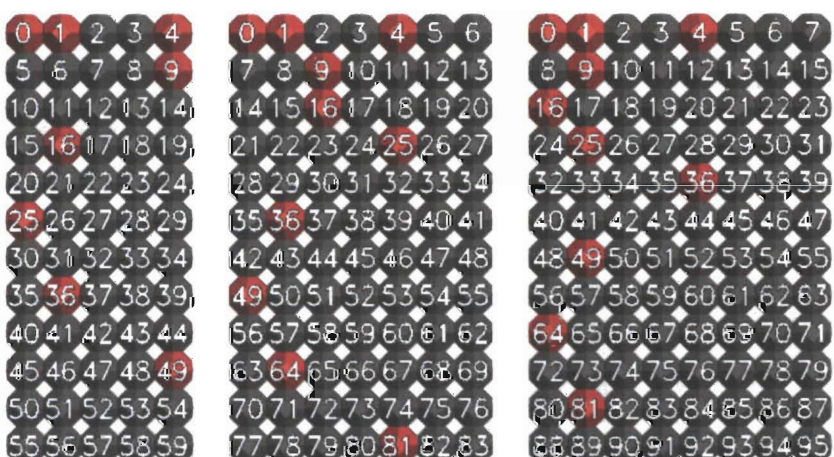


FIGURE 2.10 *Quadratic residues: the residue classes in which the squares lie.*

In which residue classes do the square numbers lie? Figure 2.10 shows us the answers for moduli 5, 7, and 8. Figure 2.10(a) is the same as Figure 2.1(e), with just the square numbers colored red.

From Figure 2.10(c) we see that

The odd squares are congruent to 1 modulo 8.

We'll see another neat proof of this after we've met the triangular numbers (Figure 2.23).

Figure 2.11 shows you that the squares are always congruent to 0, 1, or 4 modulo 5 (count the red blobs).

You may have noticed in Figure 2.10 that the squares lie on **parabolas**, which are the curves you get when you draw graphs of **quadratic** expressions, or **polynomials of degree two**, such as the algebraic expressions for the squares, triangular numbers, and other "two-dimensional" figurate numbers that we'll soon meet. But before we leave the squares, look at Figure 2.12.

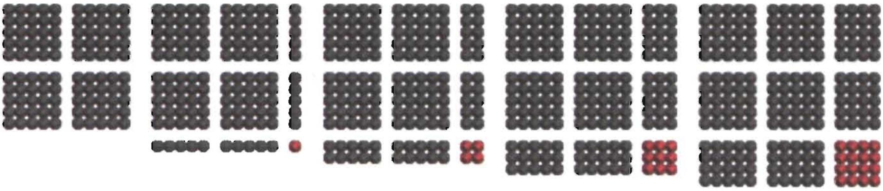


FIGURE 2.11 *Squares are congruent to 0, 1, or 4 modulo 5.*

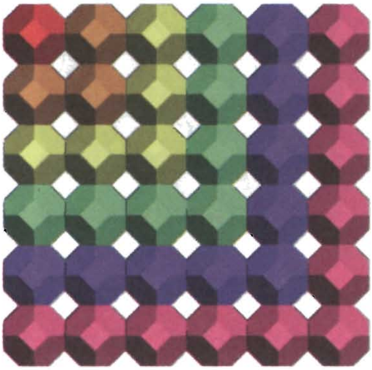


FIGURE 2.12 *Each gnomon contains an odd number of blobs of one color.*

The ancient Greeks called the piece that you can add to a figure to produce a larger figure of the same shape a **gnomon** (knower) after the shape of the gnomon of a sundial (time-knower). The gnomons of Figure 2.12 fit together to show that the sum of the first n odd numbers is the nth square, n^2. The addition of one more gnomon would illustrate the identity

$$n^2 + (2n + 1) = (n + 1)^2.$$

TRIANGULAR NUMBERS

Now write the numbers in rows of increasing length (Figure 2.13), so that the gnomons are just the rows (Figure 2.14). This gives us the **triangular numbers**.

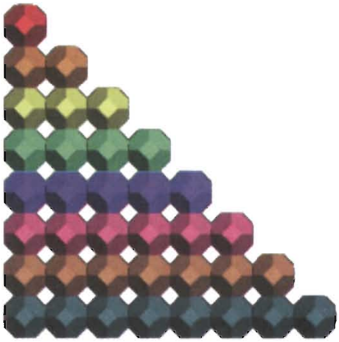


FIGURE 2.13 *Rows of increasing length yield the. . .*

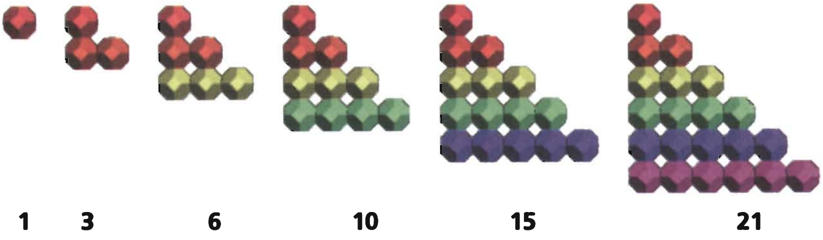


FIGURE 2.14 *. . . Triangular numbers.*

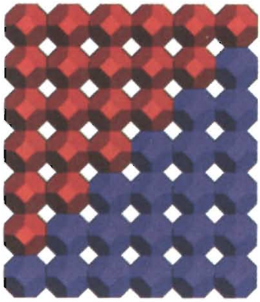


FIGURE 2.15 *Twice the triangular number Δ_n is the **pronic number** $n(n + 1)$, the product of two consecutive integers.*

We'll call the nth triangular number Δ_n. How big is it? If we put two triangular numbers of side n together, they form a rectangle, $n + 1$ by n, so the answer is $\frac{1}{2}n\,(n + 1)$. Figure 2.15 shows the case for $n = 6$.

Some readers will prefer the algebraic proofs of this fact, by induction. To prove that $1 + 2 + \cdots + n = \frac{1}{2}n(n + 1)$, we check the starting case and then suppose the result for the previous number, namely

$$1 + 2 + \cdots + (n - 1) = \frac{1}{2}(n - 1)n = \frac{1}{2}n^2 - \frac{1}{2}n.$$

Adding n to both sides, we do indeed obtain

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

Since for two consecutive numbers one is odd and the other is even, it's no surprise that their product is always divisible by 2.

We can similarly show that the sum of the first n odd numbers is n^2, by deducing

$$\begin{aligned} 1 + 3 + 5 + \cdots + (2n - 1) + (2n + 1) &= n^2 + 2n + 1 \\ &= (n + 1)^2 \end{aligned}$$

from $\quad 1 + 3 + 5 + \cdots + (2n - 1) = n^2.$

Alternatively, we could have used the same device as in Figure 2.15 to show this (Figure 2.16). In fact this "organ-pipe" method

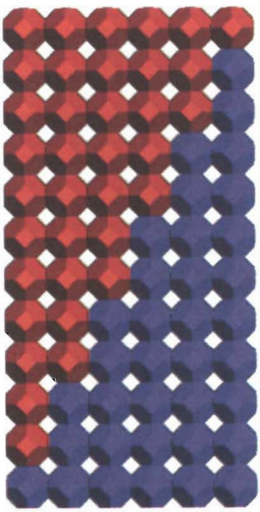


FIGURE 2.16 *The sum of the first* n *odd numbers is the square* n^2.

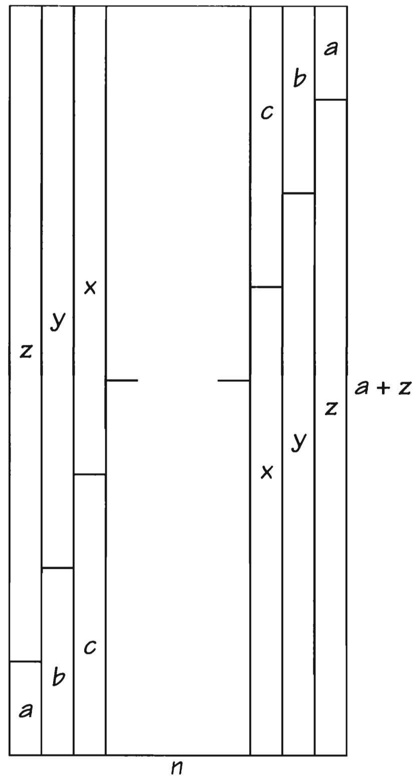


FIGURE 2.17 *Two copies of* $a + b + c + \cdots + x + y + z$ *makes a rectangle* $a + z$ *by* n.

(Figure 2.17) finds the sum of any **arithmetic progression**, or sequence of n equally spaced numbers

$$a, b, c, \ldots, x, y, z;$$

> The sum of n equally spaced numbers with first term a and last term z is n times their average:
>
> $$n \times \frac{a + z}{2}.$$

For instance the 10 term sum

$$5 + 8 + 11 + \cdots + 26 + 29 + 32 = 10 \times \frac{5 + 32}{2} = 185.$$

This works because $a + z = b + y = c + x = \cdots$.

So you can check again that the sum of 1, 2, 3, 4, . . . consecutive odd numbers, starting with 1, gives the squares (Figure 2.18(a)). If, instead of starting with 1 each time, we start where we previously left off, as in Figure 2.18(b), then we get the cubes!

(a)		(b)	
1	$= 1^2$	1	$= 1^3$
$1 + 3$	$= 2^2$	$3 + 5$	$= 2^3$
$1 + 3 + 5$	$= 3^2$	$7 + 9 + 11$	$= 3^3$
$1 + 3 + 5 + 7$	$= 4^2$	$13 + 15 + 17 + 19$	$= 4^3$
$1 + 3 + 5 + 7 + 9$	$= 5^2$	$21 + 23 + 25 + 27 + 29$	$= 5^3$
$1 + 3 + 5 + 7 + 9 + 11$	$= 6^2$	$31 + 33 + 35 + 37 + 39 + 41$	$= 6^3$
$1 + 3 + 5 + 7 + 9 + 11 + 13$	$= 7^2$	$43 + 45 + 47 + 49 + 51 + 53 + 55$	$= 7^3$

FIGURE 2.18 *Adding odd numbers gives squares or cubes.*

So, if we add up all the odd numbers in the first n rows of Figure 2.18(b), we see that the sum of the first n cubes is equal to the sum of the first Δ_n odd numbers, which we know is Δ_n^2:

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

If we had guessed this result, we could also deduce it inductively from the previous case by adding n^3 to both sides of

$$1^3 + 2^3 + \cdots + (n - 1)^3 = \left(\frac{1}{2}(n - 1)n\right)^2$$
$$= \frac{1}{4}n^4 - \frac{1}{2}n^3 + \frac{1}{4}n^2$$

Figure 2.19 shows that the sum of two *consecutive* triangular numbers is a square; Figure 2.20 has been colored in two ways, one

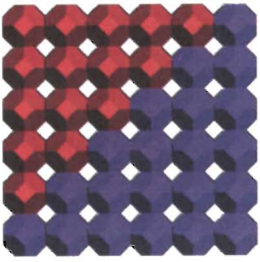


FIGURE 2.19 *Two consecutive triangles make a square.*

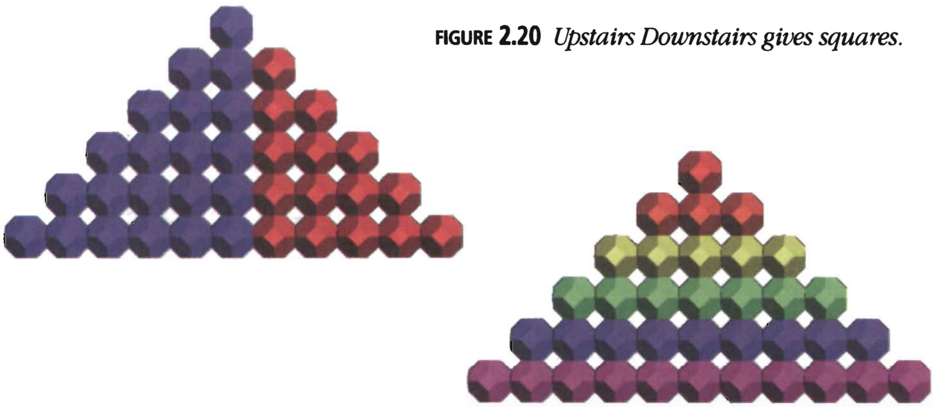


FIGURE 2.20 *Upstairs Downstairs gives squares.*

to show the two triangular numbers, and the other to show the consecutive odd numbers. In symbols, Figure 2.20 may be written as

$$\begin{aligned}\Delta_n + \Delta_{n-1} &= 1 + 2 + 3 + \cdots + n \\ &\quad + 1 + 2 + \cdots + (n-1) \\ &= 1 + 3 + 5 + \cdots + (2n-1).\end{aligned}$$

—We'll use this pattern later (in Figure 2.38) to add up the first n squares and then again (in Figure 2.49) to add up the first n cubes.

Figures 2.21, 2.22 and 2.23 show some more relations between

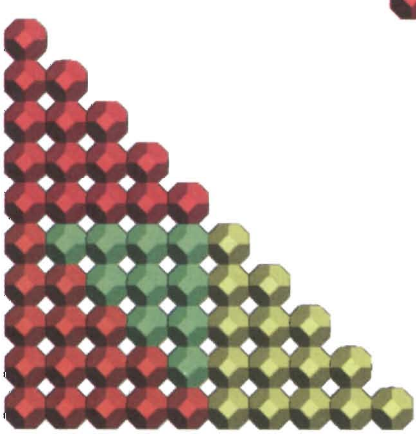


FIGURE 2.21 $3\Delta_n + \Delta_{n-1} = \Delta_{2n}$.

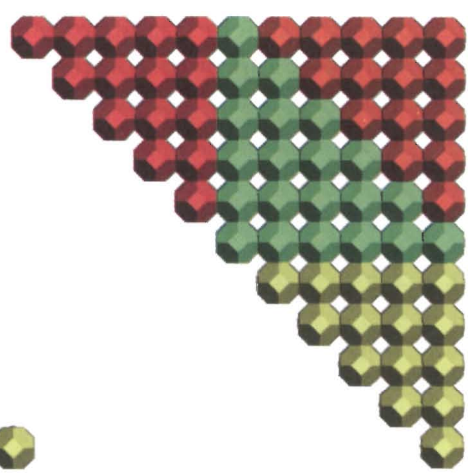


FIGURE 2.22 $3\Delta_n + \Delta_{n+1} = \Delta_{2n+1}$.

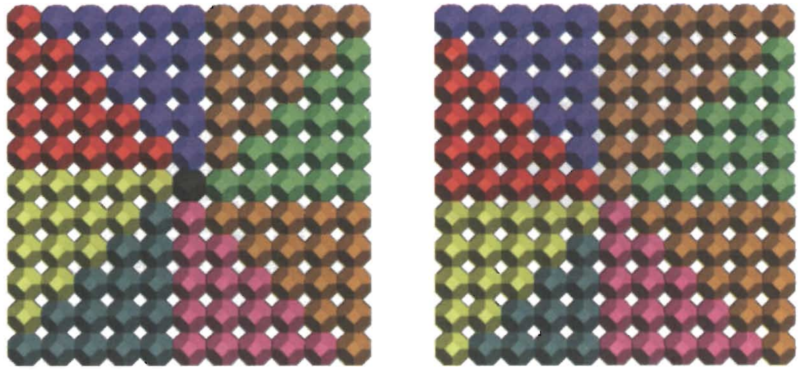


FIGURE 2.23 $(2n + 1)^2 = 8\Delta_n + 1 = \Delta_{n-1} + 6\Delta_n + \Delta_{n+1}$

triangular numbers. Figure 2.23 shows once again that an odd square is congruent to 1 modulo 8.

Which triangular numbers are also squares, for example,

$$0, 1, 36, 1225, \ldots ?$$

We'll have to wait until we've studied the Pell equation in Chapter 7 before we can answer that question.

THE POLYGONAL NUMBERS

We obtain the different kinds of polygonal numbers by adding the first n terms of appropriate arithmetic progressions starting with 1, thus:

1 + 1 + 1 + 1 + 1 + ···	gives	counting numbers	1, 2, 3, 4, 5 ···
1 + 2 + 3 + 4 + 5 + ···	gives	triangular numbers	1, 3, 6, 10, 15 ···
1 + 3 + 5 + 7 + 11 + ···	gives	square numbers	1, 4, 9, 16, 25 ···
1 + 4 + 7 + 10 + 13 + ···	gives	pentagonal numbers	1, 5, 12, 22, 35 ···
1 + 5 + 9 + 13 + 17 + ···	gives	hexagonal numbers	1, 6, 15, 28, 45 ···
1 + 6 + 11 + 16 + 21 + ···	gives	heptagonal numbers	1, 7, 18, 34, 55 ···
1 + 7 + 13 + 19 + 25 + ···	gives	octagonal numbers	1, 8, 21, 40, 65 ···

We've already met the first three kinds.

Notice that the number of sides in the polygon is *two more* than the common difference: we shall soon see, in Figure 2.24, and also in Chapter 4 on Catalan numbers, that it is two more than the number of triangles that make up the polygon. The third member of each sequence is always divisible by 3, and the fifth member by 5: is the seventh always divisible by 7?

These **polygonal numbers** get their names from arrangements of dots (Figure 2.24), which have been studied at least since the time of Pythagoras (ca. 540 B.C.).

Each sequence in Figure 2.24 can be formed from the row above by adjoining a triangle of Δ_{n-1} blobs of a new color to the left of each polygon. We already know that

$$n + \Delta_{n-1} = \Delta_n, \text{ the } n\text{th } \textbf{triangular number}, \text{ and that}$$
$$\Delta_n + \Delta_{n-1} = n + 2\Delta_{n-1} = n^2, \text{ the } n\text{th } \textbf{square}, \text{ and this continues:}$$
$$n^2 + \Delta_{n-1} = n + 3\Delta_{n-1} = \frac{1}{2}n(3n-1), \text{ the } n\text{th } \textbf{pentagonal number},$$
$$n + 4\Delta_{n-1} = n(2n-1), \text{ the } n\text{th } \textbf{hexagonal number},$$

and so on.

The p-sided polygonal number with n blobs in each side is

$$n + (p - 2)\Delta_{n-1} = \frac{1}{2}pn(n - 1) - n(n - 2).$$

For example, the nth hexagonal number is

$$n + 4\Delta_{n-1} = \Delta_n + 3\Delta_{n-1},$$

as you can see from Figure 2.25, which also shows (compare Figure 2.22) that

Every hexagonal number is a triangular number.

In fact, just the odd-sided triangular numbers give hexagonal numbers.

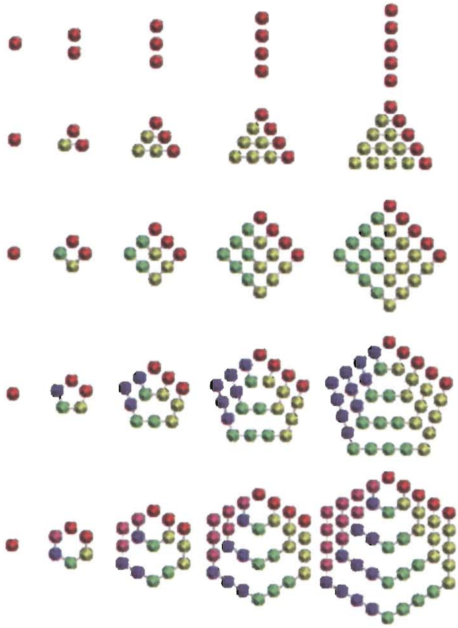


FIGURE 2.24 *Building polygonal numbers in two different ways. Adding gnomons (move from left to right in each row) increases the number of blobs in a side. Adding triangles (read down the page in each column) increases the number of sides: yellow makes triangles; green makes squares; blue makes pentagons; and violet makes hexagons.*

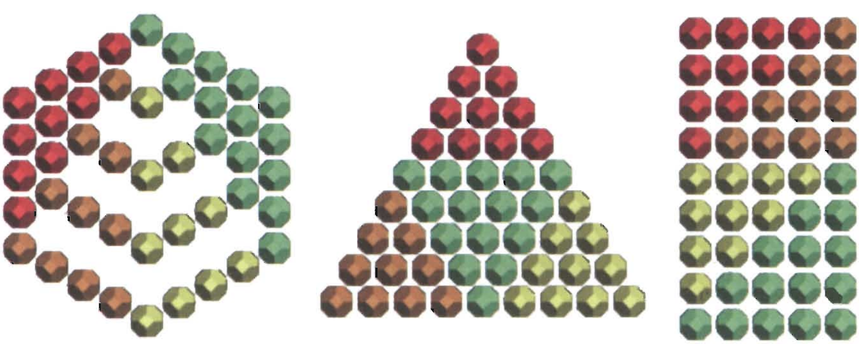


FIGURE 2.25 *The nth hexagonal number* $= 3\Delta_{n-1} + \Delta_n = \Delta_{2n-1} = n(2n - 1)$.

It's easy to prove algebraically that

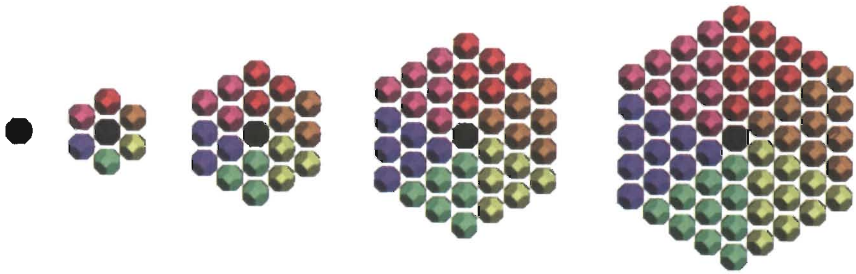
Every pentagonal number is one-third of a triangular number.

$$3 \times \frac{1}{2} n(3n - 1) = \frac{1}{2}(3n - 1)(3n).$$

Geometrically, flatten down the 'roof' of each pentagonal number in the fourth row of Figure 2.24 to make an equilateral trapezoid (bucket shape); you'll find that you can make triangular jigsaw puzzles from three copies of each.

Some people have used the name "hexagonal numbers" for those depicted in Figure 2.26, but we'll use Martin Gardner's name, **hex numbers**, to distinguish them.

FIGURE 2.26 *Hex numbers.*



1 7 19 37 61

From Figure 2.23 we can see that the nth hex number is

$$\text{hex}_n = 1 + 6\Delta_{n-1} = 1 - 3n + 3n^2.$$

Notice that $\text{hex}_{n+1} = 1 + 6\Delta_n = 1 + 3n + 3n^2$ and that the hex numbers are congruent to 1 modulo 6.

The **centred square numbers** (Figure 2.27), form a similar pattern (Figure 2.28(a)), showing that they are congruent to 1 modulo 4, which also follows from the fact that they are the sum of two consecutive squares (Figure 2.28(b)), one of which is even and the other odd.

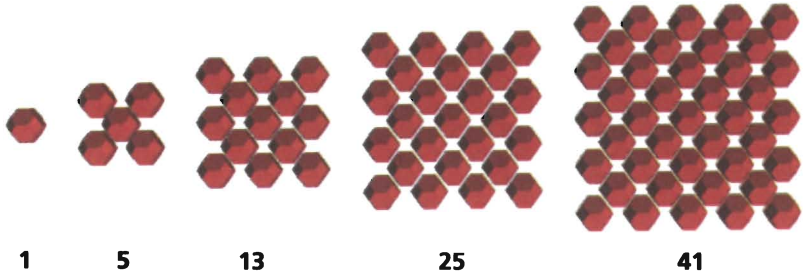


FIGURE **2.27** *The centred square numbers.*

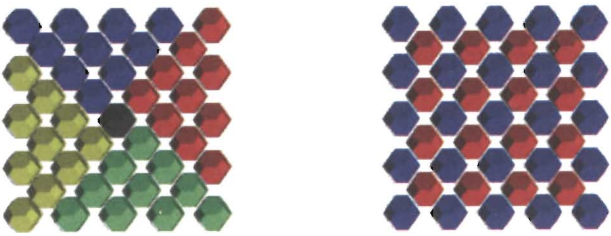
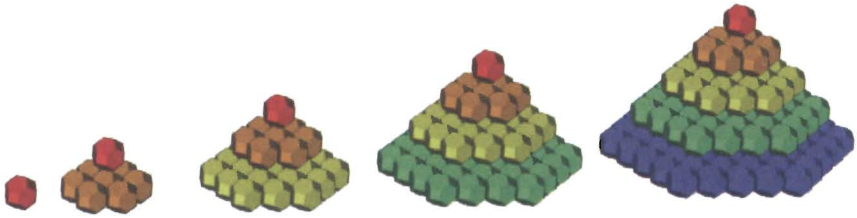


FIGURE **2.28** *The sum of consecutive squares is* $\equiv 1 \bmod 4$.

THE THIRD DIMENSION

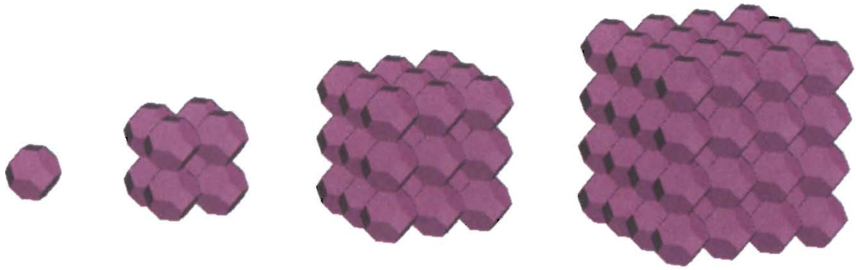
Suppose we stack the hex numbers up as **hex pyramids** (Figure 2.29). Surprise! We get the **cubes**, n^3 (Figure 2.30), perhaps the simplest of the three-dimensional figurate numbers, which we would normally make by stacking n n-by-n squares.

Why are the hex pyramids equal to the cubes? We can see this in Figure 2.31, where $n = 4$. The hexagons that we used to make the hex numbers are **projections**, or shadows, of cubes. We obtain the $(n + 1)$th cube by starting with a single red blob (Figure 2.31(a)) and building out 3 yellow rods of n blobs each (Figure 2.31(b)). Then the 3 spaces between pairs of rods each accommodate a green wall of $n \times n = n^2$ blobs (Figure 2.31(c)), and we have a nest or shell that



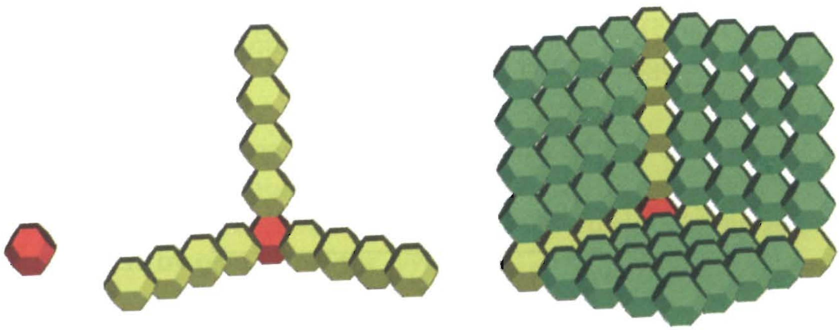
1 8 27 64 125

FIGURE 2.29 *Hex pyramids are cubes!*



$1^3 = 1$ $2^3 = 8$ $3^3 = 27$ $4^3 = 64$

FIGURE 2.30 *The cubes.*



1 $1 + 3n$ $1 + 3n + 3n^2$

FIGURE 2.31 *How a hex number builds a nest for the next cube.*

will neatly encase (3 adjacent faces of) an $n \times n \times n$ cube, making it up to $(1 + n)^3$. This is a very special case,

$$1 + 3n + 3n^2 + n^3 = (1 + n)^3,$$

of the **binomial theorem**, which we'll meet in Chapter 3.

We can stack other two-dimensional figurate numbers to make three-dimensional ones. For example, the triangular numbers can be stacked to form triangular pyramidal, or tetrahedral numbers.

TETRAHEDRAL NUMBERS

Figure 2.32 shows the first four **tetrahedral numbers**. What is the nth tetrahedral number? If you're good at three-dimensional jigsaw puzzles, then there are some ingenious ways of packing 6 copies of the nth tetrahedral number into an $n \times (n+1) \times (n+2)$ box, showing that the answer is

$$\text{Tet}_n = \frac{1}{6}n(n+1)(n+2).$$

But here's a way to see this without venturing into three dimensions. Add up all the numbers in the three copies of the triangular pattern in Figure 2.33. We get a triangular pattern of fifteen 7s (note that 7 is 2 more than 5) and similarly three times the nth tetrahedral number is the nth triangle number of $(n+2)$s, so that

> The nth tetrahedral number is
>
> $$\text{Tet}_n = \frac{1}{6}n(n+1)(n+2)$$

FIGURE 2.32 *The tetrahedral numbers.*

$$\begin{array}{ccccccc} 1 & & 1 & & 5 & & 7 \\ 1\ 2 & & 2\ 1 & & 4\ 4 & & 7\ 7 \\ 1\ 2\ 3 & + & 3\ 2\ 1 & + & 3\ 3\ 3 & = & 7\ 7\ 7 \\ 1\ 2\ 3\ 4 & & 4\ 3\ 2\ 1 & & 2\ 2\ 2\ 2 & & 7\ 7\ 7\ 7 \\ 1\ 2\ 3\ 4\ 5 & & 5\ 4\ 3\ 2\ 1 & & 1\ 1\ 1\ 1\ 1 & & 7\ 7\ 7\ 7\ 7 \end{array}$$

$$3(1 + 3 + 6 + 10 + 15) = (1 + 2 + 3 + 4 + 5) \times 7$$

$$3(\Delta_1 + \Delta_2 + \cdots + \Delta_n) = \Delta_n \times (n + 2) = \frac{1}{2} n(n + 1)(n + 2)$$

$$\Delta_1 + \Delta_2 + \cdots + \Delta_n = \frac{1}{6} n(n + 1)(n + 2).$$

FIGURE 2.33 *Adding up triangular numbers made easy.*

Tetrahedral numbers are whole numbers, so

The product of three consecutive integers is always a multiple of 6.

The third triangle of numbers in Figure 2.33 can be thought of as the fifth tetrahedral number standing on one of its edges. We can add up the numbers by reading off the layers:

$$(1 \times 5) + (2 \times 4) + (3 \times 3) + (4 \times 2) + (5 \times 1) = 35,$$

and generally

$$(1 \times n) + (2 \times (n - 1)) + (3 \times (n - 2)) + \cdots + ((n - 1) \times 2) + (n \times 1) = \text{Tet}_n.$$

Another way to look at this is to add up the SW–NE diagonals of the multiplication table (Figure 2.34; compare to Figure 2.8). Since the multiplication table is symmetrical, the tetrahedral numbers are **even**, except that the squares down the main diagonal are alternately

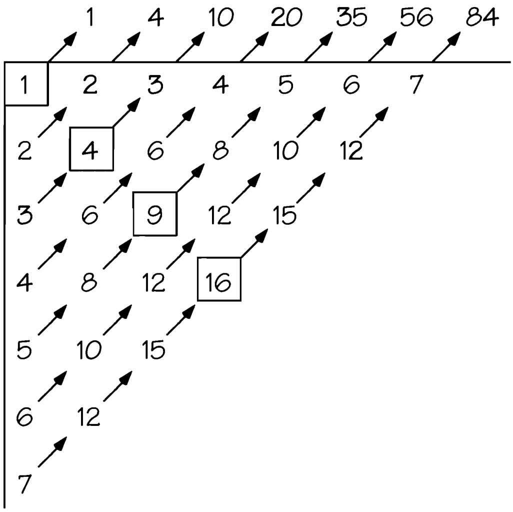


FIGURE 2.34 *The multiplication table gives the tetrahedral numbers.*

odd and even, making every fourth tetrahedral number **odd**, counting from the first:

1, 4, 10, 20, **35**, 56, 84, 120, **165**, 220, 286, 364, **455**, 560, ...

TRUNCATED TETRAHEDRAL NUMBERS

If we take the $(3n - 2)$th tetrahedral number and chop off the $(n - 1)$th tetrahedral number from each corner, we are left with the nth **truncated tetrahedral number** (Figure 2.35):

$$\text{Ttet}_n = \text{Tet}_{3n-3} - 4\text{Tet}_{n-1} = \frac{1}{6}n(23n^2 - 27n + 10).$$

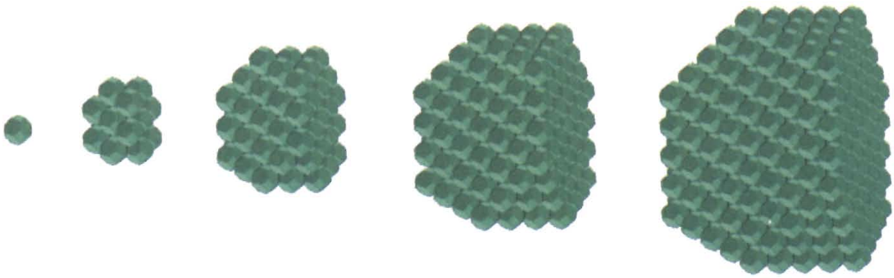


FIGURE 2.35 *The truncated tetrahedral numbers.*

SQUARE PYRAMID NUMBERS

Now let's make some square pyramids. Old war memorials sometimes have a stack of $1 + 4 + 9 + \cdots + n^2$ cannon balls as a square pyramid. This is the nth pyramidal number, Pyr_n.

You can make a different shape of "square pyramid" out of children's building blocks (stack them up with the corners of the squares one above the other; see Figure 2.36). Now try to pack six of these into a rectangular box of dimensions n by $n + 1$ by $2n + 1$. If you succeed, then you will have shown that:

> The nth pyramidal number, Pyr_n,
> $1^2 + 2^2 + 3^2 + \cdots + n^2$,
> is equal to $\frac{1}{6}n(n + 1)(2n + 1)$.

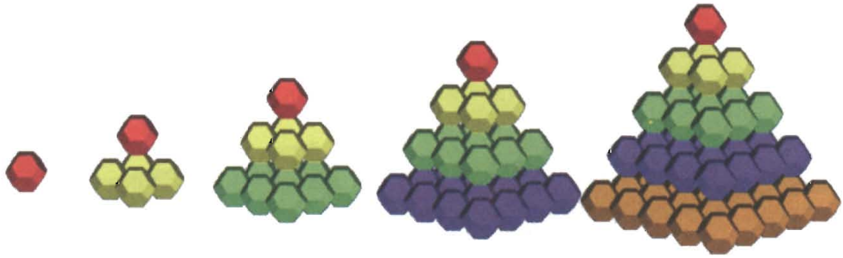


FIGURE 2.36 *The square pyramidal numbers, Pyr_n.*

But here are three easier ways to see this.

1. We saw in Figure 2.19 that the sum of two consecutive triangular numbers is a square,

$$\Delta_{n-1} + \Delta_n = n^2,$$

so the sum of two consecutive tetrahedral numbers is a square pyramidal number:

$$\mathrm{Tet}_{n-1} + \mathrm{Tet}_n = \mathrm{Pyr}_n,$$

$$\frac{1}{6}(n-1)n(n+1) + \frac{1}{6}n(n+1)(n+2) = \frac{1}{6}n(n+1)(2n+1).$$

2. Pack the layers of two (red and light blue) square pyramids into a shallow rectangular box, width $2n + 1$ and length Δ_n, the nth triangular number as in Figure 2.37(a). Partition the space that's left into strips, which fit together as in Figure 2.12 to make the square layers of a third square pyramid (Figure 2.37(b)).
   So one square pyramidal number, Pyr_n, is

$$\frac{1}{3}\Delta_n \times (2n+1) = \frac{1}{6}n(n+1)(2n+1),$$

   as before.

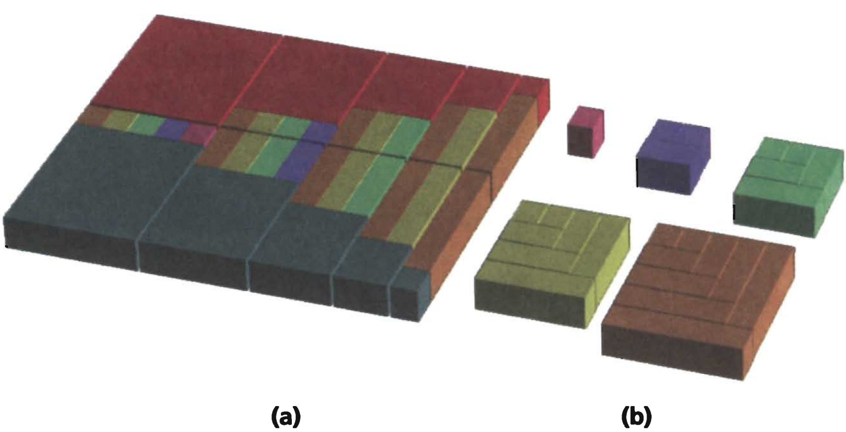


FIGURE 2.37 *Three square pyramids fit into a rectangular box.*

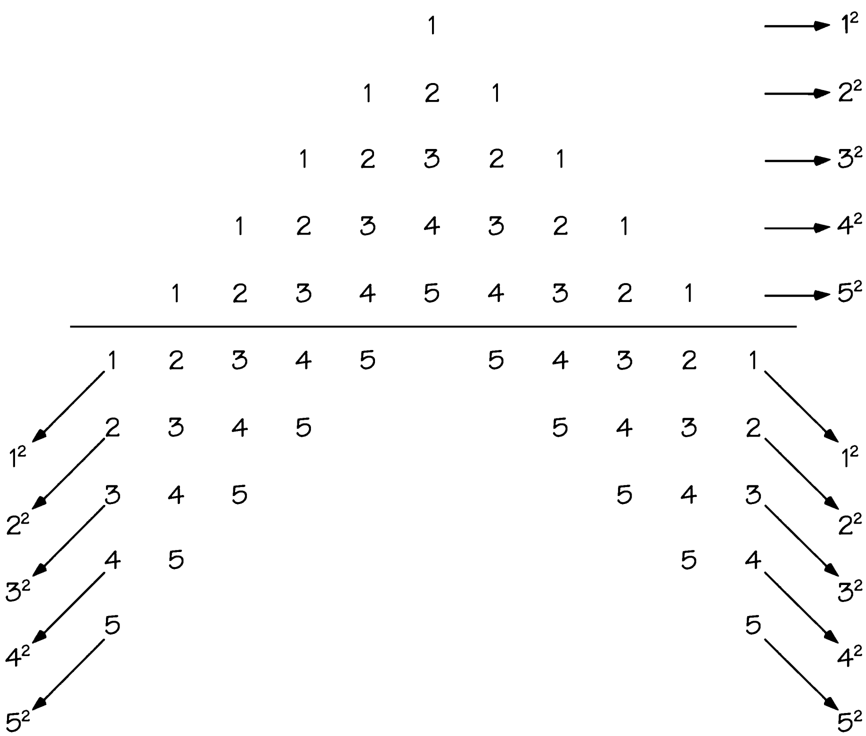


FIGURE 2.38 *Eleven copies of the fifth triangular number give three copies of the first five squares.*

3. Finally, in Figure 2.38 there are $(2 \times 5) + 1 = 11$ columns, each containing the triangular number

$$1 + 2 + 3 + 4 + 5 = \tfrac{1}{2} \times 5 \times (5 + 1).$$

The five rows above the line add up to $1^2, 2^2, 3^2, 4^2, 5^2$ by the Upstairs-Downstairs rule (Figure 2.20), while each of the two triangles below the line, when summed diagonally, also contains the first five squares:

$$3 \times (1^2 + 2^2 + 3^2 + 4^2 + 5^2) = (1 + 2 + 3 + 4 + 5) \times 11,$$

$$3 \text{ Pyr}_n = \Delta_n \times (2n + 1).$$

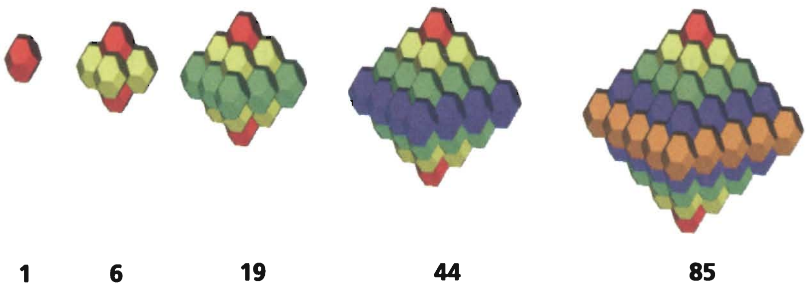


FIGURE 2.39 *The octahedral numbers.*

OCTAHEDRAL NUMBERS

The easiest way to view the **octahedral numbers** (Figure 2.39) is as double square pyramids, the sum of two consecutive square pyramids:

$$\mathrm{Oct}_n = \mathrm{Pyr}_{n-1} + \mathrm{Pyr}_n = \frac{1}{3}n(2n^2 + 1).$$

They are the sums of the numbers in the following arrays, in which the "1s" are the "poles" of the octahedron:

$$\begin{matrix} 1 \end{matrix} \quad \begin{matrix} 2 & 1 \\ 1 & 2 \end{matrix} \quad \begin{matrix} 3 & 2 & 1 \\ 2 & 3 & 2 \\ 1 & 2 & 3 \end{matrix} \quad \begin{matrix} 4 & 3 & 2 & 1 \\ 3 & 4 & 3 & 2 \\ 2 & 3 & 4 & 3 \\ 1 & 2 & 3 & 4 \end{matrix} \quad \begin{matrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 5 & 4 & 3 & 2 \\ 3 & 4 & 5 & 4 & 3 \\ 2 & 3 & 4 & 5 & 4 \\ 1 & 2 & 3 & 4 & 5 \end{matrix} \quad \begin{matrix} 6 & 5 & 4 & 3 & 2 & 1 \\ 5 & 6 & 5 & 4 & 3 & 2 \\ 4 & 5 & 6 & 5 & 4 & 3 \\ 3 & 4 & 5 & 6 & 5 & 4 \\ 2 & 3 & 4 & 5 & 6 & 5 \\ 1 & 2 & 3 & 4 & 5 & 6 \end{matrix}$$

0 1 6 19 44 85 146

The differences

1 5 13 25 41 61

between two consecutive octahedral numbers are the centred square numbers we saw in Figure 2.27.

STELLA OCTANGULA NUMBERS

If we take an octahedral number of edge n and adjoin a tetrahedral number of edge $n - 1$ to four alternate faces, we get a tetrahedral number of edge $(n - 1) + 1 + (n - 1) = 2n - 1$, as shown in Figure 2.40.

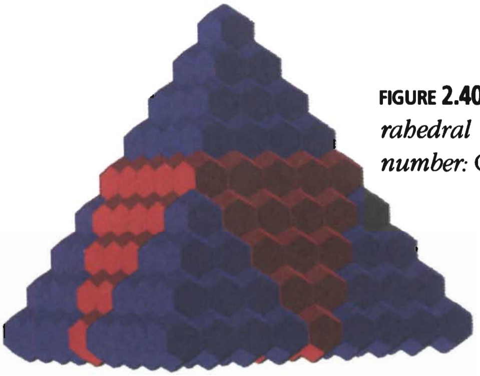
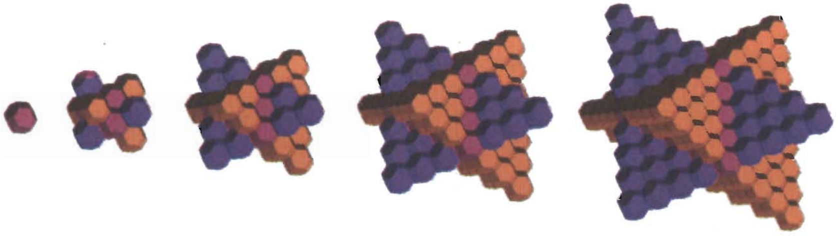


FIGURE 2.40 *An octahedral number and four tetrahedral numbers make a bigger tetrahedral number:* $\mathrm{Oct}_n + 4\mathrm{Tet}_{n-1} = \mathrm{Tet}_{2n-1}$.

If we adjoin four more such tetrahedra to the other four faces, we get a **stella octangula number** (Figure 2.41) named for Kepler's *stella octangula*:

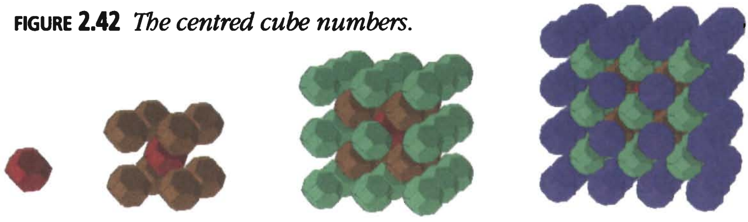
$$\mathrm{Stel}_n = \mathrm{Oct}_n + 8\mathrm{Tet}_{n-1} = n(2n^2 - 1).$$



1 14 51 124 245

FIGURE 2.41 *The stella octangula numbers.*

FIGURE 2.42 *The centred cube numbers.*



1 **9** **35** **91**

Centred Cube Numbers

The three-dimensional analogs of the centred square numbers are the (body-)**centred cube numbers** (Figure 2.42):

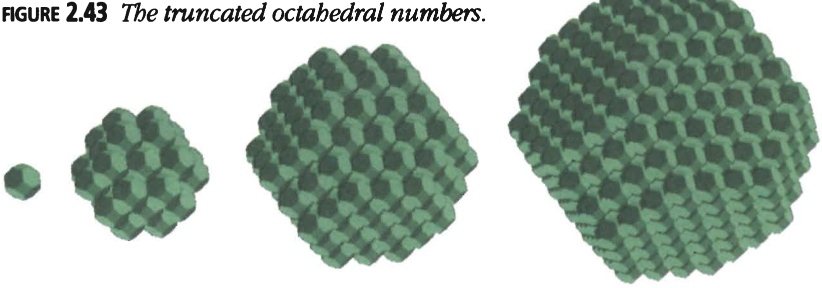
$$\text{Ccub}_n = n^3 + (n-1)^3 = (2n-1)(n^2-n+1).$$

Truncated Octahedral Numbers

Start with the $(3n-2)$th octahedral number, Oct_{3n-2}, and cut off the $(n-1)$th square pyramid, Pyr_{n-1}, from each of its six vertices. We are left with the **truncated octahedral numbers** (Figure 2.43):

$$\begin{aligned}\text{Toct}_n = \text{Oct}_{3n-2} - 6\,\text{Pyr}_{n-1} &= \frac{1}{3}(3n-2)(2(3n-2)^2+1)\\ &\quad - \frac{6}{6}(n-1)n(2n-1)\\ &= 16n^3 - 33n^2 + 24n - 6.\end{aligned}$$

FIGURE 2.43 *The truncated octahedral numbers.*



1 **38** **201** **586**

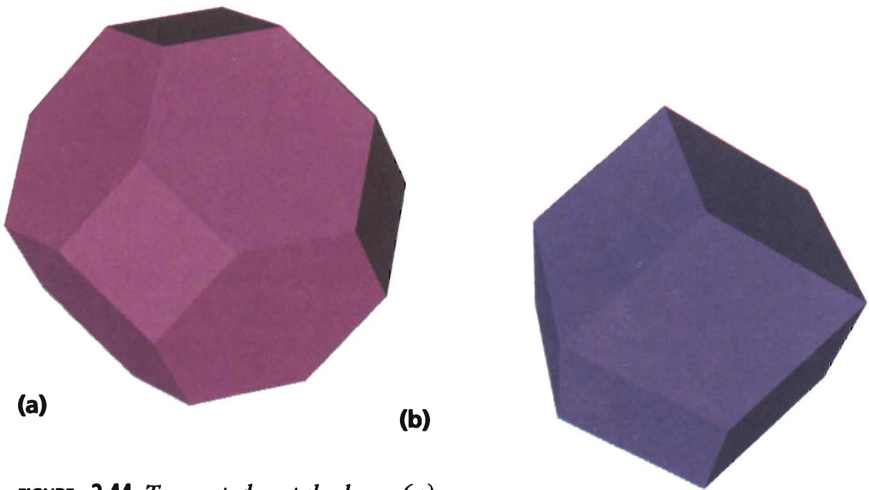
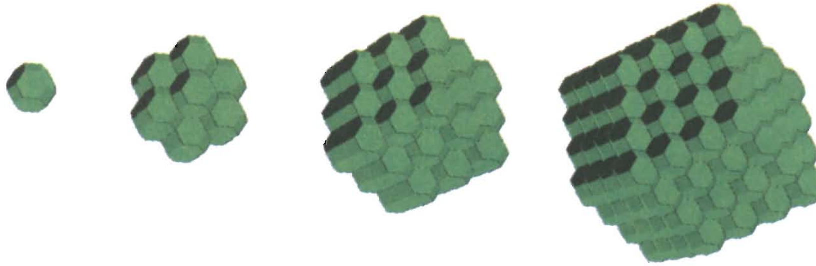


FIGURE 2.44 *Truncated octahedron (a) and rhombic dodecahedron (b).*

RHOMBIC DODECAHEDRAL NUMBERS

What are the three-dimensional analogs of the hex numbers? The nth hex number, hex_n, counts the number of cells of a honeycomb packing that are less than n steps away from a central one (the black one in Figure 2.26). Now, among the Platonic (regular) and Archimedean (semiregular) solids, the only ones that pack three-space exactly are the cube (obviously) and the truncated octahedron, Figure 2.44(a), the blob that we have used as our unit in most of the figures in this chapter. In the packing by truncated octahedra, the nexus of all cells less than n steps away from a given cell has the shape of a rhombic



1 15 65 175

FIGURE 2.45 *The rhombic dodecahedral numbers.*

dodecahedron, Figure 2.44(b). The three-dimensional nexus numbers are the **rhombic dodecahedral numbers** (Figure 2.45). One way to visualize a rhombic dodecahedral number is by appending a square pyramid to each of the six faces of a centred cube:

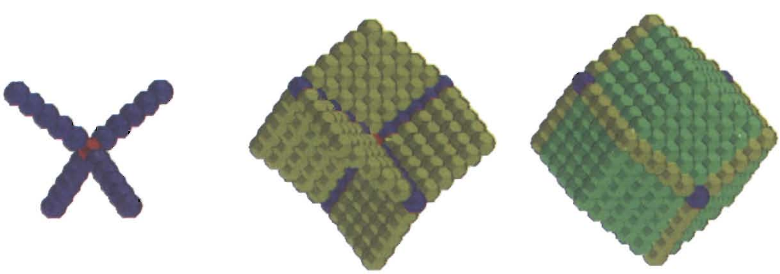
$$\text{Rho}_n = \text{Ccub}_n + 6\text{Pyr}_{n-1} = (2n - 1)(2n^2 - 2n + 1).$$

Figure 2.26 showed how $\text{hex}_{n+1} = 1 + 3n + 3n^2$. Figure 2.46 shows how $\text{Rho}_{n+1} = 1 + 4n + 6n^2 + 4n^3$. Start with one red blob and build out four rods of n blue cells from alternate hexagonal faces (Figure 2.46(a)). Then add 6 yellow walls, each of n^2 cells, one between each pair of rods (Figure 2.46(b)). Finally, insert 4 blocks of n^3 green cells, one in each solid angle formed by three walls (Figure 2.46(c)).

Similar nexus numbers are defined in all dimensions:

Dimension	$(n + 1)$th nexus number	
0	1	the unit
1	$1 + 2n$	odd numbers
2	$1 + 3n + 3n^2$	hex numbers
3	$1 + 4n + 6n^2 + 4n^3$	rhombic dodecahedral numbers
4	$1 + 5n + 10n^2 + 10n^3 + 5n^4$	the next nexus numbers

and so on. . . .



1+4n (a) 1+4+6n² (b) 1+4n+6n²+4n³ (c)

FIGURE 2.46 *Building a nexus number.*

BENDING THE NEXT NEXUS NUMBER

In Figure 2.12 we bent each one-dimensional nexus number (odd number) into a gnomon and stacked these gnomons to form squares. In Figure 2.31 we bent the hex numbers into "nests" and then stacked these to form cubes. Similarly, in four dimensions we can bend the rhombic dodecahedral numbers into nests that stack to form four-dimensional cubes, or **tesseracts**. Most of us find this a little bit difficult to visualize, but

$$\begin{aligned} \text{odd}_1 + \text{odd}_2 + \cdots + \text{odd}_n &= n^2, \\ \text{hex}_1 + \text{hex}_2 + \cdots + \text{hex}_n &= n^3, \\ \text{Rho}_1 + \text{Rho}_2 + \cdots + \text{Rho}_n &= n^4, \end{aligned}$$

and $\text{Rho}_{n+1} = 1 + 4n + 6n^2 + 4n^3$ is just what's needed to increase n^4 to $(n + 1)^4$. If you've been skipping ahead, you'll have seen the connection with the binomial theorem in the next chapter.

THE FOURTH DIMENSION

There are other figurate numbers in higher dimensions. Although it's hard to visualize jigsaw puzzles in four dimensions, it can be done! Suppose you want to stack up the first few tetrahedral numbers to make **pentatope numbers** (Figure 2.47) (the pentatope is the simplest regular figure in four dimensions).

$$\begin{aligned} 1 &= 1 = \text{Ptop}_1 \\ 1 + 4 &= 5 = \text{Ptop}_2 \\ 1 + 4 + 10 &= 15 = \text{Ptop}_3 \\ 1 + 4 + 10 + 20 &= 35 = \text{Ptop}_4 \\ 1 + 4 + 10 + 20 + 35 &= 70 = \text{Ptop}_5 \\ 1 + 4 + 10 + 20 + 35 + 56 &= 126 = \text{Ptop}_6 \end{aligned}$$

FIGURE 2.47 *Pentatope numbers.*

```
                                    5
                        4         4 4
              3       3 3       3 3 3
        2   2 2     2 2 2     2 2 2 2
      1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

                                    1
                        1         1 2
              1       1 2       1 2 3
        1   1 2     1 2 3     1 2 3 4
      1 1 2 1 2 3 1 2 3 4 1 2 3 4 5

                                    1
                        1         2 1
              1       2 1       3 2 1
        1   2 1     3 2 1     4 3 2 1
      1 2 1 3 2 1 4 3 2 1 5 4 3 2 1

                                    1
                        2         1 1
              3       2 2       1 1 1
        4   3 3     2 2 2     1 1 1 1
      5 4 4 3 3 3 2 2 2 2 1 1 1 1 1
      _____________________________

                                    8
                        8         8 8
              8       8 8       8 8 8
        8   8 8     8 8 8     8 8 8 8
      8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
```

FIGURE 2.48 *Four copies of the first five tetrahedral numbers add up to (5 + 3) times the sum of the first five triangular numbers.*

Here's a way of doing it without taking leave of our three-dimensional world. Indeed, we can even stay on our two-dimensional paper. The first row of Figure 2.48 shows the first five tetrahedral numbers (compare with Figure 2.33). These are copied out three more times, but rearranged. If you add four numbers, one from the same position in each copy, the sum is always 8. The number of positions is the sum of the first five triangular numbers, that is Tet_5, the fifth tetrahedral number.

In general, four copies of the first n tetrahedral numbers add up to $n + 3$ times the nth tetrahedral number, $\frac{1}{6}n(n + 1)(n + 2)$. So

The nth pentatope number is:

$$\mathrm{Ptop}_n = \frac{1}{4}\mathrm{Tet}_n \times (n + 3) = \frac{1}{24}n(n + 1)(n + 2)(n + 3)$$

Incidentally, we have proved that

The product of four consecutive integers is divisible by 24

SUMS OF CUBES

We could also use a four-dimensional jigsaw puzzle to pile up cubes to make a cubic pyramid, but the multiplication table of Figure 2.8 rescues us from this! Figure 2.49 is the multiplication table, equipped with gnomons as in Figure 2.12.

$1^3 \to$	1	2	3	4	5
$2^3 \to$	2	4	6	8	10
$3^3 \to$	3	6	9	12	15
$4^3 \to$	4	8	12	16	20
$5^3 \to$	5	10	15	20	25

FIGURE 2.49 *Each gnomon in the multiplication table sums to a cube and helps us to add up the cubes.*

The gnomons contain

$$\begin{aligned}
1(1) &= 1 \times 1^2 = 1^3 \\
2(1+2+1) &= 2 \times 2^2 = 2^3 \\
3(1+2+3+2+1) &= 3 \times 3^2 = 3^3 \\
4(1+2+3+4+3+2+1) &= 4 \times 4^2 = 4^3 \\
5(1+2+3+4+5+4+3+2+1) &= 5 \times 5^2 = 5^3
\end{aligned}$$

using the Upstairs-Downstairs rule (Figure 2.20). But the total of the numbers in the multiplication table is exactly the product

$$(1+2+3+4+5)(1+2+3+4+5) = \Delta_5^2$$

and so, as we saw in Figure 2.18(b):

> The sum of the first n cubes
>
> $$1^3 + 2^3 + 3^3 + \cdots + n^3$$
>
> is equal to the square of the nth triangular number,
>
> $$\Delta_n^2 = \frac{1}{4} n^2(n+1)^2$$

Figure 2.50 shows this as an easy three-dimensional jigsaw puzzle. The square of side $1+2+3+4+5 = 15$ of part (a) is chopped into rectangles that are reassembled to form five cubes (b).

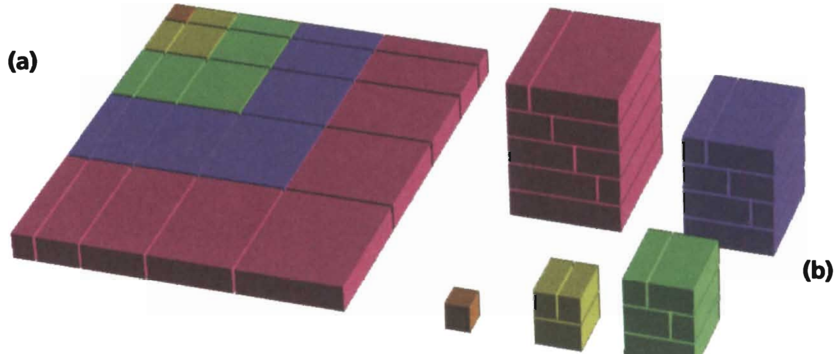


FIGURE 2.50 *A squared triangular number is the sum of cubes.*

MORE AND MORE DIMENSIONS

We can go on piling up triangular pyramids in more and more dimensions. The method that we used in Figures 2.29 and 2.45 extends to show that

$$1 + 1 + 1 + \cdots + 1 = n$$
$$1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$$
$$1 + 3 + 6 + \cdots + \frac{1}{2}n(n+1) = \frac{1}{6}n(n+1)(n+2)$$
$$1 + 4 + 10 + \cdots + \frac{1}{6}n(n+1)(n+2) = \frac{1}{24}n(n+1)(n+2)(n+3)$$
$$1 + 5 + 15 + \cdots + \frac{1}{24}n(n+1)(n+2)(n+3) \qquad \ldots \quad \ldots$$

giving the counting, triangular, tetrahedral, and pentatope numbers; and although we've run out of names, we'll never run out of dimensions—the first unnamed numbers

$$\frac{1}{120}n(n+1)(n+2)(n+3)(n+4)$$

show that:

The product of five consecutive integers is divisible by $120 = 1 \times 2 \times 3 \times 4 \times 5$

SOME VERY LARGE NUMBERS

The squares get large quite quickly; the cubes and higher powers expand even more quickly; and n^n grows more quickly than *all* the figurate numbers. However, mathematicians sometimes need even larger numbers and want some way of writing them.

Computers used to print $m \uparrow n$ so as to avoid superscripts (though today only the arrowhead seems to remain). This suggests the following handy **arrow notation**. Just as

$m \times n$ or	mn	abbreviates	$m + m + \cdots + m$
and	$m \uparrow n$	abbreviates	$mm \cdots m$
so we use	$m \uparrow\uparrow n$	to abbreviate	$m \uparrow m \uparrow \cdots \uparrow m$
then	$m \uparrow\uparrow\uparrow n$	to abbreviate	$m \uparrow\uparrow m \uparrow\uparrow \cdots \uparrow\uparrow m$
and then	$m \uparrow\uparrow\uparrow\uparrow n$	to abbreviate	$m \uparrow\uparrow\uparrow m \uparrow\uparrow\uparrow \cdots \uparrow\uparrow\uparrow m$

and so on, with n copies of m in each case and these expressions being evaluated from the right.

Although the notation $m \uparrow\uparrow \cdots \uparrow n$ was only introduced in 1976 by Donald Knuth, an essentially similar function was defined by W. Ackermann in 1928, and so we'll call

$$1\uparrow 1, \quad 2\uparrow\uparrow 2, \quad 3\uparrow\uparrow\uparrow 3, \quad 4\uparrow\uparrow\uparrow\uparrow 4, \ldots$$

the **Ackermann numbers**.

The first Ackermann number is 1, the second is $2 \uparrow\uparrow 2 = 2\uparrow 2 = 4$, and the third is

$$3^{3^{3^{\cdot^{\cdot^{\cdot}}}}}$$

where the number of threes is $3^{3^3} = 7625597484987$. It's virtually impossible to comprehend the immensity of the fourth Ackermann number, $4\uparrow\uparrow\uparrow\uparrow 4 = 4\uparrow\uparrow\uparrow 4\uparrow\uparrow\uparrow 4\uparrow\uparrow\uparrow 4$. This is

$$4\uparrow\uparrow 4\uparrow\uparrow 4\uparrow\uparrow 4 \cdots \uparrow\uparrow 4\uparrow\uparrow 4\uparrow\uparrow 4\uparrow\uparrow 4,$$

where the number of fours here is

$$4\uparrow\uparrow 4\uparrow\uparrow 4 \cdots \uparrow\uparrow 4\uparrow\uparrow 4\uparrow\uparrow 4,$$

where the number of fours *there* is

$$4\uparrow\uparrow 4\uparrow\uparrow 4\uparrow\uparrow 4$$

(the number of fours in *that* being four). This last number is

$$4^{4^{4^{\cdot^{\cdot^{\cdot^{4}}}}}}$$

in which the number of fours is

$$4^{4^{4^{\cdot^{\cdot^{\cdot^{4}}}}}},$$

where the number of fours is $4^{4^{4^4}} = 4^{4^{256}} = 4^{(\text{a 155-digit number})}$.

Our own "chained arrow" notation names some even larger numbers. In this, $a \uparrow\uparrow \cdots \uparrow\uparrow b$ (with c arrows) is called $a \to b \to c$.

$$a \to b \to \cdots \to x \to y \to 1$$

is another name for $a \to b \to \cdots \to x \to y$, and

$$a \cdots x \to y \to (z + 1)$$

is defined to be

$$\begin{array}{rl} a \cdots x & \text{if } y = 1, \\ a \cdots x \to (a \cdots x) \to z & \text{if } y = 2, \\ a \cdots x \to (a \cdots x \to (a \cdots x) \to z) \to z & \text{if } y = 3, \end{array}$$

and so on. The parentheses here may be rubbed out after the numbers inside them have been completely evaluated. The first three of our own rapidly increasing sequence of numbers

$$1,\ 2 \to 2,\ 3 \to 3 \to 3,\ 4 \to 4 \to 4 \to 4,\ \ldots$$

agree with the Ackermann numbers $1\uparrow 1$, $2\uparrow\uparrow 2$, $3\uparrow\uparrow\uparrow 3$, but $4 \to 4 \to 4 \to 4$ is very much larger than $4\uparrow\uparrow\uparrow\uparrow 4$, which is only $4 \to 4 \to 4$.

The number $10^{10^{10^{34}}}$ was once called **Skewes's number** and was said to be the largest individual number that occurred in a mathematical proof. However, the number that appears in modern improvements of Skewes's theorem has been deflated to 10^{1167} (see Chapter 5). Its role has in any case been taken over by **Graham's number**, although the theorem in which this number appears might well be deflated in the future.

Graham's number is
$4\uparrow\uparrow \ldots \uparrow\uparrow 4$, where the number of arrows is
$4\uparrow\uparrow \ldots \uparrow\uparrow 4$, where the number of arrows is
... et cetera ...
(where the number of lines like that is 64).
It lies between $3 \to 3 \to 64 \to 2$ and
$3 \to 3 \to 65 \to 2$.

Skewes's number	is less than	$4^{4^{4^{4}}} = 4\uparrow\uparrow 5$,
10^{1167}	is less than	$5^{5^{5}} = 5\uparrow\uparrow 3$, and
Graham's number	is less than	$3 \to 3 \to 3 \to 3$.

What's the largest individual number that occurs "naturally" in an undeflatable theorem? A moderately large and very special number is

8080 17424 79451 28758 86459 90496 17107 57005 75436 80000 00000,

which is the order of the so-called Monster simple group, but we are sure that this entry has already been outrun by worthier candidates.

REFERENCES

W. Ackermann. Zum Hilbertschen Aufbau der reellen Zahlen. *Math. Annalen,* 99(1928): 118–133.

Martin Gardner. *Second Book of Mathematical Puzzles and Diversions.* W. H. Freeman, New York, 1961; Chapter 4, Digital Roots.

Martin Gardner. *Time Travel and Other Mathematical Bewilderments.* W. H. Freeman, New York, 1988; Chapter 2, Hexes and Stars.

Martin Gardner. *Penrose Tiles to Trapdoor Ciphers.* W. H. Freeman, New York, 1989; Chapter 17, Ramsey Theory.

P. Goodwin. A polyhedral series or two. *Math. Gaz.,* 69(1985): 191–197.

Harvey Hindin. Stars, hexes, triangular numbers and Pythagorean triples. *J. Recreational Math.,* 16(1983-84): 191–193.

Edward Kasner and James Newman. *Mathematics and the Imagination.* G. Bell & Sons, London, 1949, p. 23.

Donald E. Knuth. Mathematics and computer science: Coping with finiteness. *Science,* (1976): 12–17.

3

What Comes Next?

We hope you like the kind of problem where someone gives you an intriguing number sequence and asks you what comes next. In this chapter we'll give you several ways to find out. Most of these involve building some kind of pattern from your numbers. Pascal's triangle is one very well-known pattern.

If you're like us, you'll also enjoy playing with number sequences for their own sake, so we'll also show you some number games that often use patterns to magically transform one sequence into another. One of the nicest of these is described next.

Moessner's Magic

Alfred Moessner discovered that some of our favorite sequences can be found in a surprising new way. Start with the counting numbers and circle every second number; then form the cumulative totals of the uncircled numbers, and you'll see the squares:

1 (2) 3 (4) 5 (6) 7 (8) 9 (10) 11 (12) 13 (14) 15 (16)
(1) (4) (9) (16) (25) (36) (49) (64)

If instead you circle every third number, total what's left, circling the last number in each block, and total the uncircled (hex) numbers, you'll see the cubes:

```
1  2 (3)  4  5 (6)  7  8 (9) 10 11 (12) 13 14 (15)
1 (3)     7 (12)    19 (27)   37 (48)    61 (75)
(1)      (8)       (27)      (64)      (125)
```

Circling every fourth number:

```
1  2  3 (4)  5  6  7 (8)  9 10 11 (12) 13 14 15 (16)
1  3 (6)     11 17 (24)   33 43 (54)   67 81 (96)
1 (4)        15 (32)      65 (108)     175 (256)
(1)          (16)         (81)         (256)
```

leads similarly to the fourth powers, and so on.

So circling the numbers

$$n + n \quad n + n + n \quad n + n + n + n \ldots$$

has led to the numbers

$$n \times n \quad n \times n \times n \quad n \times n \times n \times n \ldots .$$

If we circle each triangular number, $1 + 2 + 3 + \cdots + n$:

```
(1) 2 (3) 4  5 (6) 7  8  9 (10) 11  12  13  14 (15)
   (2)    6 (11)   18 26 (35)    46  58  71 (85)
         (6)       24 (50)       96 154 (225)
                  (24)           120 (274)
                                (120)
```

we get the **factorial numbers**, $1 \times 2 \times 3 \times \cdots \times n$, which we'll talk about soon.

What if we circle the squares?

```
(1) 2  3 (4)  5  6  7  8 (9) 10  11  12  13  14  15 (16)
    2 (5)    10 16 23 (31)   41  52  64  77  91 (106)
   (2)       12 28 (51)      92 144 208 285 (376)
             12 (40)         132 276 484 (769)
            (12)             144 420 (904)
                             144 (564)
                            (144)
```

If these numbers mystify you, notice that the squares are

$$1$$
$$1 + 2 + 1$$
$$1 + 2 + 3 + 2 + 1$$
$$1 + 2 + 3 + 4 + 3 + 2 + 1$$
$$\dots$$

and that the final circled numbers are

$$1$$
$$1 \times 2 \times 1$$
$$1 \times 2 \times 3 \times 2 \times 1$$
$$1 \times 2 \times 3 \times 4 \times 3 \times 2 \times 1$$

The general rule is that if you start by circling

$$1a, \quad 2a + 1b, \quad 3a + 2b + 1c, \quad 4a + 3b + 2c + 1d \dots,$$

then the final circled numbers are

$$1^a, \quad 2^a \times 1^b, \quad 3^a \times 2^b \times 1^c, \quad 4^a \times 3^b \times 2^c \times 1^d \dots.$$

FACTORIAL NUMBERS

How many "words" can we make from the letters A, E, T, each used just once?

AET, ATE, EAT, ETA, TAE, TEA

The first letter can be any one of the three, the second can be either one of the two remaining, and the third is then the one left over,

$$3 \times 2 \times 1 = 6 \text{ words.}$$

If you have n different letters, they can be arranged in

$$n \times (n - 1) \times (n - 2) \times \cdots \times 3 \times 2 \times 1 \text{ ways.}$$

This number is called **factorial n**, or **n factorial**. It often used to be written $\underline{|n}$, but today is usually written $n!$.

Of course, there's just one way to arrange no objects, so $0! = 1$. In general, $n!$ is the product of the numbers from 1 to n, the empty product being 1 (Figure 3.1).

$$\begin{array}{rcrcl} & = & 1 & = & 0! \\ 1 & = & 1 & = & 1! \\ 1 \times 2 & = & 2 & = & 2! \\ 1 \times 2 \times 3 & = & 6 & = & 3! \\ 1 \times 2 \times 3 \times 4 & = & 24 & = & 4! \\ 1 \times 2 \times 3 \times 4 \times 5 & = & 120 & = & 5! \\ 1 \times 2 \times 3 \times 4 \times 5 \times 6 & = & 720 & = & 6! \\ 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 & = & 5040 & = & 7! \end{array}$$

FIGURE 3.1 *The factorial numbers.*

We just saw how we can get the factorial numbers from Moessner's magic, and in fact we already met them in Chapter 2 when we piled up triangular pyramids in more and more dimensions.

ARRANGEMENT NUMBERS

Factorial n is the number of **arrangements**, or **orders**, or **permutations** of n things in a row. How many arrangements are there of r objects, chosen from n different things? The first can be any one of the n, the second can be any one of the remaining $n-1$, the third any one of remaining $n-2$, and so on, the rth being any one of $n - r + 1$. The total number of different arrangements is

$$n \times (n-1) \times (n-2) \times \cdots \times (n-r+1),$$

the product of all the numbers from 1 to n, except for those from 1 to $n - r$, so we can express this concisely using the factorial numbers:

> The number of **arrangements** of r things out of n is
> $$\frac{n!}{(n-r)!}$$

CHOICE NUMBERS

If we're only concerned with the number of **choices**, or **combinations**, of the r things chosen from the n, then we don't distinguish between the factorial r different ways in which we could have arranged them in a row. So to get the **choice numbers**, $\binom{n}{r}$, we divide the arrangement numbers by r!

The number of *choices* of r things from n is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

In this formula, you can swap r for $n - r$ without altering the value. The number of ways of choosing 5 things out of 8 is the same as the number of ways of choosing the 3 you want to leave out:

$$\binom{8}{5} = \binom{8}{3}$$

and generally,

$$\binom{n}{r} = \binom{n}{n-r}$$

This is the left-right symmetry of Pascal's triangle, see Figures 3.2 and 3.3

Suppose a class of 28 students wants to choose a soccer team of 11 players. In how many ways can they do it? We now know that this is

$$\binom{28}{11} = \frac{28 \times 27 \times 26 \times 25 \times 24 \times 23 \times 22 \times 21 \times 20 \times 19 \times 18}{1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 11}$$

$$= \frac{28!}{11!17!} = 2^2 \times 3^3 \times 5 \times 7 \times 13 \times 19 \times 23 = 21474180.$$

Now suppose you're in the class and want to know if you're on the team. In how many ways could you be included? If you're on, the other 10 must be chosen from the other 27:

$$\binom{27}{10} = 8436285 \text{ ways.}$$

In how many ways are you *not* included? All 11 have to be chosen from the other 27:

$$\binom{27}{11} = 13037895 \text{ ways.}$$

So $\binom{28}{11}$ is the sum of these two numbers, and generally, since you are either on the team of r or not,

$$\boxed{\binom{n}{r} = \binom{n-1}{r} + \binom{n-1}{r-1}}$$

PASCAL'S TRIANGLE

This is a very simple way of generating the choice numbers. Start from $\binom{0}{0} = 1$ on row 0, and $\binom{1}{0} = 1$ and $\binom{1}{1} = 1$ on row 1, and calculate successive rows by putting $\binom{n}{0} = 1$ and $\binom{n}{n} = 1$ at each end and forming each other number as the sum of the two in the row immediately above (Figure 3.2).

The first few choice numbers are shown in Figure 3.2. The array in Figure 3.3 is usually known as **Pascal's triangle**, because it was intensively studied by Blaise Pascal (1623–1662), the famous French philosopher and mathematician. It had already been described much earlier by Chinese mathematicians and by Omar Khayyám, who died in 1123.

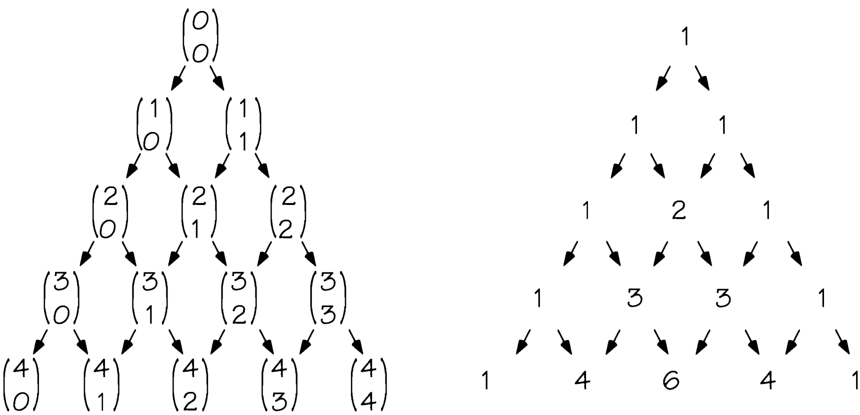


FIGURE 3.2 *Generating the choice numbers.*

Of course, we've seen some of these numbers before, in Chapter 2, when we piled up triangular pyramids in more and more dimensions. The numbers at the beginning of each row are just ones,

1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ...

The second numbers in each row are the **counting numbers**,

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ...

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 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1
1 9 36 84 126 126 84 36 9 1
1 10 45 120 210 252 210 120 45 10 1
1 11 55 165 330 462 462 330 165 55 11 1
1 12 66 220 495 792 924 792 495 220 66 12 1
1 13 78 286 715 1287 1716 1716 1287 715 286 78 13 1

FIGURE 3.3 *Pascal's numbers: the choice numbers, or binomial coefficients.*

The third numbers are the **triangular numbers**,

1, 3, 6, 10, 15, 21, 28, 36, 45, 55, 66, ...

The fourth numbers are the **tetrahedral numbers**,

1, 4, 10, 20, 35, 56, 84, 120, 165, 220, 286, ...

The fifth ones are the **pentatope numbers**,

1, 5, 15, 35, 70, 126, 210, 330, 495, 715, 1001, . . . ,

and so on. The numbers in each diagonal are the cumulative sums of those in the previous diagonal.

CHOICE NUMBERS WITH REPETITIONS

In how many ways can you choose five things from *n*, if repetitions are allowed? In other words, how many essentially different kinds of "poker hands" are there, if we ignore flushes and straights and are playing with a double deck, so that you can have five of a kind?

"Poker hand"	13 in a suit	*n* cards in a suit
all different	$\binom{13}{5} = 1287$	$\binom{n}{5}$
one pair	$13 \times \binom{12}{3} = 2860$	$n \times \binom{n-1}{3}$
two pairs	$\binom{13}{2} \times 11 = 858$	$\binom{n}{2} \times (n-2)$
three of a kind	$13 \times \binom{12}{2} = 858$	$n \times \binom{n-1}{2}$
full house (3 & 2)	$13 \times 12 = 156$	$n \times (n-1)$
four of a kind	$13 \times 12 = 156$	$n \times (n-1)$
five of a kind	$13 = 13$	n
Total	$6188 = \binom{17}{5}$	$\binom{n+4}{5}$

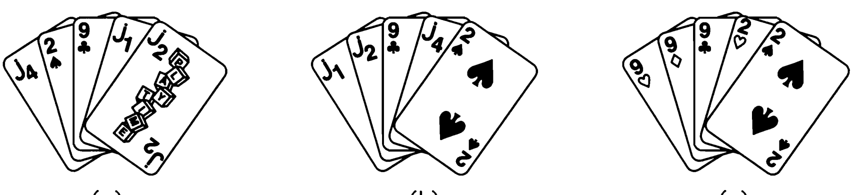
Surely such a simple answer can be found more simply? In fact, the hands correspond to the number of 5-card hands chosen from a *Sweet Seventeen* deck of 17 distinguishable cards: A, K, Q, J, 10, 9, 8, 7, 6, 5, 4, 3, and 2 and four distinguishable jokers: j_1, j_2, j_3, and j_4.

If you are dealt a Sweet Seventeen hand (Figure 3.4(a)), sort it in the usual way, high on the left, low on the right, but with any jokers in the positions corresponding to their labels (Figure 3.4(b)). To convert it into a poker hand, replace any jokers by copies of the first genuine card that follows them: Figure 3.4(c) shows the resulting full house, nines on twos.

To see why the correspondence is exact, convert your sorted poker hands (Figure 3.5(a)) into a Sweet Seventeen hand by replacing all duplicates of cards farther to the right by jokers, labeled by their position counting from the left (Figure 3.5(b)).

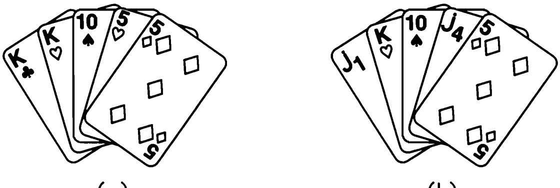
In general, to find the number of choices of r things from n different ones, but with repetitions allowed, imagine that you are playing Sweet Seventeen, but instead of a deck of 13 + 4 jokers, you have a deck of $n + (r - 1)$ jokers, and the answer is

$$\binom{n + r - 1}{r}.$$



(a) (b) (c)

FIGURE 3.4 *A "Sweet Seventeen" hand becomes a poker hand.*



(a) (b)

FIGURE 3.5 *A poker hand becomes a Sweet Seventeen hand.*

CHOICE NUMBERS ARE BINOMIAL COEFFICIENTS

Figure 3.6 will help you to do two of the algebraic manipulations in Figure 3.7, where the numbers, or **coefficients**, that appear are exactly those of Pascal's triangle.

You can see why this is so if you label the *b*s:

$$(a + b_1)(a + b_2) = a^2 + a(b_1 + b_2) + b_1b_2,$$

$$(a + b_1)(a + b_2)(a + b_3)$$
$$= a^3 + a^2(b_1 + b_2 + b_3) + a(b_1b_2 + b_1b_3 + b_2b_3) + b_1b_2b_3,$$

$$(a + b_1)(a + b_2)(a + b_3)(a + b_4)$$
$$= a^4 + a^3(b_1 + b_2 + b_3 + b_4)$$
$$+ a^2(b_1b_2 + b_1b_3 + b_2b_3 + b_1b_4 + b_2b_4 + b_3b_4)$$
$$+ a(b_1b_2b_3 + b_1b_2b_4 + b_1b_3b_4 + b_2b_3b_4) + b_1b_2b_3b_4.$$

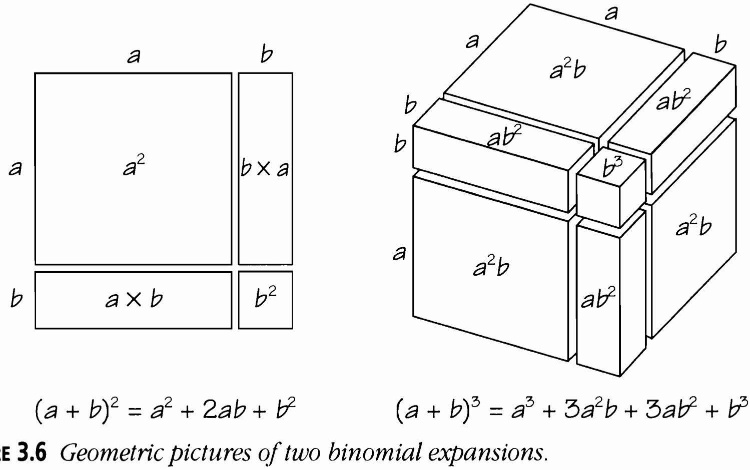


FIGURE 3.6 *Geometric pictures of two binomial expansions.*

$$
\begin{array}{rcll}
(a+b)^0 &=& 1 & 1 \\
(a+b)^1 &=& a+b & 1\ 1 \\
(a+b)^2 &=& a^2+2ab+b^2 & 1\ 2\ 1 \\
(a+b)^3 &=& a^3+3a^2b+3ab^2+b^3 & 1\ 3\ 3\ 1 \\
(a+b)^4 &=& a^4+4a^3b+6a^2b^2+4ab^3+b^4 & 1\ 4\ 6\ 4\ 1
\end{array}
$$

FIGURE 3.7 *Binomial expansions.*

Each term on the right comes from choosing either a or b from each of the binomials $(a + b)$ on the left. The number of terms with r bs and $n - r$ as is the number of choices of r bs from among the total of n bs, namely $\binom{n}{r}$. We've proved the **Binomial Theorem:**

$$
(a+b)^n = a^n + \binom{n}{1}a^{n-1}b + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{r}a^{n-r}b^r + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n
$$

Because there are two choices from each of the n binomial factors, the total number of products is 2^n. We check this by adding the rows of Pascal's triangle:

$$
\begin{array}{rcl}
1 &=& 2^0 \\
1+1 &=& 2^1 \\
1+2+1 &=& 2^2 \\
1+3+3+1 &=& 2^3 \\
1+4+6+4+1 &=& 2^4 \\
1+5+10+10+5+1 &=& 2^5 \\
1+6+15+20+15+6+1 &=& 2^6
\end{array}
$$

This is the result of setting $a = b = 1$ in the binomial theorem. If you let $a = 1$ and $b = -1$, you get the following.

$$
\begin{aligned}
1 - 1 &= 0\\
1 - 2 + 1 &= 0\\
1 - 3 + 3 - 1 &= 0\\
1 - 4 + 6 - 4 + 1 &= 0\\
1 - 5 + 10 - 10 + 5 - 1 &= 0\\
1 - 6 + 15 - 20 + 15 - 6 + 1 &= 0
\end{aligned}
$$

This is most obvious in the odd-numbered rows because of the symmetry, but it's also true in the even-numbered ones.

FRIEZE PATTERNS

In Figure 3.8(a) we've drawn a pattern bounded by a zigzag of zeros at the left and horizontal lines of zeros above and below. In Figure 3.8(b) we've used ones instead of zeros. Now fill in the question marks by the rule that the numbers *a* and *d* in each little diamond

```
   b
a     d
   c
```

add to 1 more than do *b* and *c* in Figure 3.8(a), while they *multiply* to 1 more than do *b* and *c* in Figure 3.8(b).

Some surprising things happen, as shown in Figure 3.9(a) and (b). For the *additive* pattern, part (a), the next zeros in each line form a copy of the initial zigzag, so the pattern repeats itself every seven

```
0 0 0 0 0 0 0 0 0            1 1 1 1 1 1 1 1 1
 0 ? ? ? ? ? ? ? ?            1 ? ? ? ? ? ? ? ?
  0 ? ? ? ? ? ? ? ?            1 ? ? ? ? ? ? ? ?
   0 ? ? ? ? ? ? ? ?            1 ? ? ? ? ? ? ? ?
  0 ? ? ? ? ? ? ? ?            1 ? ? ? ? ? ? ? ?
 0 ? ? ? ? ? ? ? ?            1 ? ? ? ? ? ? ? ?
  0 ? ? ? ? ? ? ? ?            1 ? ? ? ? ? ? ? ?
   0 0 0 0 0 0 0 0 0            1 1 1 1 1 1 1 1 1
```

(a) (b)

FIGURE 3.8 *Fill in these friezes, using the diamond rule.*

```
0  0  0  0  0  0  0  0
 0  1  2  5  6  4  3  0
  0  2  6 10  9  6  2  0
   0  5 10 12 10  4  1  0
  0  2  8 11 12  7  2  0
 0  1  4  8 10  8  4  0
  0  2  3  6  5  4  1  0
   0  0  0  0  0  0  0  0
```

(a)

```
1  1  1  1  1  1  1  1  1  1
 1  2  2  4  2  1  3  2  4  1
  1  3  7  7  1  2  5  7  3  1
   1 10 12  3  1  3 17  5  2  1
  1  3 17  5  2  1 10 12  3  1
 1  2  5  7  3  1  3  7  7  1
  1  3  2  4  1  2  2  4  2  1
   1  1  1  1  1  1  1  1  1  1
```

(b)

FIGURE 3.9 *Filled in frieze patterns repeat after so many steps.*

places. The ***multiplicative*** pattern, part (b), is even more surprising; all the divisions come out exactly, so that the entries are whole numbers. This time the ones in each row form an ***upside-down*** copy of the initial zigzag. We have to go a total of nine places in each row before we get an exact repetition.

The same sort of thing happens for arbitrary widths and shapes of initial zigzag, as you can verify by experiment. See if you can work out why.

For multiplicative frieze patterns, the essential observation is that for any six entries such as

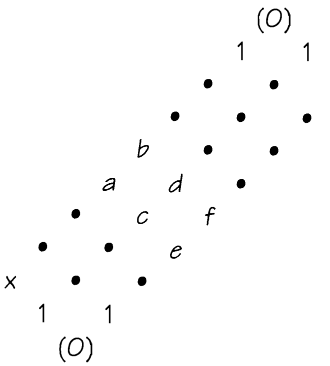
```
    b
a       d
    c       f
        e
```

we have $(a + e)/c = (b + f)/d$. Figure 3.10 shows how this implies that a number x just above the lower row of ones will reappear sometime later, just below the upper row of ones.

There are other ways of starting than by using a zigzag of 1s. In fact, you can use any diagonal sequence of numbers

$$1 = a_0, a_1, \ldots, a_n = 1 \text{ such that } a_i \text{ divides } a_{i-1} + a_{i+1}.$$

How many such sequences are there? We'll tell you the answer in the next chapter.



$$\frac{x+0}{1} = \ldots = \frac{a+e}{c} = \frac{b+f}{d} = \ldots = \frac{0+y}{1}$$

so $x = y$.

FIGURE 3.10 *How multiplicative frieze patterns reflect.*

HOW MANY REGIONS?

Every now and again someone comes up with a trick sequence where the general rule is not what it seems at first sight.

How many regions are there inside each of the six circles in Figure 3.11? The nth circle has n spots on the perimeter, joined in all possible ways, the spots having been chosen so that not more than two lines pass through any inside point.

We see that the answers are 1, 2, 4, 8, 16 for the first five circles. They are obviously powers of two. If you exercise just a little care in the last one, you can make it come to 32. And you might like to check that with 10 spots the answer comes to 256.

Let's try to prove that the answer for n spots is 2^{n-1}. Since 2^{n-1} is the number of subsets of the numbers $1, 2, \ldots, n-1$, we can do this by giving a rule that attaches one such subset to each region.

We label the points (0), $1, 2, \ldots, n-1$ counterclockwise around the circle (Figure 3.12). Because we want to end up with a subset of $\{1, 2, \ldots, n-1\}$, the digit 0 has only temporary status: we always omit it at the end.

Here's how to find the region corresponding to a set of numbers. The *meeting point* for a set of four numbers a, b, c, d ($a<b<c<d$) is where the chords ac, bd intersect. The meeting point of a set of two numbers, $c < d$, is the smaller of the two. Go to the meeting

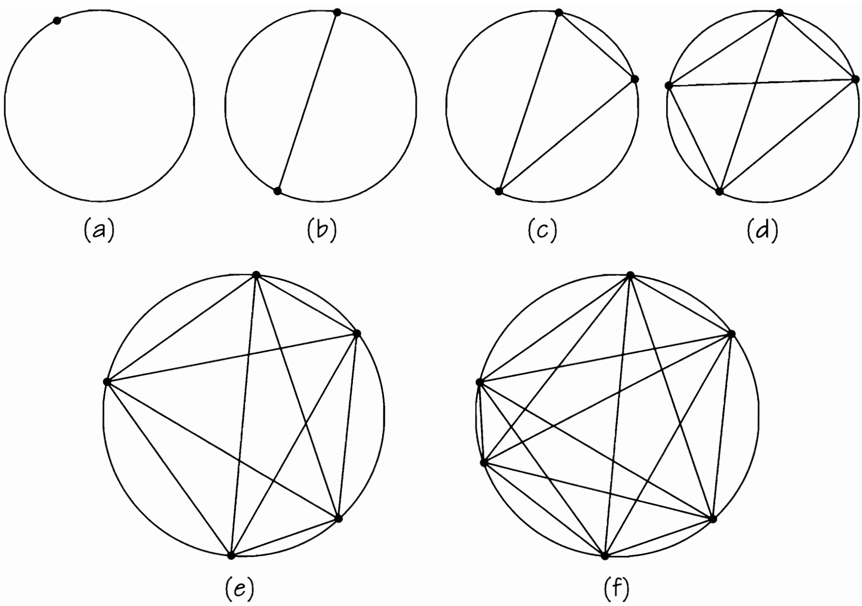


FIGURE 3.11 *A deceptive sequence.*

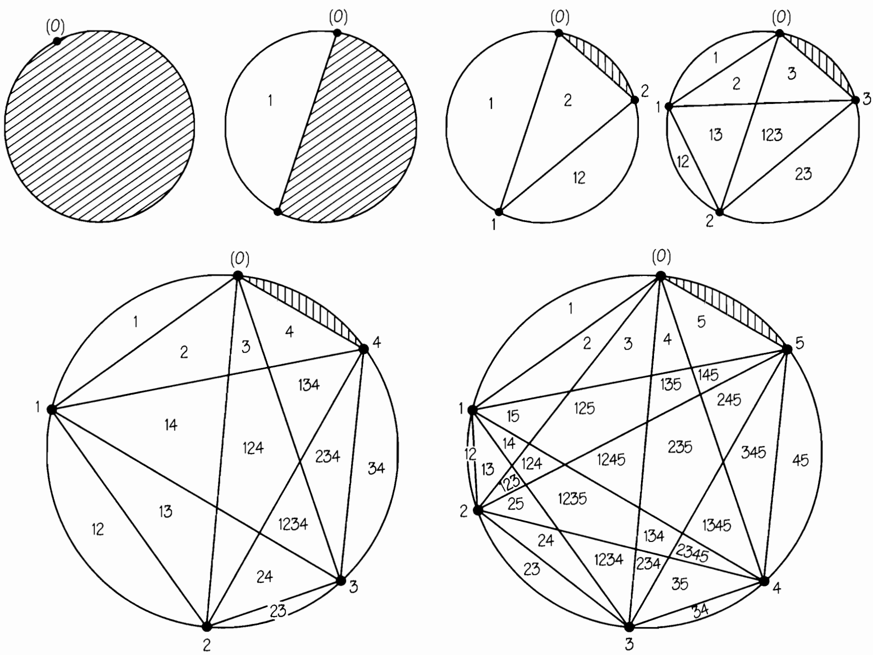


FIGURE 3.12 *The regions of Figure 3.11 all labeled.*

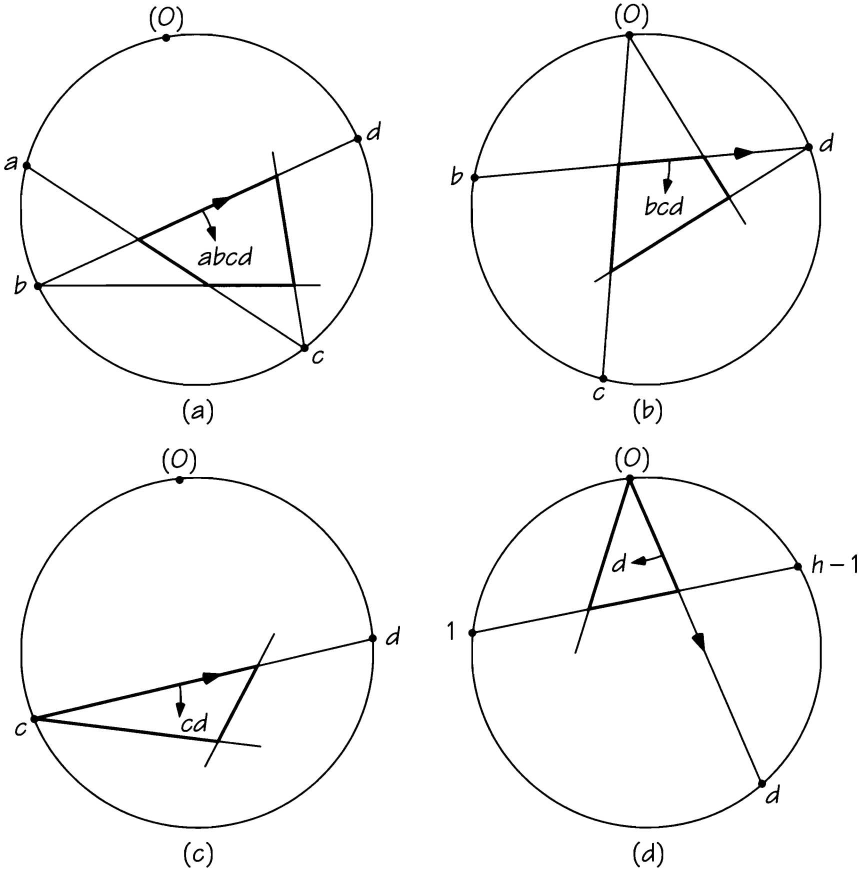


FIGURE 3.13 *Finding the region, given the label.*

point, look along the ray toward *d*, the largest number in the set. Then the region is on your right. For example, if $0 < a < b < c < d$, then the label *abcd* is assigned to the region shown in Figure 3.13(a); but if $a = 0$, as in Figure 3.13(b), then the region is labeled *bcd*, dropping the 0. If $0 < c < d$, we give the label *cd* to the region shown in Figure 3.13(c); and if $c = 0$, as in Figure 3.13(d), we simply label the region *d*. This leaves unlabeled the region on your *left* when you stand at 0 and look along the ray toward $n - 1$. This corresponds to the empty set and is shown shaded in Figures 3.13(d) and 3.12.

HOW CAREFUL WERE YOU?

We told you that you'd get the answer 32 for the 6-spot circle if you used just a little care. However, using just a bit more care, you'll find the answer 31. What you *won't* find is a region labeled 12345. The labels contain just 0, 1, 2, 3, or 4 numbers, so the number of regions is *not* 2^{n-1}, but rather the sum of the first five terms

$$\binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \binom{n-1}{3} + \binom{n-1}{4}$$

in the binomial expansion of $(1 + 1)^{n-1}$. This is *all* the terms for n up to 5. When $n = 6$, the last term

$$\binom{6-1}{5} = 1,$$

is missing, and for larger n, the answer falls increasingly short. The correct answers are, for

$n =$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
regions	1	2	4	8	16	31	57	99	163	256	386	562	794	1093

GUESSING THE NEXT TERM OF A SEQUENCE

As we've just seen, in some problems it's easy to guess the wrong answer. But if you always guess wrong, you'll fail that vital intelligence test. Here are some techniques that may guide you toward the right answer.

DIFFERENCING

Let's take the sequence we just found and try to work out the next term, supposing that we *hadn't* been able to find the general formula above.

We prepare a **difference table** in which each entry is the difference between the two entries just above it (in the sense "right minus left"). You'll see that in our case,

values	1	2	4	8	16	31	57	99	163
first differences	1	2	4	8	15	26	42	64	
second differences	1	2	4	7	11	16	22		
third differences	1	2	3	4	5	6			
fourth differences	1	1	1	1	1				

the fourth differences are all the same. We can guess that this pattern continues forever and use this guess to work out the next term by working upward (Figure 3.14).

So this correctly suggests that for a circle with 10 spots, the number of regions should be 256. We verify this by noting that exactly half of the 2^9 subsets of $\{1, 2, \ldots, 9\}$ have at most four members.

Some easy algebra shows that if we started with the sequence of values of a dth-degree polynomial, then its first difference sequence will be the values of some degree $d-1$ polynomial, its second difference sequence will be the values of some degree $d-2$ polynomial, and its dth difference sequence will be the values of some de-

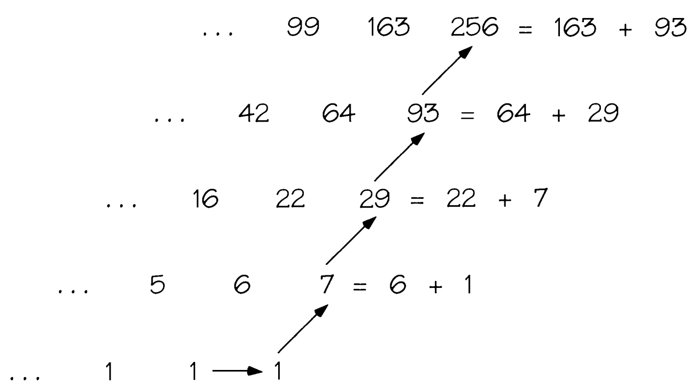


FIGURE 3.14 *Calculating the next term, assuming that the fourth differences are constant.*

gree zero polynomial. In other words, its *d*th differences will all have the same constant value.

For example, let's difference the sequence of cubes:

values	0	1	8	27	64	125	216	343	512	729	1000
first differences	1	7	19	37	61	91	127	169	217	271	
second differences	6	12	18	24	30	36	42	48	54		
third differences	6	6	6	6	6	6	6	6			

The *n*th differences of *n*th powers are all equal to the *n*th factorial number, *n*!.

NEWTON'S USEFUL LITTLE FORMULA

When you *do* find constant differences, how do you work out the polynomial? You can find the answer by forming the differences for the binomial coefficients, $\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots$

$\binom{n}{0}$

$n =$ 0	1	2	3	4	5
1	1	1	1	1	1
0	0	0	0	0	
0	0	0	0		
0	**0**	**0**			

$\binom{n}{1}$

$n =$ 0	1	2	3	4	5
0	1	2	3	4	5
1	1	1	1	1	
0	0	0	0		
0	**0**	**0**			

$\binom{n}{2}$

$n =$ 0	1	2	3	4	5
0	0	1	3	6	10
0	1	2	3	4	
1	1	1	1		
0	**0**	**0**			

$\binom{n}{3}$

$n =$ 0	1	2	3	4	5
0	0	0	1	4	10
0	0	1	3	6	
0	1	2	3		
1	**1**	**1**			

If you look at the bold numbers, you'll see that the sequence with difference table of shape

$n =$	0	1	2	3	4	5	6
	A	?	?	?	?	?	?
		B ?	?	?	?	?	
			C ?	?	?	?	
			D	**D**	**D**	**D**	

is

$$A\binom{n}{0} + B\binom{n}{1} + C\binom{n}{2} + D\binom{n}{3},$$

that is,

$$A + Bn + \frac{1}{2}Cn(n - 1) + \frac{1}{6}Dn(n - 1)(n - 2).$$

For example, when we differenced the cubes, we found

$$A = 0, \quad B = 1, \quad C = 6, \quad D = 6,$$

and

$$0\binom{n}{0} + 1\binom{n}{1} + 6\binom{n}{2} + 6\binom{n}{3} =$$

$$0 + n + 3n\,(n - 1) + n(n - 1)(n - 2) = n^3.$$

A similar pattern works for higher-degree polynomials. If, starting from the *t*th term, you get a table in which the *d*th-degree differences are all *K*:

	$n =$	t	$t+1$	$t+2$	$t+3$ ...
values		A	?	?	? ...
first differences			B	?	? ? ...
second differences				C	? ? ...
....					....
d th. differences				K	K K ...

then the nth term is

$$A\binom{n-t}{0} + B\binom{n-t}{1} + C\binom{n-t}{2} + \cdots + K\binom{n-t}{d}.$$

This correctly gives

$$1\binom{n-1}{0} + 1\binom{n-1}{1} + 1\binom{n-1}{2} + 1\binom{n-1}{3} + 1\binom{n-1}{4}$$

$$= \frac{1}{24}(n^4 - 6n^3 + 23n^2 - 18n + 24)$$

for our "regions" problem, in which $t, A, B, \ldots, K$ were all 1.

HOW MANY TRIANGLES?

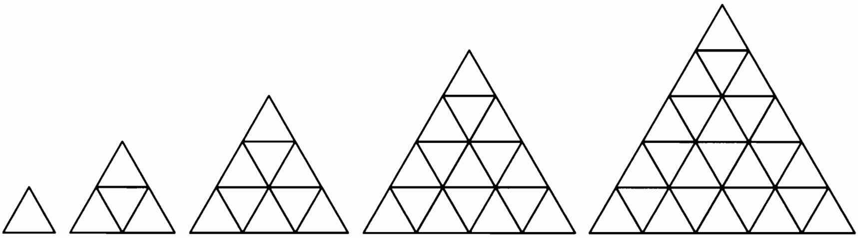


FIGURE 3.15 *How many triangles in each of these diagrams?*

A good way to find the nth term is to count the first few very carefully, and then make a difference table. If you draw more figures than we've done in Figure 3.15, and count all the triangles (including the upside-down ones) you'll find the numbers and their differences are

1 5 13 27 48 78 118
4 8 14 21 30 40
4 6 7 9 10
2 1 2 1

The third differences alternate between 2 and 1 so the answer is given alternately by two expressions:

$$\frac{n(n+2)(2n+1)}{8} \qquad \frac{n(n+2)(2n+1)-1}{8}$$

for even n. for odd n.

Even when your sequence does not come from a polynomial, differencing is often informative. Figures 3.16 and 3.17 show how the powers of 2, and the Fibonacci numbers (which we'll learn more about in the next chapter), repeat themselves when differenced.

```
2^n  1   2   4   8   16   32   64   128  256  512  1024...
       1   2   4   8   16   32   64   128  256  512 ...
         1   2   4   8   16   32   64   128  256...
```

FIGURE 3.16 *Difference table for the powers of 2.*

```
0   1   1   2   3   5   8   13   21   34   55   89...
  1   0   1   1   2   3   5   8   13   21   34...
   -1   1   0   1   1   2   3   5   8   13...
```

FIGURE 3.17 *Difference table for the Fibonacci numbers.*

So if, as in Figure 3.18, some row of another difference table becomes powers of 2, the original sequence differs from the powers of 2 only by the values of polynomial. This is similar for other sequences of powers and for the Fibonacci numbers as well.

```
1   3   6   11   20   37   70   135  264...
  2   3   4   9   17   33   65   129...
    1   2   4   8   16   32   64...
```

FIGURE 3.18 *A sequence not very different from the powers of 2.*

JACKSON'S DIFFERENCE FANS

Robert Jackson suggests that if you've completed a difference table and still don't understand the sequence, you should turn the paper through an angle of 60°, say, and start again and perhaps repeat this several times to make a *fan* of difference tables.

Jackson found that this fanning process converts powers of 4 to powers of 3 to powers of 2 to powers of 1 to powers of 0 (Figure 3.19). In general, the sequence k^n, multiplied by any polynomial in n, is reduced to zeros by a k-fold difference fan.

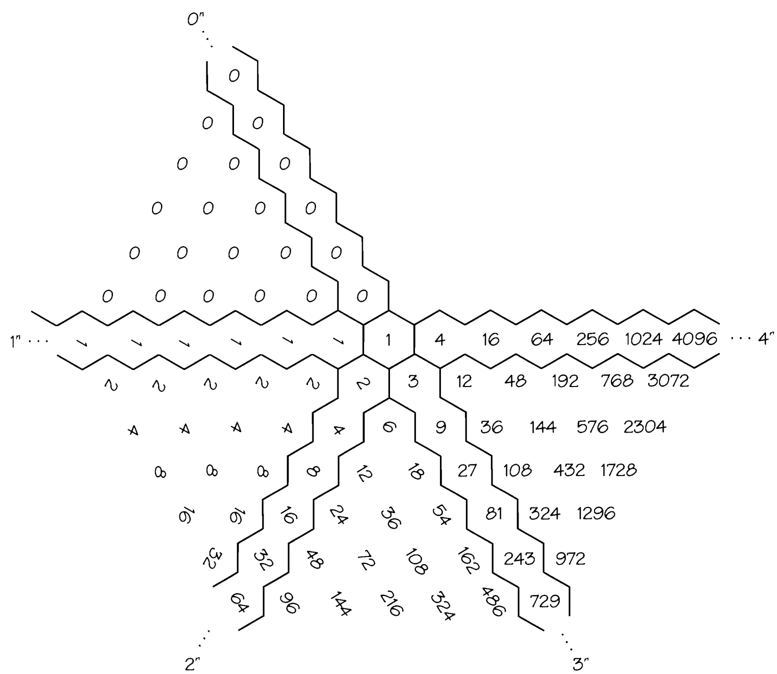


FIGURE 3.19 *A fourfold difference fan.*

It's a good idea to try a difference fan on any sequence that you suspect is given by a simple formula involving powers of the first few numbers. Figure 3.20 shows what happens to one such sequence.

NUMBER WALLS

If you suspect that your sequence is not a polynomial, but is computed by the kind of "bootstrapping" rule that defines the Fibonacci numbers (where each term is the sum of the two previous ones),

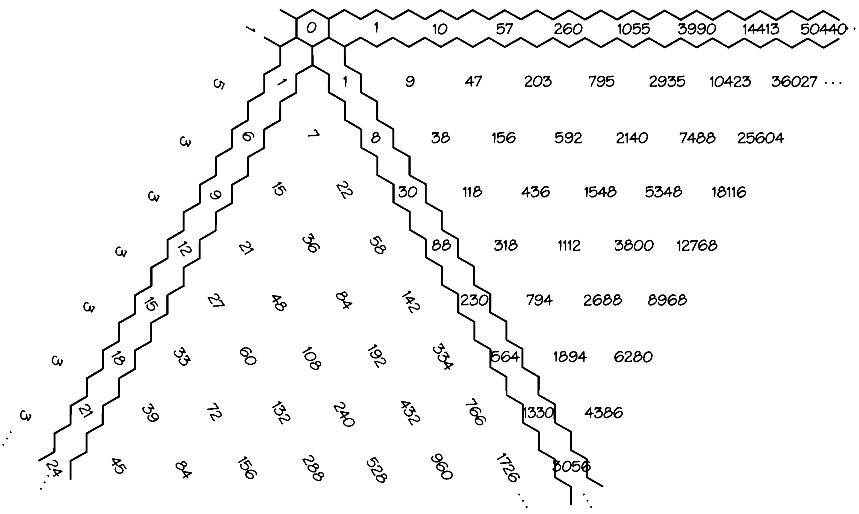


FIGURE 3.20 *A threefold difference fan for the sequence* $n(3^n - 2^n)$.

then you can detect this by a **number wall** (or "quotient-difference table") rather than an ordinary difference table.

The top row of a number wall (e.g., Figure 3.21) is a string of ones, and beneath these we place the terms of your suspect sequence. Further entries are computed by the rule that for every cross of five adjacent numbers,

$$\begin{matrix} & N & \\ W & X & E \\ & S & \end{matrix} \quad \text{we have} \quad X^2 = NS + EW.$$

Then, if some row consists entirely of zeros, your suspect sequence is indeed a bootstrapping sequence, in which each term can be computed as the sum of fixed multiples of the previous *k* terms, where

1	1	1	1	1	1	1
0	1	5	19	65	211	665
	1	6	36	216	1296	
		0	0	0		

FIGURE 3.21 *A number wall or quotient-difference table.*

k is the number of intermediate rows in the wall (between the top row of ones and the bottom row of zeros).

So Figure 3.21 shows that 0, 1, 5, 19, 65, 211, 665 is part of a bootstrapping sequence in which each term is the sum of fixed multiples of its two predecessors (a "second-order recurrence"). You can now find further terms by working north-eastward using

$$E = (X^2 - NS)/W$$

as in Figure 3.22. In fact, the typical term here is $3^n - 2^n$.

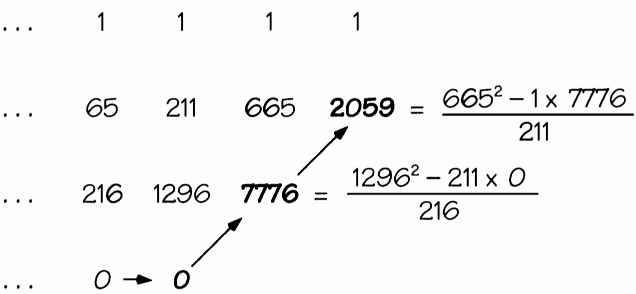


FIGURE 3.22 *Calculating the next entries in a quotient-difference table.*

WALLS HAVE WINDOWS

Our rule for number walls isn't complete, because sometimes you'll have to divide zero by zero! Fred Lunnon first told us about the remarkable fact that the zeros in a number wall form square "windows" bordered by geometric progressions. Figure 3.23 shows an example. To get the numbers just below a window, you must use another rule, explained in Figure 3.24.

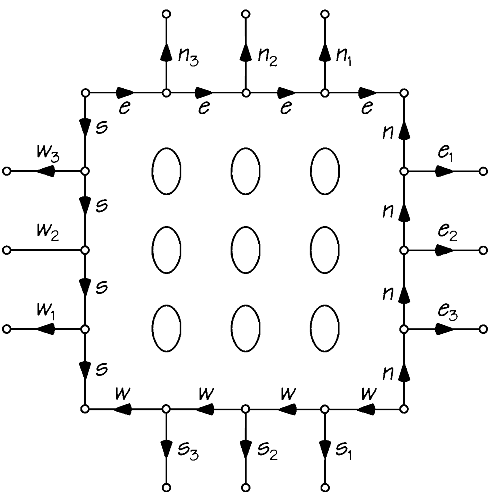
It's easier to work past an isolated zero

$$\begin{array}{ccccc} & & N' & & \\ & & N & & \\ W' & W & \boxed{0} & E & E' \\ & & S & & \\ & & S' & & \end{array}$$

using the fact that $S'N^2 + N'S^2 = E'W^2 + W'E^2$.

1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	–4	1	–2	0	–1	0	0	1	2	4	7	12	20
	12	–7	4	–2	1	0	0	1	0	2	1	4	
		1	–1	1	–1	1*	–1*	1	–1	1	–1		
			0	0	0	0*	0*	0	0	0			

FIGURE 3.23 *Zeros appear in windows in the quotient-difference table for the Fibonacci numbers minus one.*



The numbers n, s, e, w, n_1, e_2, . . . on the arrows are the ratios of the entries, "head/tail".

$ns = ew$ for even windows

$ns = -ew$ for odd windows

$$\frac{s_1}{s} = \frac{n_1}{n} - \frac{w_1}{w} + \frac{e_1}{e}$$

$$\frac{s_2}{s} = \frac{n_2}{n} + \frac{w_2}{w} - \frac{e_2}{e}$$

$$\frac{s_3}{s} = \frac{n_3}{n} - \frac{w_3}{w} + \frac{e_3}{e}$$

FIGURE 3.24 *How to work past a window.*

As you can see from Figure 3.23, the number wall for the Fibonacci numbers minus one has some windows, so the more complicated rule is needed to work out the starred entries.

All entries in the number wall for a sequence of whole numbers will be whole numbers. Just as in frieze patterns, this provides a useful check for your arithmetic.

What do you do if you come across a sequence of numbers and don't know what they are? You can look in Sloane & Plouffe's wonderful *Encyclopedia of Integer Sequences,* or you can email to sequences@research.att.com with a line that reads, say,

lookup 1 1 2 3 5 8 12

or whatever the sequence is that you're interested in.

REFERENCES

J. H. Conway, and H. S. M. Coxeter. Triangulated polygons and frieze patterns. *Math. Gaz.,* **57**(1973): 87-94, 175-183; *MR* **57**, no. 1254-5.

Martin Gardner. *Mathematical Magic Show.* Math. Assoc. Amer., 1990; Chapter 4, Factorials.

Martin Gardner. *Mathematical Circus.* Math. Assoc. Amer., Spectrum Series, 1992; Chapter 13, Fibonacci and Lucas Numbers; Chapters 15, Pascal Triangle.

Calvin T. Long. On the Moessner theorem on integral powers. *Amer. Math. Monthly,* **73**(1966): 846-851 (and papers of O. Perron, I. Paasche, and H. Salié quoted there).

Calvin T. Long. Strike it out—add it up. *Math. Mag.,* **66**(1982): 273-277.

Yosho Mikami. *The Development of Mathematics in China and Japan,* 1913.

Alfred Moessner. Eine Bemerkung über die Potenzen der natürlichen Zahlen. *S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss.,* 1951, **29**(1952); *MR* **14**, 353b.

Ivan Paasche. Ein neuer Beweis des Moessnerschen Satzes. *S.-B. Math.-Nat. Kl. Bayer. Akad. Wiss.,* **1952**(1953): 1-5; *MR* **14**, 8469.

Ivan Paasche. Ein zahlentheoretische-logarithmischer "Rechenstab." *Math. Naturwiss. Unterr.,* **6**(1953/54): 26-28.

Ivan Paasche. Beweis des Moessnerschen Satzes mittels linearer Transformationen. *Arch. Math.,* **6**(1955): 194-199; *MR* **17**, 237d.

Paasche, Ivan. Eine Verallgemeinerung des Moessnerschen Satzes, *Compositio Math.,* **12**(1956): 263-270.

Joe Roberts. *Lure of the Integers.* Math. Assoc. of America, Spectrum series, 1992.

N. J. A. Sloane and S. Plouffe. *The Encyclopedia of Integer Sequences.* Academic Press, New York, 1995.

4

Famous Families of Numbers

Many families of numbers arise again and again in many different mathematical problems: Often they have been named after the mathematicians who investigated them. In this chapter we'll meet Bell and Stirling, Ramanujan, Catalan, Bernoulli and Euler, Fibonacci and Lucas.

Bell Numbers and Stirling Numbers

Many of these numbers arise from counting arrangements of various kinds. How many ways are there of arranging n objects in groups? If your objects are clearly distinguishable, the answer is usually called the **Bell number**, after Eric Temple Bell, who was known as a mathematician, a mathematical historian, and the author of several detective stories, under the name John Taine.

Figure 4.1 shows all the ways of grouping 4 pieces of luggage and shows that the fourth Bell number, b_4, is 15. The number of group-

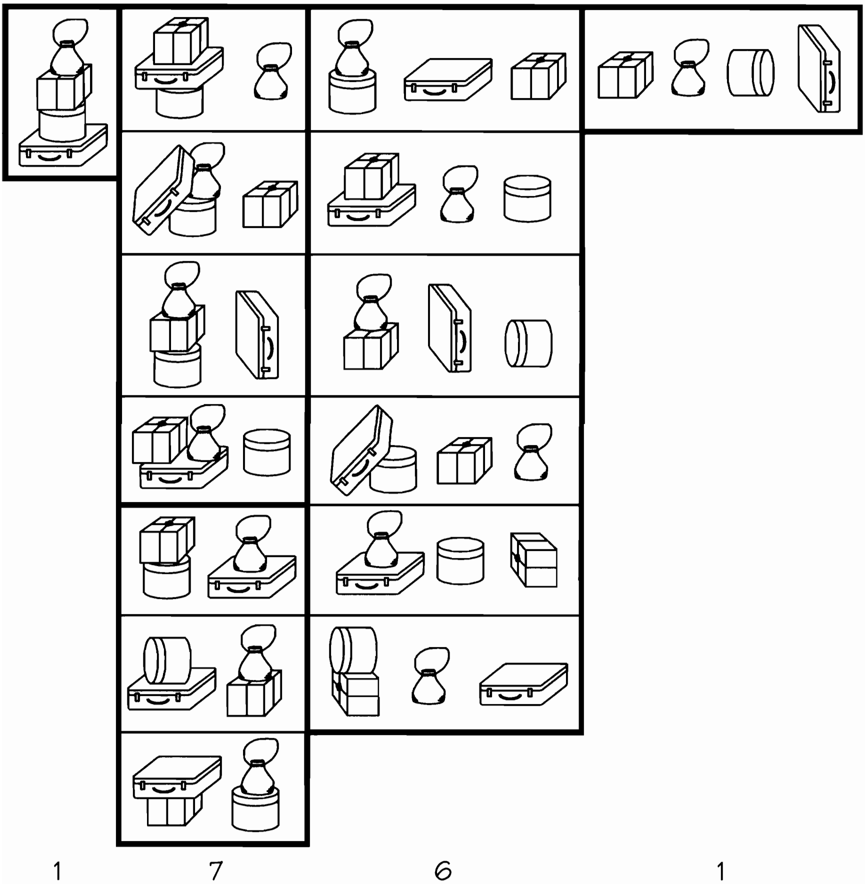


FIGURE 4.1 *The Bell number $b_4 = 15$, and the Stirling numbers $\left\{ {4 \atop k} \right\} = 1, 7, 6, 1$.*

ings of n distinct things into exactly k groups is the **Stirling set number**, $\left\{ {n \atop k} \right\}$, traditionally called the Stirling number of the second kind. The four columns in Figure 4.1 show that

$$\left\{ {4 \atop 1} \right\} = 1, \left\{ {4 \atop 2} \right\} = 7, \left\{ {4 \atop 3} \right\} = 6, \left\{ {4 \atop 4} \right\} = 1.$$

Here are some more Stirling set numbers and Bell numbers. The

n	Stirling set numbers, $\left\{ n \atop k \right\}$						Bell numbers, b_n
1	1						1
2	1	1					2
3	1	3	1				5
4	1	7	6	1			15
5	1	15	25	10	1		52
6	1	31	90	65	15	1	203

TABLE 4.1 *Stirling set numbers and Bell numbers.*

*n*th **Bell number,** b_n, is the sum of the Stirling set numbers in each row.

$$b_n = \left\{ n \atop 1 \right\} + \left\{ n \atop 2 \right\} + \cdots + \left\{ n \atop n \right\}$$

and is the total number of ways of arranging n objects into groups.

The **Stirling cycle numbers,** $\left[n \atop k \right]$ (traditionally, Stirling numbers of the first kind) count the permutations of n objects that have just k cycles. For example, the 6 permutations of 3 objects are classified

1 cycle	2 cycles	3 cycles
(123) (132)	(12)(3) (13)(2) (23)(1)	(1)(2)(3)

$$\left[n \atop 1 \right] + \left[n \atop 2 \right] + \cdots + \left[n \atop n \right] = n!$$

They appear in the following table:

n	Stirling cycle numbers, $\left[n \atop k \right]$					$n!$
1	1					1
2	1	1				2
3	2	3	1			6
4	6	11	6	1		24
5	24	50	35	10	1	120

TABLE 4.2 *Stirling cycle numbers.*

The triangles in Tables 4.1 and 4.2 are computed by variants of the Pascal's triangle rule

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1},$$

namely

$$\left\{ {n+1 \atop k} \right\} = k\left\{ {n \atop k} \right\} + \left\{ {n \atop k-1} \right\}$$

and

$$\left[{n+1 \atop k} \right] = n\left[{n \atop k} \right] + \left[{n \atop k-1} \right].$$

Algebraically, $\left\{ {n \atop k} \right\}$ and $\left[{n \atop k} \right]$ are the coefficients of x^n and x^k respectively, in $1/(1-x)(1-2x)\cdots(1-kx))$ and $(1+x)(1+2x)\cdots(1+nx)$.

PARTITION NUMBERS AND COMPOSITIONS; RAMANUJAN'S NUMBERS

If your n objects are indistinguishable, then the number of ways of grouping them is called the **partition number**, $p(n)$, so Figure 4.2 shows that $p(4) = 5$ (the five areas of Figure 4.1).

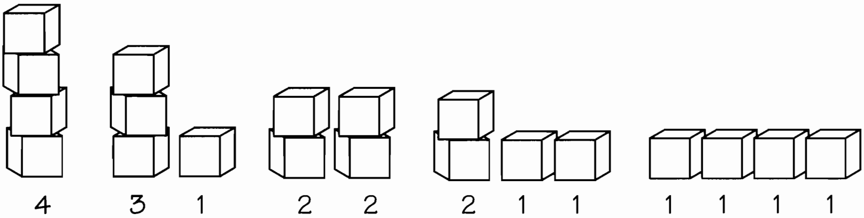


FIGURE **4.2** *The five partitions of 4.*

Here are some more partitions and partition numbers. Superscripts in the partitions indicate repetitions, for instance, $2^2 1^3$ means 22111.

n	$p(n)$	partitions
0	1	
1	1	1
2	2	2, 11
3	3	3, 21, 111
4	5	4, 31, 22, 211, 1111
5	7	5, 41, 32, 311, 221, 2111, 11111
6	11	6, 51, 42, 411, 33, 321, 3111, 222, 2211, 21111, 111111
7	15	7, 61, 52, 51^2, 43, 421, 41^3, 3^21, 32^2, 321^2, 31^4, 2^31, 2^21^3, 21^5, 1^7
8	22	8, 71, 62, 61^2, 53, 521, 51^3, 4^2, 431, 422, 421^2, 41^4, 3^22, 3^21^2, 32^21, 321^3, 31^5, 2^4, 2^31^2, 2^21^4, 21^6, 1^8
9	30	
10	42	
11	56	

There's no simple exact formula for $p(n)$, but there's a remarkable approximate formula guessed by the famous Indian mathematician, Srinivasa Ramanujan:

$$p(n) \text{ is approximately } \frac{1}{4n\sqrt{3}}\, e^{\pi\sqrt{2n/3}}.$$

This was later proved by Hardy & Ramanujan and later modified by Rademacher to an exact formula:

$$p(n) = \frac{1}{\pi\sqrt{2}} \sum_{k=1}^{\infty} A_k(n) k^{1/2} \left[\frac{d}{dx} \frac{\sinh((\pi/k)(\frac{2}{3}(x - \frac{1}{24}))^{1/2})}{(x - \frac{1}{24})^{1/2}} \right]_{x=n}$$

where $A_k(n) = \sum_{\substack{h \bmod k \\ (h,k)=1}} \omega_{h,k} e^{-2\pi inh/k}$ with $\omega_{h,k}$ a certain 24th root of unity.

In counting partitions, we are not concerned with the *order* of the parts. If you *do* consider the order, the answer is much simpler. There are exactly 2^{n-1} ordered partitions (or **compositions**, as MacMachon called them) of n. For example, the 16 compositions of 5 are

5, 41, 14, 32, 23, 311, 131, 113, 221, 212, 122, 2111, 1211, 1121, 1112, 11111.

(A strip of 5 squares can be cut into shorter strips of squares, at any of 4 possible places, so there are 2^4 ways of doing it.)

Perhaps the strangest formula for the partition numbers is Euler's way of computing them, using pentagonal numbers:

$$p(n) - p(n-1) - p(n-2) + p(n-5) + p(n-7) - p(n-12) - p(n-15) + \cdots = 0^n$$

($0^n = 1$ if $n = 0$ and $0^n = 0$ if n is positive.) Here the signs alternate in pairs and the numbers subtracted from n are the pentagonal numbers that we met in Chapter 2, including those for negative values of k. These are much easier to remember if you recall from Chapter 2 that they are one-third of those triangular numbers that are divisible by 3:

triangles	0	1	3	6	10	15	21	28	36	45	55	66	78	91	105	120	...
pentagons	0	*	1	2	*	5	7	*	12	15	*	22	26	*	35	40	...
and signs	+		−	−		+	+		−	−		+	+		−	−	

The signs are plus for pairs of terms having the same parity, minus for pairs with different parity. For example, for $n = 12$ we get

$$p(12)-p(12-1)-p(12-2)+p(12-5)+p(12-7)-p(12-12)=0$$

or $p(12) - 56 - 42 + 15 + 7 - 1 = 0$, so that $p(12) = 77$.

The partition numbers may be defined algebraically by the formula

$$\frac{1}{(1-x)(1-x^2)(1-x^3)\cdots} = 1 + p(1)x + p(2)x^2 + p(3)x^3 + \cdots$$

The **Ramanujan numbers**, $\tau(n)$ are similarly defined by

$$x[(1-x)(1-x^2)(1-x^3)\cdots]^{24} = \tau(1)x + \tau(2)x^2 + \tau(3)x^3 + \cdots$$

Ramanujan found many astonishing properties of these numbers: For instance, $\tau(m)\tau(n) = \tau(mn)$ whenever m and n have no common factor, and $\tau(n)$ is congruent to the sum of the eleventh powers of the divisors of n, modulo 691.

A PACK OF PROBLEMS

Orpheus with his lute made *trees*
And the *mountain* tops that *freeze*
Bow themselves . . .

In the next few pages we'll describe lots of problems that all have the same answer, and then we'll show you why the answer is always the same.

FRIEZE PATTERNS

How many different diagonals are possible in a frieze pattern with $n+1$ rows? The answers for $n = 1, 2, 3$ are 1, 2, 5, respectively (Figure 4.3).

1 1 1 ...
1 1 1 ...

1 1 1 1 1 ...
1 2 1 2 1 ...
1 1 1 1 1 ...

1 1 1 1 1 1 1 ...
1 2 2 1 3 1 2 ...
1 3 1 2 2 1 3 ...
1 1 1 1 1 1 1 ...

FIGURE 4.3 *Frieze patterns with n + 1 rows.*

CHOPPING POLYGONS INTO TRIANGLES

How many ways are there of chopping a given (n+2)-sided polygon into n triangles? **(Figure 4.4)**

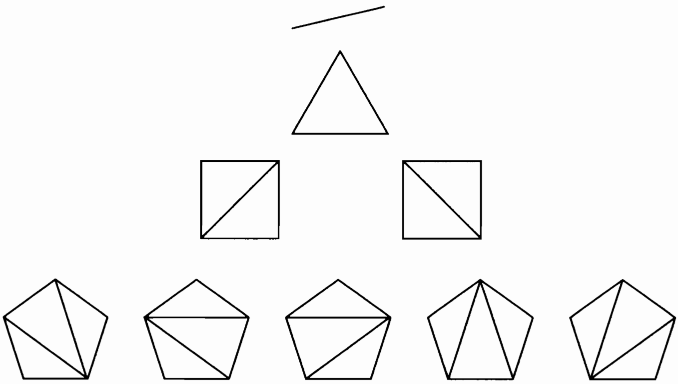


FIGURE 4.4 *Chopping polygons into triangles.*

BIFURCATING TREES

How many rooted plane binary trees are there with n internal nodes? (Figure 4.5)

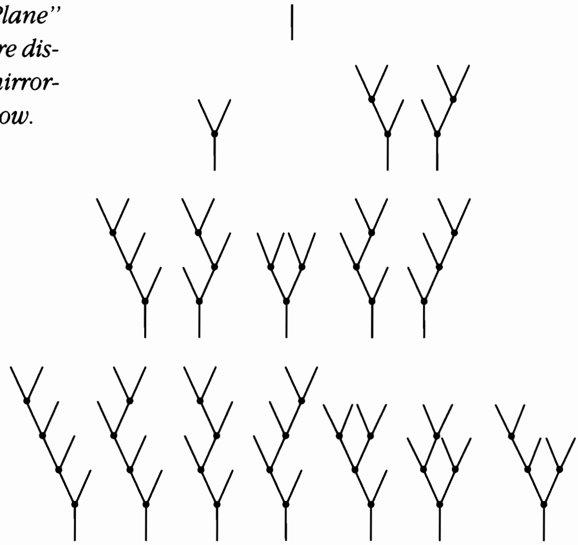


FIGURE 4.5 *Binary trees: "Plane" means that left and right are distinguishable. Add the mirror-images of trees in the last row.*

EVALUATING LADDERED EXPONENTS

How many values can you expect from an n-fold exponential? (Figure 4.6)

$$3^2 = 9$$

$$(4^3)^2 = 4^6, \quad 4^{(3^2)} = 4^9,$$

$$((5^4)^3)^2 = 5^{24}, \quad 5^{(4^{(3^2)})} = 5^{262144}, \quad (5^{(4^3)})^2 = 5^{128}, \quad (5^4)^{(3^2)} = 5^{36}, \quad 5^{((4^3)^2)} = 5^{4096}.$$

FIGURE 4.6 *Laddered exponents give varied values.*

ROOTED PLANE BUSHES

How many rooted plane bushes are there with n edges? (Figure 4.7)

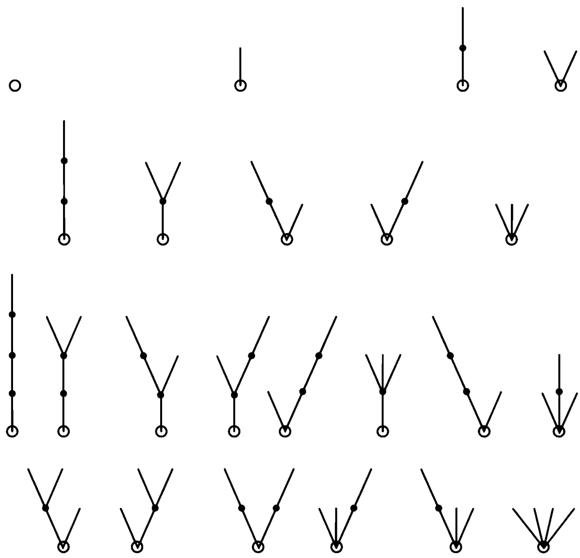


FIGURE 4.7 *Rooted plane bushes.*

MOUNTAINS

How many mountains can you draw with n upstrokes and n down-strokes? (Figure 4.8)

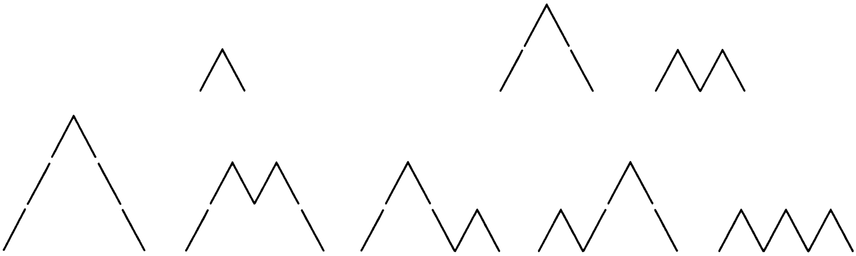


FIGURE 4.8 *Mountains.*

PARENTHESES

How many ways are there of arranging n pairs of parentheses? (Figure 4.9)

$n = 1$: ( ); $n = 2$: (( )), ( ) ( );

$n = 3$: ((( ))), (( ) ( )), (( )) ( ), ( ) (( )), ( ) ( ) ( )

FIGURE 4.9 *Arrangements of n pairs of parentheses.*

HANDS ACROSS THE TABLE

How many noncrossing handshakes are possible with n pairs of people? (Figure 4.10)

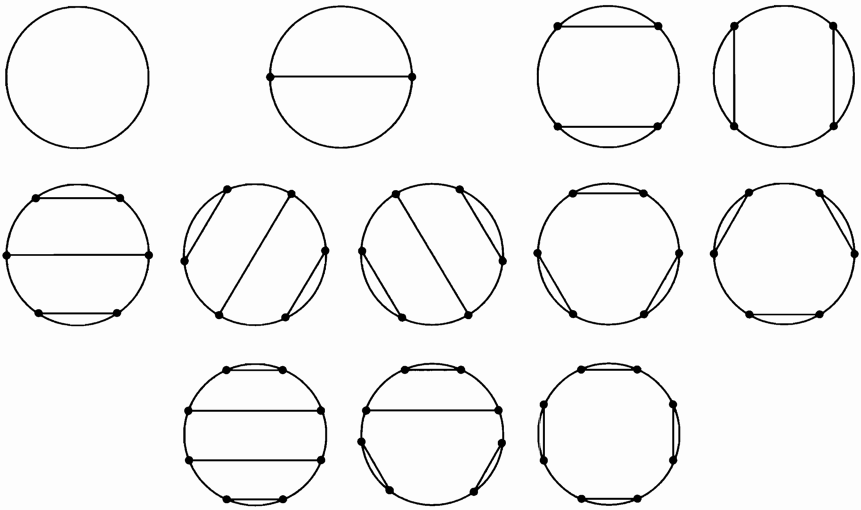


FIGURE 4.10 *Shaking hands without crossing. The last line gives 14 ways by rotating.*

CATALAN NUMBERS

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 1
1 7 21 35 35 21 7 1
1 8 28 56 70 56 28 8 1

FIGURE 4.11 *The central binomial coefficients.*

Look at the middle numbers in Pascal's triangle (Figure 4.11):

$$1, 2, 6, 20, 70, 252, 924, 3432, 12870, 48620, \ldots.$$

It seems that we can divide them by

$$1, 2, 3, 4, 5, 6, 7, 8, 9, 10, \ldots.$$

to give the whole number sequence

$$1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, \ldots.$$

A typical middle number is the binomial coefficient, $\binom{2n}{n}$, so our guess is that

$$\frac{1}{n+1}\binom{2n}{n} = \frac{2n!}{n!(n+1)!} = \frac{1}{2n+1}\binom{2n+1}{n} = c_n,$$

say, is a whole number. These numbers are called **Catalan numbers**.

Now we'll show that all the preceding problems have the same answer.

FRIEZES AND POLYGONS

The correspondence between friezes and polygons is fairly easy to describe. The frieze pattern (Figure 4.12) corresponds to the triangulated polygon of Figure 4.13 because the row just below the top

1 1 1 1 1 1 ...

1 2 2 1 3 1 ...

1 3 1 2 2 1 ...

1 1 1 1 1 1 ...

FIGURE 4.12 *A frieze pattern with five different diagonals.*

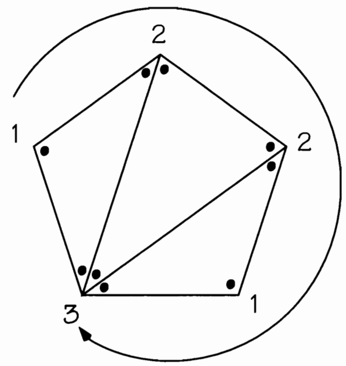


FIGURE 4.13 *A dissected polygon with the numbers of triangles at each vertex.*

row of ones in Figure 4.12 tells you the numbers of triangles at the vertices of the polygon.

Alternatively, in Figure 4.14, leave the lower left vertex unlabeled, and label all vertices that are joined to it with 1s. Then label other vertices so that the number at the last vertex of any triangle is the sum of the other two. Read around the labels and you have the five different diagonals of Figure 4.12.

A more complicated pattern and polygon appeared in Chapter 3. Conway and Coxeter have given a full explanation of why this works.

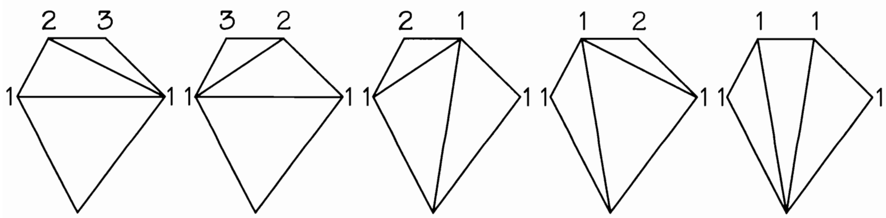


FIGURE 4.14 *Polygons labeled with frieze diagonals.*

POLYGONS AND TREES

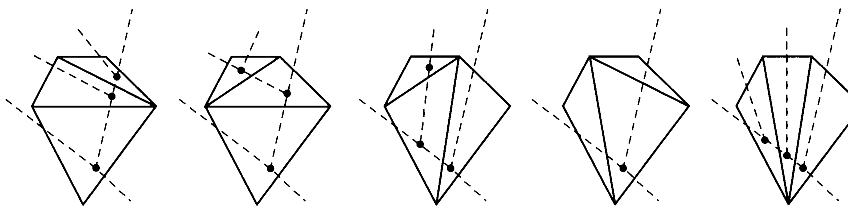


FIGURE 4.15 *Binary trees make paths through polygons.*

Figure 4.15 shows how chopped polygons correspond to plane binary trees.

TREES AND EXPONENTIATIONS

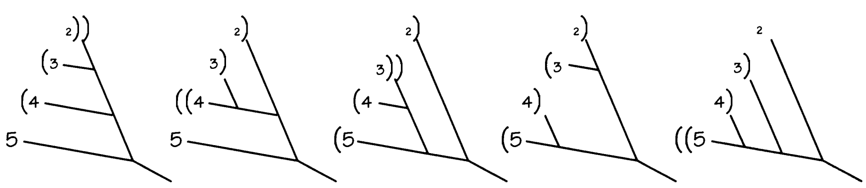


FIGURE 4.16 *Binary trees are related to orders of operation when exponentiating.*

Figure 4.16 relates binary trees to exponential expressions, and Figure 4.17 associates the latter to bushes.

EXPONENTIALS AND BUSHES

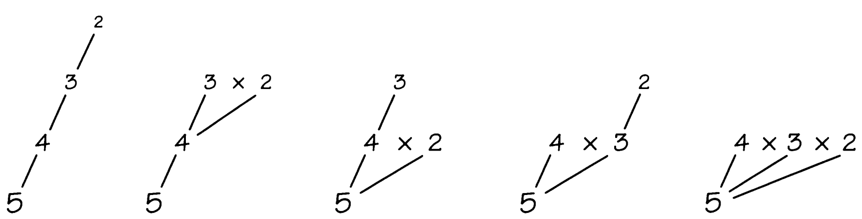


FIGURE 4.17 *Parentheses have been eliminated, using* $(a^b)^c = a^{b\times c}$.

BUSHES AND MOUNTAINS

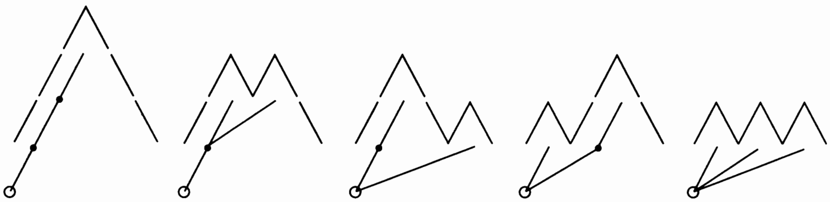


FIGURE 4.18 *Shapes of bushes and shapes of mountains.*

MOUNTAINS AND PARENTHESES

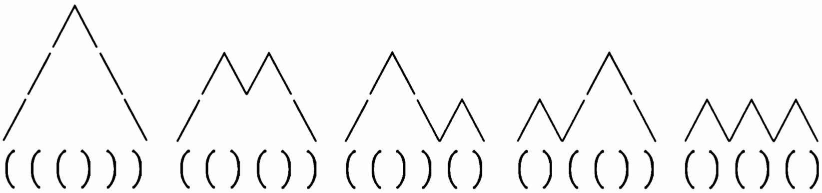


FIGURE 4.19 *Patterns of parentheses and shapes of mountains.*

PARENTHESES AND SHAKING HANDS

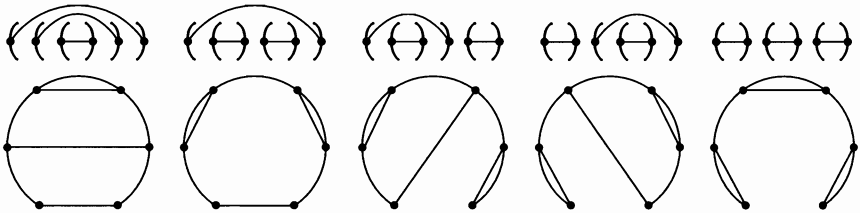


FIGURE 4.20 *Parentheses show which pairs of people shake hands.*

Figure 4.18 links bushes to mountains; Figure 4.19 links mountains to nests of parentheses; and Figure 4.20 connects these with handshakes.

So all these varied objects (and indeed many more) are all counted by the same numbers! To see that they are the Catalan numbers, it's perhaps easiest to count the mountains.

MOUNTAINS AND CATALAN NUMBERS

If we add in an extra upstroke, there are $7!/4!3! = 35$ sequences of 4 upstrokes and 3 downstrokes, but if we continue these patterns periodically (Figure 4.21) we get only 5 different infinite sequences, which break naturally at the dashed edges to reveal the 5 different mountains with 3 upstrokes and 3 downstrokes.

In general, the mountains with n upstrokes and n downstrokes correspond to the

$$\frac{1}{2n+1}\binom{2n+1}{n}$$

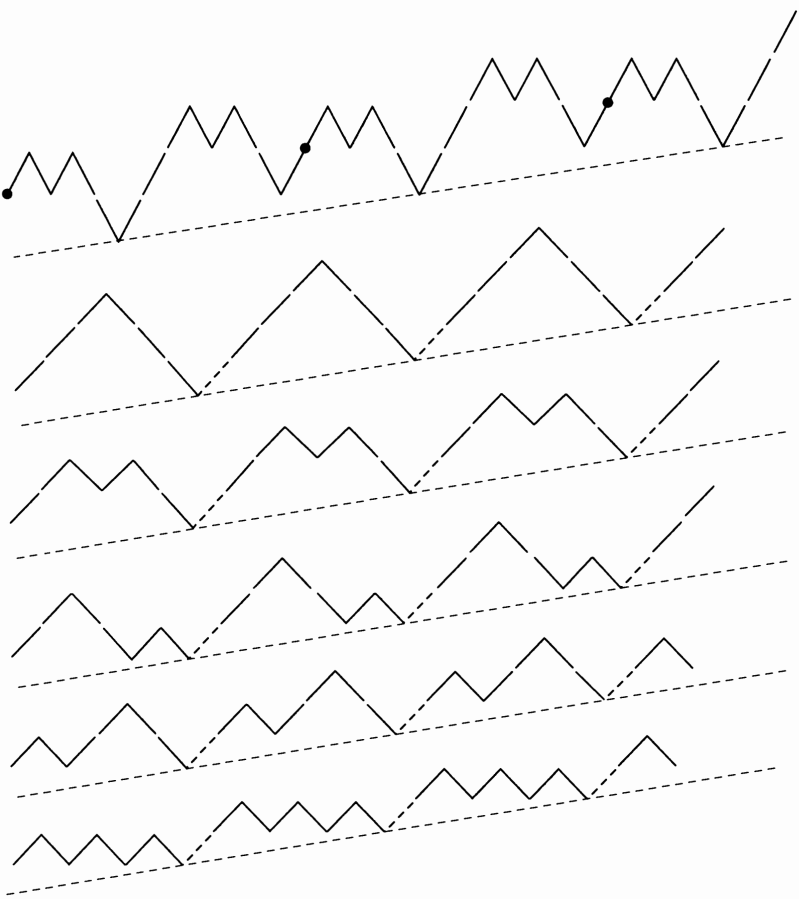


FIGURE 4.21 *Mountain ranges with 4 upstrokes and 3 downstrokes.*

different cyclic patterns of $n + 1$ upstrokes and n downstrokes. This is one of the three formulas we gave for c_n, the nth Catalan number.

We're sure you've now seen enough of the Catalan numbers, so we'll turn to a new topic.

FAULHABER'S FORMULA

We've already seen the formulas for the sums of numbers up to n, and of their squares and cubes, in Chapter 2:

$$1^0 + 2^0 + \cdots + n^0 = n,$$

$$1^1 + 2^1 + \cdots + n^1 = \frac{1}{2}[n^2 + n],$$

$$1^2 + 2^2 + \cdots + n^2 = \frac{1}{3}\left[n^3 + \frac{3}{2}n^2 + \frac{1}{2}n\right],$$

$$1^3 + 2^3 + \cdots + n^3 = \frac{1}{4}[n^4 + 2n^3 + n^2].$$

Johann Faulhaber, known in his day as The Great Arithmetician of Ulm but now almost forgotten, worked out the formula for sums of higher powers in his *Academiae Algebrae* (1631):

$$1^{k-1} + 2^{k-1} + \cdots + n^{k-1} =$$

$$\frac{1}{k}\left[n^k + \binom{k}{1}n^{k-1}\times\frac{1}{2} + \binom{k}{2}n^{k-2}\times\frac{1}{6} + \binom{k}{3}n^{k-3}\times 0 + \binom{k}{4}n^{k-4}\times\frac{-1}{30} + \cdots\right].$$

Observe that the expression in brackets is exactly like the formula for the binomial theorem, except that there is no constant term and the other terms are multiplied by certain constants:

$$1 \quad \frac{1}{2} \quad \frac{1}{6} \quad 0 \quad \frac{-1}{30} \quad 0 \quad \frac{1}{42} \quad 0 \quad \frac{-1}{30} \quad 0 \quad \frac{5}{66} \quad 0 \quad \ldots.$$

BERNOULLI NUMBERS

These constants in Faulhaber's formula are known as the **Bernoulli numbers**, since they are intensively discussed in Jacob (James) Bernoulli's posthumous masterpiece, the *Ars Conjectandi* (1713), even though Bernoulli gives full credit to Faulhaber.

To remind us of the connection with the binomial theorem, we'll use the names

$$B^0 = 1,\ B^1 = \frac{1}{2},\ B^2 = \frac{1}{6},\ B^3 = B^5 = B^7 = \cdots = 0,$$

$$B^4 = B^8 = \frac{-1}{30},\ B^6 = \frac{1}{42},\ B^{10} = \frac{5}{66},\ \ldots,$$

just as if the Bernoulli numbers were powers (which, of course, they aren't).

Faulhaber's formula can be written formally as

$$1^{k-1} + 2^{k-1} + \cdots + n^{k-1} = \frac{\text{``}(n+B)^k - B^k\text{''}}{k}.$$

Expressions inside quotation marks should be written as a sum of terms, each of which is a power of B times some number, and the powers of B should then be interpreted as Bernoulli numbers.

Jacob Bernoulli boasted that he found the sum of the tenth powers of the first thousand integers *intra semiquadrantem horae* (in $7\frac{1}{2}$ minutes)! Now that you know Faulhaber's formula, you can check this in even less time:

$$\frac{\text{``}(x+B)^{11} - B^{11}\text{''}}{11} =$$

$$\frac{1}{11}(x^{11}+11B^1x^{10}+55B^2x^9+330B^4x^7+462B^6x^5+165B^8x^3+11B^{10}x)$$

with $x = 1000$ and

$$B^1 = \frac{1}{2},\ B^2 = \frac{1}{6},\ B^4 = -\frac{1}{30},\ B^6 = \frac{1}{42},\ B^8 = -\frac{1}{30},\ B^{10} = \frac{5}{66},$$

you have

$$
\begin{aligned}
x^{11}/11 &= 90\,909\,090\,909\,090\,909\,090\,909\,090\,909\,090 \cdot 909090\ldots \\
B^1x^6 &= 500\,000\,000\,000\,000\,000\,000\,000\,000\,000 \\
5B^2x^9 &= 833\,333\,333\,333\,333\,333\,333\,333\,333 \cdot 333333\ldots \\
30B^4x^7 &= -1\,000\,000\,000\,000\,000\,000\,000 \\
42B^6x^5 &= 1\,000\,000\,000\,000\,000 \\
15B^8x^3 &= -500\,000\,000 \\
B^{10}x &= 75 \cdot 757575\ldots \\
\hline
\text{total} &= 91\,409\,924\,241\,424\,243\,424\,241\,924\,242\,500 \cdot 0
\end{aligned}
$$

How do you find the Bernoulli numbers? You can define and compute them by the equations

$$B^2 - 2B^1 + 1 = B^2, \quad \text{whence} \quad B^1 = \frac{1}{2},$$

$$B^3 - 3B^2 + 3B^1 - 1 = B^3, \quad \text{whence} \quad B^2 = \frac{1}{6},$$

$$B^4 - 4B^3 + 6B^2 - 4B^1 + 1 = B^4, \quad \text{whence} \quad B^3 = 0,$$

$$B^5 - 5B^4 + 10B^3 - 10B^2 + 5B^1 - 1 = B^5, \quad \text{whence} \quad B^4 = \frac{-1}{30},$$

and, in general,

"$(B - 1)^k = B^k$" (for $k \neq 1$) whence B^{k-1} can be computed if you already know $B^1, B^2, \cdots, B^{k-2}$. On the other hand $(B - 1)^1$ and B^1 are *not* equal. In fact

$$(B - 1)^1 = B^1 - 1 = -\frac{1}{2}, \text{ whereas } B^1 = +\frac{1}{2},$$

so, in the expressions of "$(100 + B)^k$" and of "$(99 + B)^k$", all the terms are equal except for the bold ones:

$$\text{“}(100 + B)^k\text{”} = 100^k + \mathbf{kB^1100^{k-1}} + \binom{k}{2}B^2100^{k-2} + \cdots,$$

$$\text{“}(99 + B)^k\text{”} = \text{“}(100 + B - 1)^k\text{”} =$$

$$100^k + \mathbf{k(B-1)^1100^{k-1}} + \binom{k}{2}(B - 1)^2100^{k-2} + \cdots.$$

Subtracting them, we get

$$\text{“}(100 + B)^k - (99 + B)^k\text{”} = k \cdot 100^{k-1}.$$

Now we add a hundred such equations and cancel lots of terms:

$$\begin{aligned}
\text{“}(100 + B)^k - (99 + B)^k\text{”} &= k \cdot 100^{k-1}\\
\text{“}(99 + B)^k - (98 + B)^k\text{”} &= k \cdot 99^{k-1}\\
&\cdots\cdots\cdots\cdots\\
\text{“}(2 + B)^k - (1 + B)^k\text{”} &= k \cdot 2^{k-1}\\
\text{“}(1 + B)^k - B^k\text{”} &= k \cdot 1^{k-1}\\
\hline
\text{“}(100 + B)^k - B^k\text{”} &= k(1^{k-1} + 2^{k-1} + \cdots + 99^{k-1} + 100^{k-1}),
\end{aligned}$$

which is Faulhaber's formula.

Bernoulli numbers arise in a wide variety of analytical and combinatorial contexts, all over mathematics. As you can see, they aren't whole numbers. A surprising discovery of von Staudt and Clausen tells us that in fact if $2, 3, \ldots, p$ are all the prime numbers that are 1 more than a divisor of $2n$, then

$$B^{2n} = N - \frac{1}{2} - \frac{1}{3} - \cdots - \frac{1}{p}$$

for some whole number N, which is 1 for $2n \leq 12$. For instance, $B^{12} = -691/2730$ and

$$B^{12} = 1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{5} - \frac{1}{7} - \frac{1}{13} = \frac{-691}{2730} = -.2\dot{5}3113\dot{5}$$

Together with Fermat's Little Theorem this implies that the period of the decimal for B^{2n} has length dividing $2n$, and starts one digit later than the decimal point.

Note that $B^3, B^5, B^7, \ldots$ are all zero, and, although B^2, B^4, B^6 are small, ultimately B^{2n} becomes very large: In fact, it equals

$$\frac{2(2n)!}{(2\pi)^{2n}}\left(1 + \frac{1}{2^{2n}} + \frac{1}{3^{2n}} + \cdots\right).$$

Analytically, the coefficient of x^{2n-1} in the expansion of $\cot(x/2)$ is $\pm B^{2n}/(2n)!$.

EULER NUMBERS AND ZIGZAG ARRANGEMENTS

What is the number, Z_n, of zigzag arrangements of 1, 2, . . . , n, that is, arrangements in which the numbers alternatively rise, then fall? Table 4.4 shows the first few.

n	Z_n	zigzag arrangements
1	1	1
2	1	none; 12
3	2	231; 132; none
4	5	none; 3412; 1423, 2413; 1324, 2314
5	16	24351, 25341, 34251, 35241, 45231; 14352, 15342, 34152, 35142, 45132; 14253, 15243, 24153, 25143; 13254, 23154; none
6	61	
7	272	

TABLE 4.4 *Euler numbers and zigzag arrangements.*

The semicolons separate the arrangements according to their last number, r. The number of **zigzag arrangements** of 1, 2, . . . , n with last number r is the rth entry in the nth row of the **zigzag triangle**

$$
\begin{array}{c}
1 \\
0 \rightarrow 1 \\
1 \leftarrow 1 \leftarrow 0 \\
0 \rightarrow 1 \rightarrow 2 \rightarrow 2 \\
5 \leftarrow 5 \leftarrow 4 \leftarrow 2 \leftarrow 0 \\
0 \rightarrow 5 \rightarrow 10 \rightarrow 14 \rightarrow 16 \rightarrow 16 \\
61 \leftarrow 61 \leftarrow 56 \leftarrow 46 \leftarrow 32 \leftarrow 16 \leftarrow 0 \\
0 \rightarrow 61 \rightarrow 122 \rightarrow 178 \rightarrow 224 \rightarrow 256 \rightarrow 272 \rightarrow 272
\end{array}
$$

which is computed in the 'boustrophedon' manner hinted at by the arrows. Each row after the first is found by adding the numbers from the previous row, alternately from left to right and right to left.

The left border contains the **zig numbers**, Z_{2n} which are traditionally called the **Euler numbers**; they are also known as **secant numbers** in view of the formula

$$\sec x = 1 + \frac{1 \cdot x^2}{2!} + \frac{5 \cdot x^4}{4!} + \frac{61 \cdot x^6}{6!} + \frac{1385 \cdot x^8}{8!} + \cdots$$

and the right border contains the **zag** or **tangent numbers**, Z_{2n+1}, since

$$\tan x = \frac{x}{1!} + \frac{2 \cdot x^3}{3!} + \frac{16 \cdot x^5}{5!} + \frac{272 \cdot x^7}{7!} + \frac{7936 \cdot x^9}{9!} + \cdots$$

the total number of arrangements of 1, 2, . . . , n in which there are just $k - 1$ rises has also been named after Euler, but to distinguish it, we call it the **Eulerian number**, $A(n, k)$:

$$A(n, k) = \sum_{j=0}^{k} (-1)^j \binom{n+1}{j} (k-j)^n.$$

FIBONACCI NUMBERS

Leonardo of Pisa (ca. 1200) wondered how many pairs of rabbits would be produced in the nth generation, starting from a single pair and supposing that any pair of rabbits of one generation produces one pair of rabbits for the next generation and one for the generation after that, and then they die.

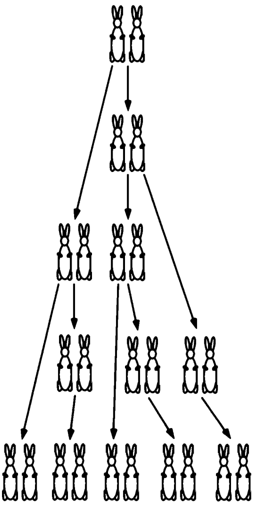


FIGURE 4.22 *A pair of rabbits and their progeny.*

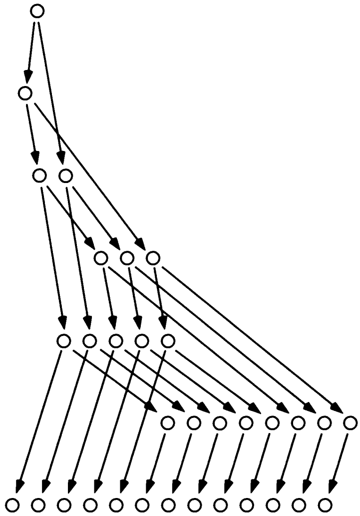


FIGURE 4.23 *Forming the Fibonacci numbers.*

If there are f_n pairs of rabbits in the nth generation, then

$$f_1 = 1 \text{ (the original pair)},$$

$$f_2 = 1 \text{ (their immediate progeny)},$$

$$f_{n+2} = f_n + f_{n+1},$$

since we get a pair in generation $n + 2$ for each pair in generation n or generation $n + 1$ (see Figures 4.22 and 4.23).

$$f_0 = 0\ 1\ 1\ 2\ 3\ 5\ 8\ 13\ 21\ 34\ 55\ 89\ 144\ 233\ 377\ 610\ \ldots$$

are called **Fibonacci numbers**, since Leonardo's father was nicknamed Bonacci, and so Leonardo was Fibonacci (*filius Bonacci* = "son of the good-natured one"). Fibonacci numbers arise in so many ways, it's almost unbelievable: their manifestations seem as numerous as Leonardo's rabbits. There is even a mathematical periodical, the *Fibonacci Quarterly*, devoted entirely to the subject. We'll only mention a few of the more striking properties.

The **Lucas numbers**, l_n,

$$l_0 = 2\ 1\ 3\ 4\ 7\ 11\ 18\ 29\ 47\ 76\ 123\ 199\ 322\ 521\ 843\ 1364\ \ldots$$

(defined by the same rule, but with a different start) are related to the Fibonacci numbers in many ways.

$$\begin{aligned} f_{2n} &= f_n l_n & l_{2n} &= l_n^2 - 2(-1)^n \\ f_0 + f_1 + \cdots + f_n &= f_{n+2} - 1 & l_0 + l_1 + \cdots + l_n &= l_{n+2} - 1 \\ l_n &= f_{n-1} + f_{n+1} & 5f_n &= l_{n-1} + l_{n+1} \\ 2f_{m+n} &= f_m l_n + f_n l_m & 2l_{m+n} &= l_m l_n + 5 f_m f_n \end{aligned}$$

$$f_{n+1} = \binom{n}{0} + \binom{n-1}{1} + \binom{n-2}{2} + \cdots$$

This last relation, noticed by Lucas, shows that you can read the Fibonacci numbers from Pascal's triangle (Figure 4.24).

Kepler pointed out that the ratios of consecutive Fibonacci numbers approach 1.618... . The exact limit is the **golden number**,

$$\tau = \frac{1 + \sqrt{5}}{2} = 1.6180339887498948482\ \ldots$$

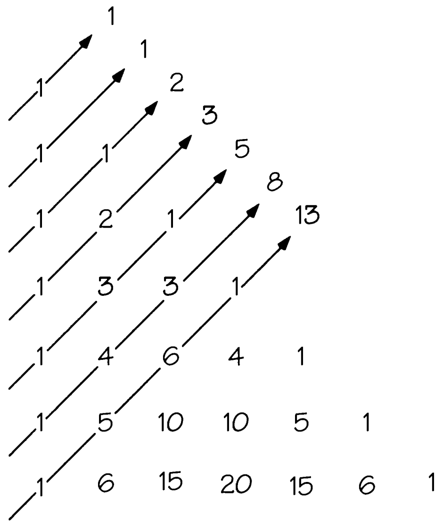


FIGURE 4.24 *Summing diagonals in Pascal's triangle gives Fibonacci numbers.*

which was known to the Greeks. The ratios of consecutive Lucas numbers tend to the same limit, while the ratio, l_n/f_n, between corresponding Lucas and Fibonacci numbers, tends to $\sqrt{5}$. In fact, there are formulas for f_n and l_n in terms of the golden number, which we'll give in Chapter 7.

Leonardo's rabbit problem was, of course, not too true to life. However, the Fibonacci numbers really do occur in nature! You can read them off pineapples or cacti, off pinecones and sunflowers: they also control the arrangements of leaves of almost all plants.

PHYLLOTAXIS

The botanical name for leaf arrangement is *phyllotaxis*. Look at the florets in the sunflower's head in Figure 4.25. They appear to form two systems of spirals, radiating from the center. Although it looks symmetrical, the numbers of clockwise and counterclockwise spirals are in fact not equal. If you count them carefully, you'll find 55 clockwise spirals and 34 counterclockwise ones.

FIGURE 4.25 *Sunflower. (Courtesy of D.R. Fowler and P. Prusinkiewicz.)*

The pineapples in Figure 4.26 and the daisy in Figure 4.27 exhibit a similar phenomenon. In Figure 4.28 which is a tracing of Figure 4.27, we emphasize the 21 such spirals going one way and 34 going the other way by numbering the petals. You can see such spirals on many other plants, such as cauliflowers (Figure 4.29), pinecones and certain kinds of cactus. There are usually two systems of florets, seeds, twigs, petals, or whatever,[1] going in opposite directions, and the numbers of spirals in these systems are consecutive Fibonacci numbers. Why is this so?

We'll describe what happens at an earlier stage in the plant's life. We regard the tip of a plant as a cone and consider the initial arrival of "buds" on this cone. The cone may be very flat, as in the sunflower (Figure 4.30(a)), or pointed, as in the stem (Figure 4.30(b)), or somewhere in between, as in the pineapple of Figure 4.30(c).

[1] "Everything is a leaf"—Goethe.

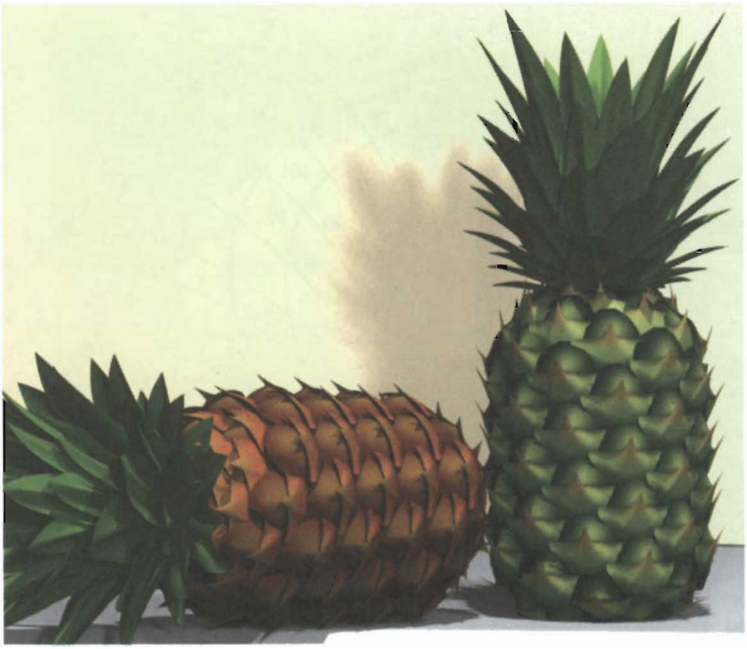


FIGURE 4.26 *Spirals on pineapples. (Photo courtesy of D.R. Fowler and A. Snider.)*

FIGURE 4.27 *Daisy capitulum. (Computer generated by Deborah R. Fowler on an IRIS workstation using an L-grammar algorithm.)*

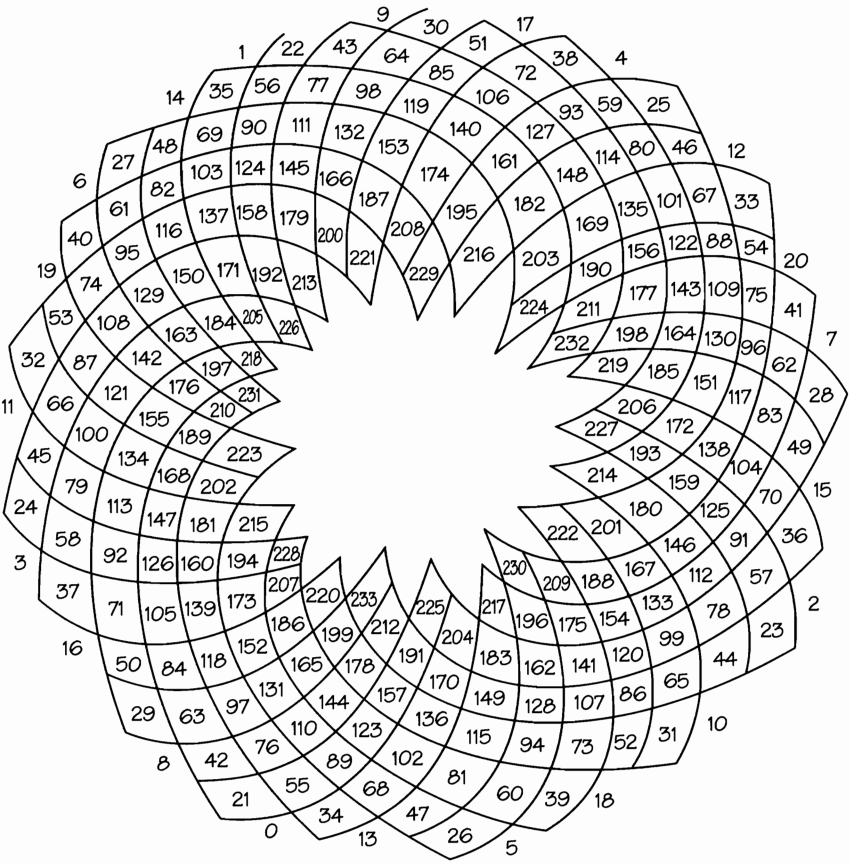


FIGURE 4.28 *The florets of Figure 4.25 numbered to show arithmetic progressions with Fibonacci differences 21 and 34.*

These buds may become seeds in a seed case, twigs on a branch, petals of a flower, and so forth. However, at this stage of its life, one bud inhibits or repels others. This may be because they're crowded and physically push each other apart, or maybe they are competing for essential nutrients, or perhaps they are deliberately secreting some inhibiting substance.

We regard the growth of our plant as taking place at the tip of the cone. Since the tip is continually advancing by new growth, a given portion of the plant moves steadily downward and outward relative to the tip (Figure 4.31(a)).

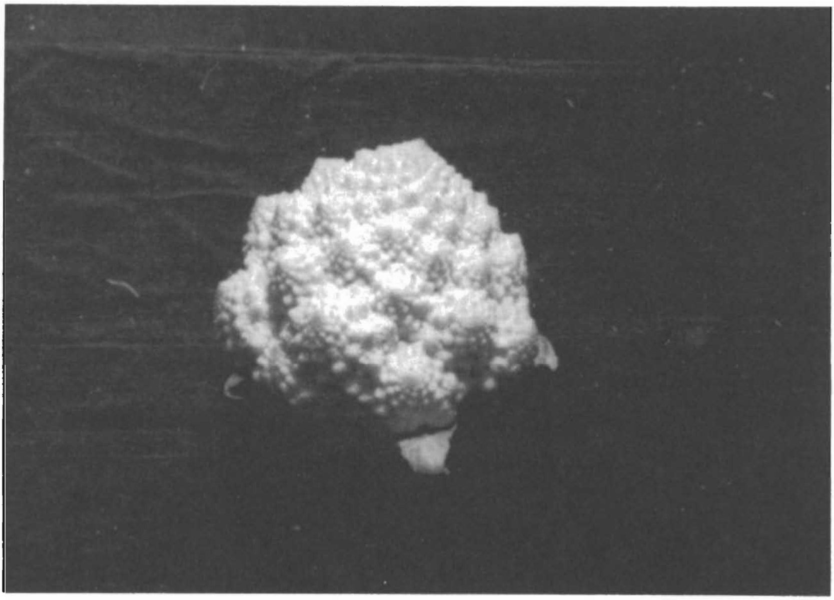


FIGURE 4.29 *The inflorescence of a cauliflower. (Photo courtesy of E. Thirian.)*

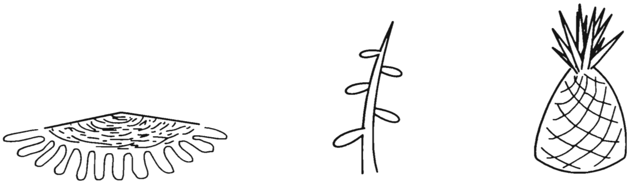


FIGURE 4.30 *The "cones" of various plants.*

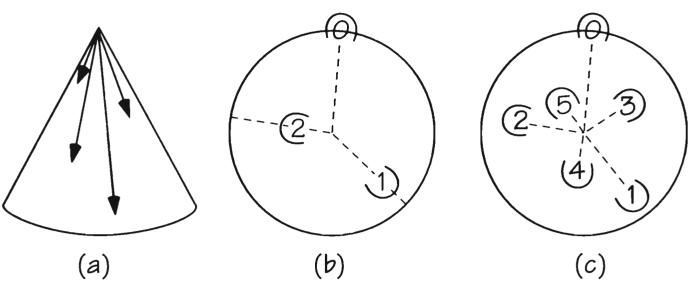


FIGURE 4.31 *Buds distributing themselves on a conical plant.*

SAY, BUD, WHERE DO YOU THINK YOU ARE GOING?

We'll number the buds 0, 1, 2, . . . , in order of their appearance. In the plan view of Figure 4.31(b), we've put bud number 0 at the 12 o'clock position, and since it was born first, it has already arrived at the perimeter. Buds 0 and 1 separate the cone into larger and smaller sectors. Bud 2 finds it easier to exist in the larger sector, forcing 3 into the smaller one. Whereabouts in these sectors will they be? Since number 1 is newer and nearer the tip than number 0, it is likely to be more inhibiting, so numbers 2 and 3 will be slightly nearer to 0 than they are to 1. We've put them at about 9 o'clock and 2 o'clock in Figures 4.31(b) and (c). At this stage we have four sectors, the largest of which is between buds 1 and 2. We expect number 4 to be born in this sector, slightly nearer to 1, since 2 is more recent.

This process might seem rather rough, but it is in fact extremely stable. Suppose, for instance, that bud number 1 was born more or less exactly opposite number 0, leaving two equal sectors open for number 2. Number 2 would randomly choose one of these, and, as it grew, it would push away from 0 and 1 (but slightly more from 1) and so enlarge its own sector. Similar phenomena happen at later stages in the process. The reader should also understand that the action takes place at a very early stage in the development, so that it is likely to be in an otherwise uniform environment, maybe only a few millimeters across.

Figure 4.32 shows what happens when many buds have developed. This time we've taken an idealized plant in which the cone is nearly cylindrical, and we've unrolled the cylinder. Of course, a real plant won't have the buds in exactly the "right" places, but neither can they be too far off. The dotted ellipses in Figure 4.32 show how the spheres of influence of buds 17, 20, and 22 leave a ready-made hole for bud number 25. It won't be born too far from the center of this hole, and even if it's a little out of position, as it grows it will jostle itself into the right position. Once the process is established, it's really very hard for it to go wrong.

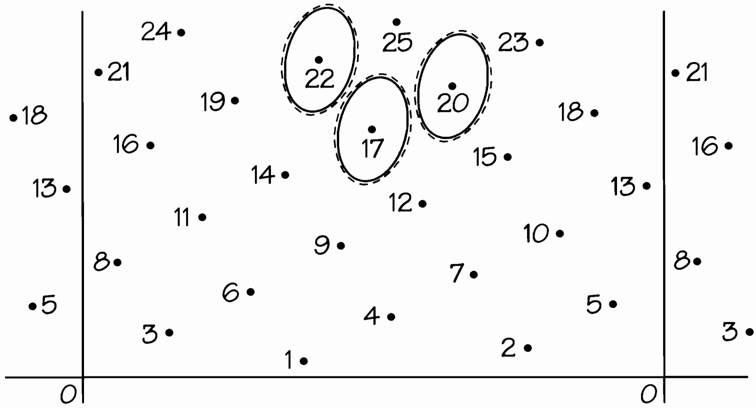


FIGURE 4.32 *The first twenty-five buds.*

In the perfect version of the process, each new bud advances by the same angle, the angle being that which divides a complete turn in a golden section. In our figures this is 0.618 of a counterclockwise revolution, or 0.382 of a clockwise one.

How many spirals are there? These spirals are very much in the eye of the beholder. You tend to link up neighboring buds into a chain. Thus, in Figure 4.32, you probably mentally link the buds into lines with difference 3 that slope up to the right, and lines of difference 5 that slant up to the left.

On the other hand, if you squint up the page from the bottom, you may find it easier to organize them into lines of difference 8, or even of 13. The successive Fibonacci numbers arise as those multiples of the basic angle that are nearest to whole numbers of revolutions. The particular Fibonacci numbers you notice depend on how squashed the vertical scale is, compared with the horizontal. Figures 4.33 and 4.34 are just Figures 4.32 with the vertical axis more and more squashed, and cut into domains by assigning each point to its nearest bud.

Although it looks very much like a plant, Figure 4.35 was actually produced just by applying this rule from a simple mathematical formula that squashes the vertical axis by a gradually varying amount.

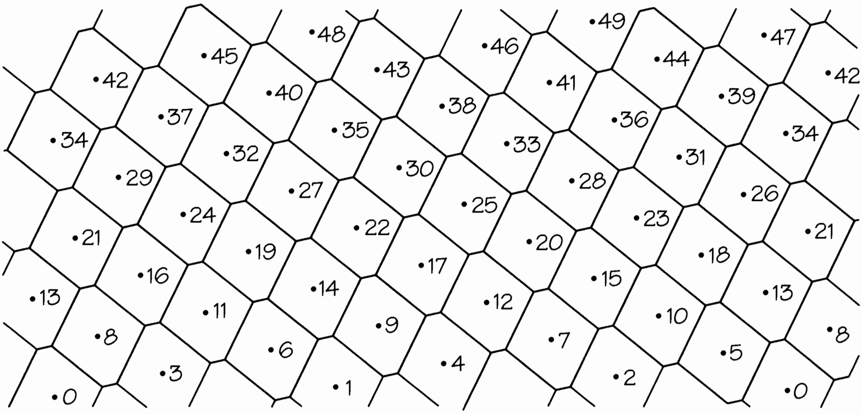


FIGURE 4.33 *Here region n touches n ± 5, n ± 8, and n ± 13.*

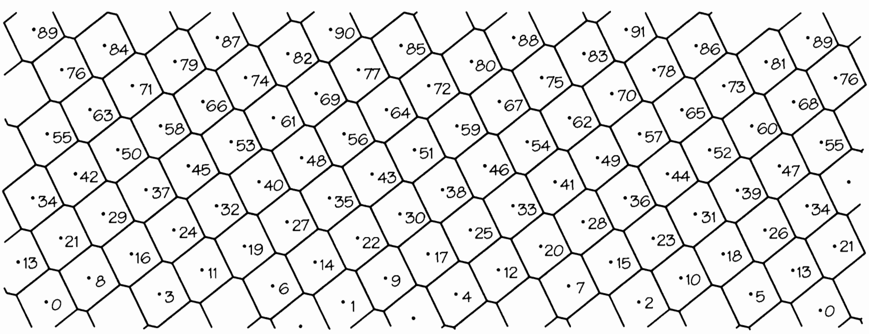


FIGURE 4.34 *Here n touches n ± 8, n ± 13, and n ± 21.*

In Figure 4.32, the numbers 3, 5, and 8 are most obvious; in Figure 4.33, with the scale multiplied by a half, 5 and 8 are most noticeable; in Figure 4.34, with scale only one-fifth, 8, 13, (and 21) are most obvious. At the foot of Figure 4.35, the numbers 3, 5, and 8 are prominent; at the top, 13 and 21 begin to predominate.

Before you next eat a pineapple, try to find the correct numbering

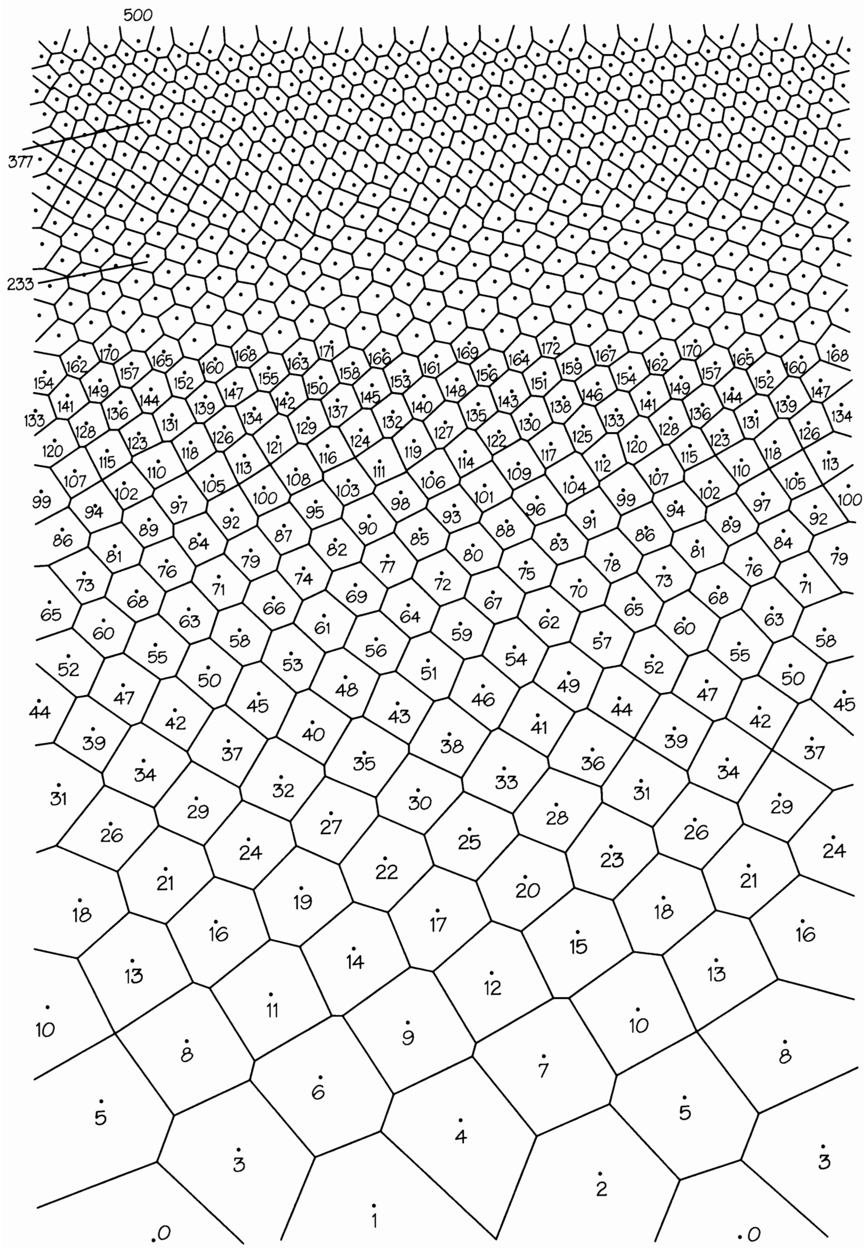


FIGURE 4.35 *Variable growth rate accentuates different Fibonacci numbers. (Diagram courtesy of Len Bos.)*

of the buds on it. The pattern is usually easiest to follow about halfway up, but with a little care you can work backward to the base and even identify number 0.

The leaves on the stems of plants exhibit remnants of this process. The embryo leaf buds in such cases were originally stacked quite tightly around the stem, as in our figures. In a later phase of growth the stem elongates so that all that is left of the original arrangement is a tendency for each leaf to spiral about 0.382 (or roughly 2/5) of a revolution from its predecessor.

The petals of a rose exhibit the same phenomenon in the way they overlap their predecessors, but they are usually so tightly packed that the structure is hard to see without dissecting it. Figure 4.36 shows a flower on which we've numbered the petals individually.

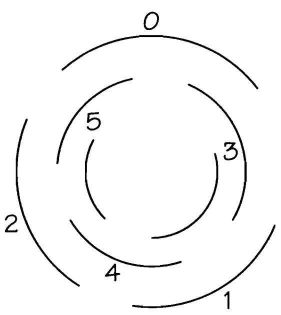


FIGURE 4.36 *The arrangement of the petals of a flower.*

We may even speculate that the exact number (when it is small) of petals on a flower might be determined by such a mechanism. The following table shows in fractions of a revolution the angle defining the position of petal number n (this is just the fractional part of $n\tau$) and also the smallest angle between this petal and any of its predecessors. If the latter were the only relevant parameter, it would estimate the likelihood of the emergence of petal number n.

Petal number	0	1	2	3	4	5	6	7	8	
Position angle	0	.618	.236	.854	.472	.090	.708	.326	.944	...
Smallest angle		.382	.236	.146	.146	.090	.090	.090	.056	...

The position angles are just the fractional parts of successive multiples

$$0,\ 1.618\ldots,\ 3.236\ldots,\ 4.854\ldots$$

of the **golden number**

$$\tau = \frac{1+\sqrt{5}}{2} = 1.6180339887494\ldots .$$

On this view you will see that the fifth petal (number 4) finds things just as easily (0.146) as its predecessor, while the sixth and later ones have a substantially harder time (0.090). Without being able to count, but merely by setting a given level of inhibition, a flower can arrange for the total number of petals to be a small Fibonacci number, but not any other number!

It is not so easy to control things in order to get a larger Fibonacci number exactly, since the influence of petal number 0 becomes rapidly less noticeable. In fact, when a flower has a large number of petals, the exact number of petals usually depends on the particular specimen.

The reader should beware of trying to force this theory too far, because plants use many other mechanisms in their development. Some emit pairs of leaves simultaneously on opposite sides of the stem. Successive pairs may then rotate by 0.191 (half of 0.382) of a revolution, but they are just as likely to rotate by a right angle. Many flowers have exactly six petals. For these it is usually the case that the petals are organized as two generations of three petals each. This effect shows very clearly on a narcissus; actually, the first three of its "petals" are *sepals* rather than petals.

One occasionally sees sport pinecones in which the numbers of spirals are the doubles of Fibonacci numbers. Presumably this happened because buds 0 and 1 emerge almost simultaneously on opposite sides and the subsequent process was bisected.

The next time you eat a head of cauliflower, notice that not only does it have Fibonacci numbers of spirals of florets, but these in turn often have spirals of subflorets. And on your walk through the forest,

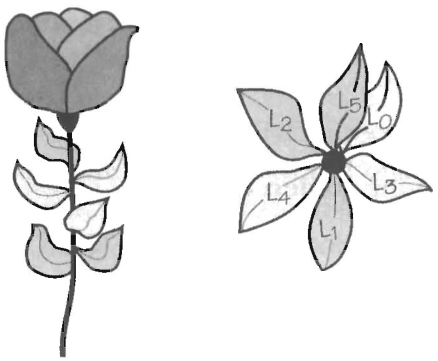


FIGURE 4.37 *Leaves round a stem.*

look down the stem of a plant. If two leaves appear to be above one another (Figure 4.37), they are probably a Fibonacci number apart.

The arrangement of the larger branches on a tree is often more random, but there are some species of mountain trees in which the Fibonacci organization can be seen throughout the entire structure, even in the roots, if you dig down to them.

REFERENCES

V. I. Arnold. Bernoulli-Euler updown numbers associated with function singularities, their combinatorics and arithmetics, *Duke Math. J.* 63(1991): 537-555.

V. I. Arnold. Snake calculus and combinatorics of Bernoulli, Euler and Springer numbers for Coxeter groups. *Russian Math. Surveys* 47(1992): 3-45.

E. R. Berlekamp, J. H. Conway, and R. K. Guy. *Winning Ways for your Mathematical Plays,* Academic Press, 1982, 475-478, 501.

N. G. de Bruijn & B. J. M. Morselt. A note on plane trees. *J. Combin. Theory* 2(1967): 27-34.

B. L. Burrows & R. F. Talbot. Sums of powers of integers. *Amer. Math. Monthly* **91**(1984): 394-403.

J. H. Conway & H. S. M. Coxeter. Triangulated polygons and frieze patterns. *Math. Gaz.* **57**(1973): 87-94, 175-183; *MR* **57** # 1254-5.

A. W. F. Edwards. A quick route to sums of powers. *Amer. Math. Monthly* **93**(1986): 451-455.

Antal E. Fekete. **Apropos** two notes on notation. *Amer. Math. Monthly* **101**(1994): 771-778.

Martin Gardner. *Time Travel and other Mathematical Bewilderments*, W. H. Freeman, New York, 1988, chapter 20, Catalan Numbers.

Johann Wolfgang von Goethe. *Versuch die Metamorphosa der Planzen zu erklären*, Gotha, 1790.

H. W. Gould. *Bell & Catalan Numbers: Research Bibliography of Two Special Number Sequences*, Sixth Edition, Morgantown, WV, 1985.

Ronald L. Graham, Donald E. Knuth & Oren Patashnik. *Concrete Mathematics, A Foundation for Computer Science*, Addison-Wesley, Reading, MA, 1989.

J. Karamata, Théorèmes sur la sommabilité exponentielle et d'autres sommabilités s'y rattachant. *Mathematica* (cluj), **9**(1935): 164–178.

A. A. Kirillow. Variations on the triangular theme in S. G. Gindikin & E. B. Vinberg (editors). *E. B. Dynkin Seminar on Lie Groups. Amer. Math. Soc.*, 1995.

Donald E. Knuth. Two notes on notation. *Amer. Math. Monthly* **99**(1992): 403-422.

D. E. Knuth & T. J. Buckholtz. Computation of tangent, Euler and Bernoulli numbers. *Math. Comput.* **21**(1967): 663-688.

Mark Krusemeyer. A parenthetical note (to a paper of Richard Guy). *Math. Mag.* **69**(1996).

Michael J. Kuchinski. *Catalan Structures and Correspondences.* M.Sc Thesis, West Virginia University, 1977.

Luo Jian-Jin. An Tu Ming, the first discoverer of the Catalan numbers. *Neimenggu Daxue Xuebo* **19**(1988): 239–245; *MR* **90e**:01007.

J. Millar, N. J. A. Sloane & N. E. Young, A new operation on sequences: The boustrophedon transform, 95-09-25 preprint.

Th. Motzkin. Relations between hypersurface cross-ratios, and a combinatorial formula for partitions of a polygon, for permanent preponderance, and for non-associative products. *Bull. Amer. Math. Soc.* **54**(1948): 352–360; *MR* **9**: 489–490.

C. D. Oakley & R. J. Wisner. Flexagons. *Amer. Math. Monthly* **64**(1957): 143–154; *MR* **19**: 241b.

Przemyslaw Prusinkiewicz & Aristid Lindenmayer. *The Algorithmic Beauty of Plants*, with James S. Hanan, F. David Fracchia, Deborah R. Fowler, Martin J. M. de Boer & Lynn Mercer, Springer-Verlag, 1990.

H. Rademacher. On the partition function $p(n)$. *Proc. London Math. Soc.* (2), **43**(1937): 241–254.

Ibrahim A. Salama & Lawrence L. Kupper. A geometric interpretation for the Eulerian numbers. *Amer. Math. Monthly* **93**(1986): 51–52.

J. O. Shallit, E2972. Another appearance of the Catalan numbers. *Amer.*

Math. Monthly **89**(1982): 698; solution David M. Wells, **93**(1986): 217-218.

Ian Stewart. Mathematical Recreations: Daisy, Daisy, give me your answer, do. *Sci. Amer.* **272**#1(Jan 1995): 96-99.

Peter Guthrie Tate. *Proc. Royal Soc. Edinburgh* **VII**(1870-71): 391-394.

D'Arcy Wentworth Thompson. *On Growth and Form*, Cambridge Univ. Press, 1952.

C. W. Wardlaw. *Organization and Evolution in Plants.* Longmans Green, 1965.

5

The Primacy of Primes

Although arithmeticians have studied prime numbers for thousands of years, there are even more open problems today than there have ever been before. Most of the positive integers can be expressed as the product of smaller ones; such products are called **composite numbers**.

$$4 = 2 \times 2,\ 6 = 2 \times 3,\ 8 = 2 \times 4,\ 9 = 3 \times 3,\ 10 = 2 \times 5,\ 12 = 3 \times 4$$

are examples of composite numbers.

The number 1 is in a class all by itself and is called the **unit**. The remaining numbers that are bigger than 1, but not the product of smaller numbers,

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47,$$
$$53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, \ldots,$$

are called **prime numbers**. Perhaps the greatest mystery about prime numbers is that although they are quite simply defined they behave quite irregularly.

Eratosthenes (276–194 B.C.), librarian of the great library in Alexandria, was one of the most brilliant men of the ancient world.

Perhaps his most noteworthy achievement was the measurement of the radius of the Earth at a time when few people believed that it was round, by comparing the lengths of the shadows of flagpoles at noon in Alexandria and in Syene (Aswan). But number theorists will always remember him for his wonderful prime number *sieve*. Just as the farmer winnows the valuable wheat from the useless chaff, so Eratosthenes used his sieve to separate the precious primes from their common composite companions. Here's how it works.

Write down the numbers in order, putting 1 in a box to show that it's the unit

$\boxed{1}$ 2 3 4 5 6 7 8 9 10 11 12 13 . . . .

Circle the first remaining number, which is 2, and strike out every second number thereafter:

$\boxed{1}$ ② 3 ~~4~~ 5 ~~6~~ 7 ~~8~~ 9 ~~10~~ 11 ~~12~~ 13 . . . .

Circle the next remaining number, namely 3, and strike out all subsequent multiples of that number

$\boxed{1}$ ② ③ ~~4~~ 5 ~~6~~ 7 ~~8~~ ~~9~~ ~~10~~ 11 ~~12~~ 13 . . . .

If you continue in this way at each stage, circling the first remaining number and striking out its higher multiples, the numbers you circle will be the prime numbers (Figure 5.1).

Lots of numbers get struck out more than once. For instance, we struck out 6 as a multiple of 3, even though it had already been struck out as a multiple of 2. In fact, when we're coping with a prime number p, its multiples by numbers smaller than p will already have been dealt with, and the first one that hasn't been will be

$$p \text{ times } p = p^2$$

When we dealt with 2 and 3, leaving 5 as the next prime, the remaining numbers,

5, 7, 11, 13, 17, 19, 23

below $5^2 = 25$ were therefore already known to be prime.

The sieving process is made easier if you write the numbers in a tabular array, with rows of some fixed length (compare Figures 2.1, 2.4, 2.5, 2.6, 2.7 in Chapter 2), for then the multiples of any fixed number will form an orderly pattern that helps to get them right.

In Figure 5.1 we have drawn just the odd numbers below $361 = 19^2$, and the various straight lines strike out the multiples of 3, 5, 7, 11, 13, and 17. So the remaining numbers are 1 and the odd primes below 360.

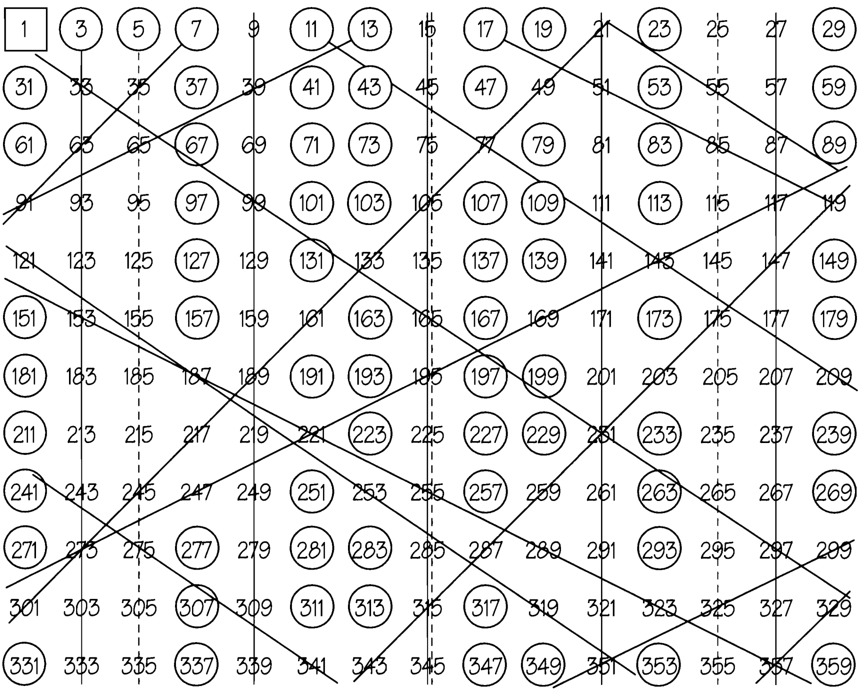


FIGURE 5.1 *A sieve of Eratosthenes.*

As a result of a collaborative effort during the nineteenth century, tables were published giving the prime numbers and the factors of other numbers for each successive million, up to ten million. This work was crowned in 1909 by Derrick Norman Lehmer's ***Factor Table for the First Ten Million***. In 1914 Lehmer also published a list of primes less than ten million, extracted from his factor table.

All these tables were computed using the sieve of Eratosthenes. Indeed, the paper on which Lehmer did his calculation was supported on a very long table, equipped with rollers at each end, and for the smaller primes he made paper stencils with holes through which he could mark their multiples. Some of these were between 15 and 20 feet in length.

It is interesting to note that the Austrian arithmetician J. P. Kulik (1773–1863) spent 20 years of his life preparing a table to 100 million by hand, but this was never published, and the volume for numbers from 12,642,600 to 22,852,800 has been lost.

The number 1 used to be counted as a prime; it appears in D. N. Lehmer's list of primes, for example. But there are so many ways in which it differs from the proper prime numbers that mathematicians now put it in its special class. For instance, if you started the sieving process with 1, you'd just strike out everything else!

ARITHMETIC MODULO P

NUMBERS MOD P

When p is prime, the numbers modulo p form a fruitful field for investigation.

It's clear that modulo any number you can add, subtract, or multiply, but the fascinating thing about primes is that you can also divide. Modulo a prime number, it's sensible to talk about fractions!

FRACTIONS MOD P

If we work modulo 7, then

$$2 \times 4 \equiv 1,\ 3 \times 5 \equiv 1, \text{ and } 6 \times 6 \equiv 1,$$

and so it's quite alright to say that

$$\frac{1}{2} \equiv 4, \quad \frac{1}{3} \equiv 5, \quad \frac{1}{4} \equiv 2, \quad \frac{1}{5} \equiv 3, \quad \frac{1}{6} \equiv 6, \quad \text{mod } 7.$$

Now since $\frac{1}{3}$ is 5, $\frac{2}{3}$ should be 2×5, which is 3 mod 7, and you can check that

$$\frac{2}{3} \equiv 3, \quad \frac{3}{4} \equiv 6, \quad \frac{2}{5} \equiv 6, \quad \frac{3}{5} \equiv 2, \quad \frac{4}{5} \equiv 5, \quad \frac{5}{6} \equiv 2, \quad \text{mod } 7.$$

What about $1/7$? Since 7 is the same as 0 mod 7, this would be dividing by zero, which is illegal!

> Not even modulo 7
> shalt thou divide by zero!

If you are working modulo a number that *isn't* prime, there are *more* things by which you can't divide. For instance, modulo 15, every multiple of 3 is one of 0, 3, 6, 9, or 12, and since 1 does not appear, *there is no number* $\frac{1}{3}$ *mod* 15, even though 3 is not 0 modulo 15.

Here's a fairly easy way to find the numbers $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, . . . , modulo p. Is there a number $\frac{1}{8}$ mod 101? The first multiple of 8 that exceeds 101 is $8 \times 13 = 104 \equiv 3$ mod 101, and so we can certainly reduce the size of the problem:

8 times 13 gives 3.

Now the first multiple of 3 after 101 is $3 \times 34 = 102 \equiv 1$ mod 101, whence

3 times 34 gives 1

and so

8 times 13×34 will give 1.

This tells us the answer:

$$\frac{1}{8} \equiv 13 \times 34 \text{ mod } 101.$$

This is surprisingly easy to work out: we know that three 34s are 1, so twelve 34s are 4 and thirteen 34s must be 38:

$$\frac{1}{8} \text{ is } 38 \text{ mod } 101.$$

In the same way we can find all the numbers $\frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{100}$ mod 101. If you want $\frac{1}{68}$ mod 101, you start with 68 and then repeatedly multiply by the first number that enables you to subtract 101 (and so get a smaller answer):

$$68 \xrightarrow{\times 2} 136 \equiv 35 \xrightarrow{\times 3} 105 \equiv 4 \xrightarrow{\times 26} 104 \equiv 3 \xrightarrow{\times 34} 102 \equiv 1,$$

so $\frac{1}{68}$ is $2 \times 3 \times 26 \times 34$, which is just $2 \times 26 = 52$, since we already know that 3×34 is 1.

$$\frac{1}{68} \equiv 52 \text{ mod } 101.$$[1]

You can divide by any number, other than zero, modulo a prime number,

but for numbers that aren't prime there will be some other divisions that you can't do.

Can a Number Be Split into Prime Factors in More Than One Way?

Euclid's famous book, *The Elements of Geometry* (ca. 300 B.C.), not only deals with geometry but also provides (in Book IX) a theoretical foundation for number theory. His most fundamental discovery about prime numbers is equivalent to the assertion that every number fac-

[1]Modulo a *nonprime* this method sometimes succeeds, but it can also fail by reaching zero. For instance, modulo 102,

$$68 \xrightarrow{\times 2} 136 \equiv 34 \xrightarrow{\times 3} 102 \equiv 0,$$

$$5 \xrightarrow{\times 21} 105 \equiv 3 \xrightarrow{\times 34} 102 \equiv 0.$$

In fact there *isn't* a number $\frac{1}{68}$ mod 102, but there *is* a number $\frac{1}{5}$, namely 41.

tors uniquely into prime numbers. This is not at all obvious, and you might perhaps think that it's not true for 1001, since

$$1001 = 7 \times 143 = 11 \times 91.$$

The explanation's easy. Although 7 and 11 are primes, $143 = 11 \times 13$ and $91 = 7 \times 13$ are not, and the only factorization of 1001 into numbers that are *all* primes is

$$1001 = 7 \times 11 \times 13 = 11 \times 7 \times 13 = \cdots$$

(the order doesn't matter). It's easy to show that this always happens if you know

EUCLID'S PRINCIPLE:

A prime number can't divide a product unless it divides one of the factors.

For if $n = a \times b \times c \times \ldots$ and p *doesn't* divide any of a, b, c, . . . , then, modulo p, there are numbers $1/a$, $1/b$, $1/c$, . . . , and so there is a number $1/n = 1/a \times 1/b \times 1/c \times \ldots$, which shows that n can't be divisible by p.

Euclid's principle is what stops a number from having two really different factorizations. This is because any prime in one factorization must divide *some* prime in the other, and so must actually *be* that prime. We can cancel this prime and repeat the argument. The two factorizations can only differ in the order in which the primes are arranged.

THERE ARE ALWAYS NEW PRIMES!

It's not obvious that the sequence of primes continues indefinitely. There might come a stage, perhaps, when the sieving process stops because all the numbers have been struck out. However, Euclid also proved that the primes do indeed continue forever.

Imagine that all the primes you know are

$$2, 3, 5, 7, 11, 13.$$

Then we'll show that there must be another one. Multiply your primes together and add 1 to get the larger number

$$2\times3\times5\times7\times11\times13 \quad + \quad 1 = 30031.$$

This number is certainly bigger than 1. What is the smallest number, bigger than 1, that divides it exactly? This must be a prime, otherwise one of its factors would be a smaller candidate. But it can't be one of the old primes, 2, 3, 5, 7, 11, or 13, since each of these leaves remainder 1 when divided into 30031. So we've found a new prime.

Sometimes the big number here is already prime, but sometimes, as in the previous example, it isn't:

$$1 + 1 = 2 \text{ is prime}$$

$$2 + 1 = 3 \text{ is prime}$$

$$2\times3 + 1 = 7 \text{ is prime}$$

$$2\times3\times5 + 1 = 31 \text{ is prime}$$

$$2\times3\times5\times7 + 1 = 211 \text{ is prime}$$

$$2\times3\times5\times7\times11 + 1 = 2311 \text{ is prime}$$

but

$$2\times3\times5\times7\times11\times13 + 1 = 30031 = 59\times509$$

and

$$2\times3\times5\times7\times11\times13\times17 + 1 = 510511 = 19\times97\times277,$$

while

$$2\times3\times5\times7\times11\times13\times17\times19 + 1= 9699691 = 347\times27953.$$

The next few prime numbers of the form $2\times3\times5\times \cdots \times p + 1$ are for $p = 31, 379, 1019, 1021$, and 2657.

Although we've known since Euclid that the primes get as large as you like, it was quite a long time before mathematicians could explicitly point to some very big ones.

MERSENNE'S NUMBERS

Before the rise of learned journals, there were several people who made it their concern to communicate any scientific discoveries they became aware of to a large number of correspondents. Father Marin Mersenne (1588–1648) performed this service for the mathematicians of his day. Indeed, much of Fermat's work was first "published" via Mersenne's letters.

In a letter to Frenicle de Bessy, Mersenne discussed the possibilities of prime numbers of the form $2^n - 1$ and made the surprising assertion that $2^n - 1$ was prime for $n = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257$ and for no other n below 257. Mersenne's statement aroused a great deal of interest because the numbers are so large that for the next 200 years nobody was able to confirm or deny it. This interest continues today, and numbers of the form $2^n - 1$ have been named for Mersenne. Did he, or Fermat, perhaps have techniques that were unknown to anyone else?

In 1876, Edouard Lucas managed to prove that $2^{127} - 1$ was indeed prime, and this remained the largest known prime for over 70 years. However, over this period several mistakes were found in Mersenne's statement, and it finally became apparent that it had been no more than an educated guess.

It's easy to see that $2^n - 1$ is definitely *not* prime if n itself is not prime. The binary expansion of $2^{15} - 1$, for instance, is

$$111\ 111\ 111\ 111\ 111,$$

which is obviously divisible by the binary numbers $111 = 2^3 - 1$ and $11111 = 2^5 - 1$ and so can't be prime.

But there are prime numbers p for which $2^p - 1$ is not prime. For instance,

$$2^{11} - 1 = 2047 = 23 \times 89$$

in decimal, or

$$11111111111 = 10111 \times 1011001$$

in binary. There are special techniques that make it comparatively easy to test a number $2^p - 1$ for primality and there's a good chance that, as you are reading this, the largest known prime is a Mersenne number. However, in 1951, Miller and Wheeler held the record with $180(2^{127} - 1)^2 + 1$, and around 1989–1992, J. Brown et al. held it with

$$391581 \times 2^{216193} - 1.$$

Table 5.1 lists the Mersenne primes with $p < 100000$. The primes $p = 110503$, 132049, 216091, 756839, and 859433 also give Mersenne primes.

p	$2^p - 1$	p	(digits)	p	(digits)	p	(digits)
2	3	31	(10)	1279	(386)	9941	(2993)
3	7	61	(19)	2203	(664)	11213	(3376)
5	31	89	(27)	2281	(687)	19937	(6002)
7	127	107	(33)	3217	(969)	21701	(6533)
13	8191	127	(39)	4253	(1281)	23209	(6987)
17	131071	521	(157)	4423	(1332)	44497	(13395)
19	524287	607	(183)	9689	(2917)	86243	(25962)

TABLE 5.1 *Mersenne primes* $2^p - 1$ *(or their numbers of decimal digits) with* $p < 100{,}000$. *Of these p, Mersenne missed* 61, 89, *and* 107, *and wrongly put in* 67 *and* 257.

PERFECT NUMBERS

The ancients were particularly intrigued by the equations

$$6 = 1 + 2 + 3$$

$$28 = 1 + 2 + 4 + 7 + 14$$

which show that both 6 and 28 are the sum of *all* the smaller numbers that divide evenly into them. They called them *perfect*. It was suggested that God made the world in 6 days because 6 was a perfect number.

Two thousand years before Mersenne, Euclid (Book IX, Prop. 36) discovered an interesting connection between perfect numbers and what we now call Mersenne primes.

> If $M = 2^p - 1$ is a Mersenne prime,
> then the Mth triangular number,
> $$\Delta_M = \frac{1}{2} M(M + 1)$$
> is a perfect number.

For instance, 31 is a Mersenne prime, and the 31st triangular number is $16 \times 31 = 496$, whose smaller divisors are

$$\left.\begin{array}{llllll} 1, & 2, & 4, & 8, & 16, & \text{with sum } 31 \\ 31, & 2 \times 31, & 4 \times 31, & 8 \times 31, & & \text{with sum } 15 \times 31 \end{array}\right\} \text{total } 16 \times 31.$$

The same thing happens for every Mersenne prime.

Are there any other perfect numbers? The great Swiss mathematician Leonard Euler (1707–1783) showed that all the even perfect numbers are of Euclid's form. All we know about the odd ones is that they must have at least 300 decimal digits and many factors. There probably aren't any!

FERMAT'S NUMBERS

Which of the numbers

$$11, 101, 1001, 100001, 1000001. \ldots$$

are primes? Well, 11 divides the numbers that are 1 more than odd powers of 10:

$$1001 = 11 \times 91; \; 100001 = 11 \times 9091; \; 10000001 = 11 \times 909091; \; \ldots$$

then 101 divides the ones that are 1 more than odd powers of 100:

$$1000001 = 101\times9901;\ 10000000001 = 101\times99009901;\ \ldots$$

and 1001 divides those that are 1 more than odd powers of 1000:

$$1000000001 = 1001\times999001;$$
$$1000000000000001 = 1001\times999000999001,$$

and so on. Thus the only ones that have any chance of being prime are those with no odd factors in the exponent. In fact,

$$10^1 + 1 = 11 \text{ is prime, and}$$
$$10^2 + 1 = 101 \text{ is prime, but}$$
$$10^4 + 1 = 73 \times 137 \text{ is composite, as are}$$
$$10^8 + 1 = 17 \times 5882353,$$
$$10^{16} + 1 = 353 \times 449 \times 641 \times 1409 \times 69857,$$
$$10^{32} + 1 = 19841 \times 976193 \times 6187457 \times 834427406578561,$$
$$10^{64} + 1 = 1265011073 \times 15343168188889137818369$$
$$\times 515217525265213267447869906815873, \text{ and}$$
$$10^{128} + 1 = 257 \times 15361 \times 453377 \times \text{a prime of 116 digits.}$$

Indeed, we don't know if this sequence contains any primes other than 11 and 101.

We can do such calculations in any base. It remains true that

$b^n + 1$
can only be prime if n is a power of 2.

In 1640, Pierre de Fermat guessed that the base 2 examples were all prime. Today the numbers $2^{2^m} + 1$ are called **Fermat numbers**. In fact, as Fermat knew,

$$\begin{aligned} 2^1 + 1 &= 3, \\ 2^2 + 1 &= 5, \\ 2^4 + 1 &= 17, \\ 2^8 + 1 &= 257, \\ \text{and } 2^{16} + 1 &= 65537 \end{aligned}$$

are all primes, but in 1732 Euler found that the next Fermat number is composite:

$$2^{32} + 1 = 4294967297 = 641 \times 6700417.$$

In 1880, F. Landry, at age 82, showed that

$$2^{64} + 1 = 274177 \times 67280421310721,$$

and in 1975 Brillhart and Morrison found that

$$2^{128} + 1 = 59649589127497217 \times 5704689200685129054721.$$

In 1981 Richard Brent and John Pollard factored

$$\begin{aligned} 2^{256} + 1 = {} & 1238926361552897 \\ & \times 93461639715357977769163558199606896584051237541638188580280321. \end{aligned}$$

Even before these factorizations were found, it was known that these numbers were composite, as are (see note on page 148)

$$2^{2^9} + 1, \quad 2^{2^{10}} + 1, \quad 2^{2^{11}} + 1, \quad \ldots, \quad 2^{2^{23}} + 1.$$

$2^{2^{24}} + 1$, with 5050446 digits, is currently the first one in doubt.

We know that Euclid was interested in perfect numbers, whose theory involves the Mersenne primes, but we don't know whether he ever thought about the Fermat primes—but he should have!

The great German mathematician Carl Friedrich Gauss (1777–1855) proved as a young man the surprising fact that if p is a Fermat prime, then a regular polygon with p sides can be constructed with ruler and compass, using Euclid's rules. Euclid gave constructions for the triangle and pentagon, but before Gauss no-one had constructed any larger prime polygons. It's said that Gauss requested that a regular

17-gon be inscribed on his tombstone. This wasn't done, but there is a regular 17-gon on a monument to Gauss in Braunschweig.

Gauss is surely the mathematician who has made the deepest contributions to number theory, as well as to many other branches of mathematics and the sciences (with Weber he designed an electric telegraph).

It's easy to construct a regular 85-gon, using constructions for the 5-gon and 17-gon, and since angles can be bisected you can construct regular 170-gons, 340-gons, . . . and, more generally,

$$2^k pqr \cdots \text{-gons},$$

where p, q, r, . . . are distinct Fermat primes. Gauss asserted that these were the only regular polygons constructible with ruler and compass.

The only known such polygons with an odd number of sides are those for which the number divides

$$2^{32} - 1 = 3 \times 5 \times 17 \times 257 \times 65537 = 4294967295.$$

The divisors are

1, 3, 5, 15, 17, 51, 85, 255, 257, 771, 1285, 3855, 4369, 13107, 21845, 65535, 65537, 196611, 327685, 983055, 1114129, 3342387, 5570645, 16711935, 16843009, 50529027, 84215054, 252645135, 286331153, 858993459, 1431655765, 4294967295.

William Watkins noticed that you can get their binary expansions just by reading the first 32 rows of the Pascal triangle modulo 2!

Thus, the binary number

1	is	1
1 1	is	3
1 0 1	is	5
1 1 1 1	is	15
1 0 0 0 1	is	17

It's quite probable that there are no more such odd polygons, because it seems likely that

$$3, 5, 17, 257 \text{ and } 65537$$

are the only prime Fermat numbers.

We can show that $2^{32} + 1$ is composite, using congruences mod $p = 641$. Since

$$p = 625 + 16 = 5^4 + 2^4 \text{ and } p - 1 = 5 \times 128 = 5 \times 2^7,$$

$$5 \times 128 \equiv -1 \quad \text{and} \quad 2^4 \equiv -5^4, \quad \text{modulo } p$$

Thus,

$$2^{32} = 2^4 \times 2^{28} \equiv -5^4 \times 128^4 \equiv -(-1)^4 = -1.$$

How can people show that numbers are not prime without being able to find any factors? Another of Fermat's discoveries provides the answer.

FERMAT'S TEST FOR PRIMES

Fermat showed that any odd prime number p must satisfy

FERMAT'S TEST:

$b^{p-1} \equiv 1 \bmod p$, for any b not divisible by p.

So if a number *doesn't* satisfy this condition, it can't be prime. We'll explain why the test works in Chapter 6, but here we'll use it to tell us that 91 isn't prime. If it *were* prime, then, by the test, 2^{90} would be congruent to 1, mod 91. But, working mod 91,

$$2^6 = 64 \equiv -27,$$

$$\text{so } 2^{12} \equiv (-27)^2 = 729 \equiv 1$$

since $8 \times 91 = 728$.

$$2^{84} = (2^{12})^7 \equiv 1^7 = 1 \text{ and}$$

$$2^{90} = 2^{84} \times 2^6 \equiv 1 \times (-27),$$

which is not congruent to 1.

Fermat's test looks harder than the calculations needed to find the factorization $91 = 7 \times 13$, but for large numbers the situation is reversed. To find the factors of a 50-digit number, the naive way might need more than 10^{23} trial divisions, which would take thousands of years, even on a supercomputer. But Fermat's test may prove your number composite using only a few hundred multiplications of 50-digit numbers, done today in a fraction of a second.

With a computer it's now fairly easy to test numbers for primality, but factoring is still hard (although there are much quicker ways than trial division).

The number $341 = 11 \times 31$ passes Fermat's test to base 2, even though it isn't prime! Using congruences, this is easy. Since $2^5 = 32$, we have

$$2^5 \equiv +1 \pmod{31},$$

$$2^5 \equiv -1 \pmod{11},$$

and so

$$2^{10} \equiv +1 \quad (\text{mod } 11 \text{ and } 31).$$

We can deduce from this that 2^{10} is congruent to 1 (mod 341) and therefore so is 2^{340}.

So Fermat's test is *necessary* for primality, but it's not *sufficient*. Since $3^{340} \equiv 56 \pmod{341}$, the base 3 Fermat test shows that 341 is composite, but there are some special numbers, the **Carmichael numbers**, that are composite, although they satisfy the Fermat test for many bases.

The smallest Carmichael number is 561, which satisfies Fermat's test for all bases b not divisible by 3, 11, or 17. This is because b^{560} is a power of each of b^2, b^{10} and b^{16}, which, by Fermat's test, are congruent to 1 modulo 3, 11, and 17, respectively.

Sir John Wilson (1741–1793) gave another test for primes:

WILSON'S TEST:

p is a prime precisely when $(p-1)! \equiv -1 \bmod p$.

Unlike Fermat's test, Wilson's is both necessary and sufficient for primality. In the next chapter we'll show that both these tests work. Unfortunately, the calculations needed to find $(p - 1)!$ modulo p are even more time-consuming than testing p by trial divisors.

HOW FREQUENT ARE THE PRIMES?

Up to 10 there are 4 primes, so 1 in 2.5 numbers is prime.

up to 100	there are	25	primes, i.e. 1 in	4
10^3		168		6
10^4		1229		8.1
10^5		9592		10.4
10^6		78498		12.7
10^7		664579		15
10^8		5761455		17.3
10^9		50847534		19.8

It seems that, up to 10^n, roughly 1 in every $2.3n$ of the numbers is a prime. What happens in general?

LEGENDRE'S LOGARITHMIC LAW

> Of the numbers up to N,
> roughly 1 in every l is prime,
> where l is the natural logarithm of N.

The natural logarithm of N is 2.30258509 . . . times the base 10 logarithm of N and is roughly equal to the Nth **harmonic number** (see Chapter 9):

$$H_N = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{N}.$$

In 1896, a full century after Adrienne Marie Legendre (1752–1833) guessed the approximate formula $N/\ln N$ for the number of primes up to N, Jacques Hadamard and Charles-Jacques de la Vallée-

Poussin conclusively established it. They both lived for more than 50 years after producing their simultaneous but independent proofs.

In the meantime, Gauss and Riemann had made improved guesses, expressed in terms of natural logarithms that we'll meet in Chapter 9.

A Little Improvement

Gauss guessed that Legendre's idea should be modified slightly.

> Of the numbers near N,
> roughly $\frac{1}{\ln N}$ are prime.

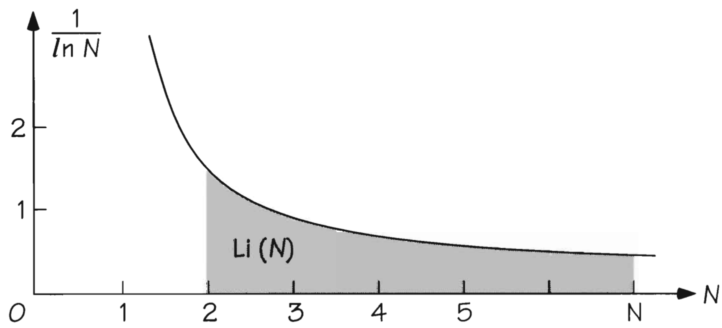


FIGURE 5.2 *Gauss's guess.*

This leads to a little improvement, Li(N), on Legendre's formula. Gauss's guess, Li(N), is the shaded area under the graph of $1/\ln N$ in Figure 5.2.

Riemann's Remarkable Refinement

In 1859, the great German mathematician Georg Friedrich Bernhard Riemann (1826–1866) gave an often better estimate.

The number of primes up to N is roughly

$$R(N) = \mathrm{Li}(N) - \frac{1}{2}\,\mathrm{Li}(N^{\frac{1}{2}}) - \frac{1}{3}\,\mathrm{Li}(N^{\frac{1}{3}}) - \frac{1}{5}\,\mathrm{Li}(N^{\frac{1}{5}}) + \frac{1}{6}\,\mathrm{Li}(N^{\frac{1}{6}}) - \cdots.$$

The coefficient of $\frac{1}{n}\,\mathrm{Li}(N^{1/n})$ is the nth **Möbius number**, $\mu(n)$, which is 0 if n has a repeated factor and otherwise 1 or -1, if it has an even or odd number of prime factors.

It's now known that Gauss's guess, Li(N), is a better estimate than Legendre's, $N/\ln N$, and it seems that Riemann's refinement is usually better still. In fact, under some very plausible assumptions, Rubinstein & Sarnak have shown that it is better exactly 99.07% of the time. However, in 1914, John Edensor Littlewood had already proved that occasionally Riemann's refinement is worse than Gauss's guess, and his student, Skewes, showed that it had to happen before

$$10^{10^{10^{34}}} \text{ (Skewes's number).}$$

We don't know any particular N for which this happens; but Sherman Lehman has proved that it happens for at least 10^{500} numbers with 1166 or 1167 decimal digits.

In the same posthumous papers, Riemann made a famous conjecture that today is called the **Riemann hypothesis**. If this is true, the error in the approximation is less than some multiple of n^k for any $k > 1/2$, but the proved estimates are not nearly as good as that. We don't even know if the error is less than $n^{0.99}$! And it wouldn't be if the Riemann zeta-function had some zero $x+yi$ with $x > 0.99$ (see Chapter 9). Riemann's "hypothesis" is the most tantalizing of the unsolved problems of mathematics.

How Good Are the Guesses?

Table 5.2 compares Legendre's, Gauss's, and Riemann's guesses.

N	Number of Primes up to N, $\pi(n)$	... and the errors in: Legendre's Law $\frac{N}{\ln N} - \pi(N)$	Gauss's Guess $\mathrm{Li}(N) - \pi(N)$	Riemann's Refinement $R(N) - \pi(N)$
10	4	0	2	
10^2	25	−3	5	1
10^3	168	−23	10	0
10^4	1229	−143	17	−2
10^5	9592	−906	38	−5
10^6	78498	−6116	130	29
10^7	664579	−44158	339	88
10^8	5761455	−332774	754	97
10^9	50847534	−2592592	1701	−79
10^{10}	455052511	−20758029	3104	−1828
10^{11}	4118054813	−169923159	11588	−2318
10^{12}	37607912018	−1416705183	38263	−1476
10^{13}	346065536839	−11992858452	108971	−5773
10^{14}	3204941750802	−102838308636	314890	−19200
10^{15}	29844570422669	−891604962452	1052619	73218
10^{16}	279238341033925	−7804289844393	3214632	327052

TABLE 5.2 *Errors (to the nearest integer) in formulas for* π(n).

WHICH NUMBERS ARE SUMS OF TWO SQUARES?

Fermat found the fascinating fact that a prime number, p, can be written as the sum of two squares just if $p + 1$ isn't divisible by 4. The expression is then unique; for instance,

$$2 = 1^2+1^2,\quad 5 = 2^2+1^2,\quad 13 = 3^2+2^2,\quad 17 = 4^2+1^2,$$

$$29 = 5^2+2^2,\quad 37 = 6^2+1^2,\quad 41 = 5^2+4^2, \ldots,$$

But 3, 7, 11, 19, 23, 31, 43, . . . are *not* sums of two squares. This last part is easy to see: Square numbers leave remainder 0 or 1 on division by 4 (Chapter 2), so the sum of two of them can only leave remainder 0, 1, or 2. The major part of the proof of Fermat's fact is really a property of Gauss's complex prime numbers, and we'll explain it in Chapter 8.

To see if an arbitrary positive number is the sum of two squares, factor it into prime powers:

$$p^a q^b r^c \ldots .$$

Then it *is* the sum of two squares just if $p^a + 1$, $q^b + 1$, $r^c + 1, \ldots$ are none of them divisible by 4.

Thus, for $1000 = 2^3 5^3$, we look at $2^3 + 1 = 9$ and $5^3 + 1 = 126$. Neither is divisible by 4, so 1000 is the sum of two squares. Can you find them? Notice that the representation is no longer necessarily unique: $1000 = 900 + 100 = 676 + 324$.

On the other hand, $999 = 3^3 \times 37$ and $3^3 + 1 = 28$ is divisible by 4, so 999 is *not* the sum of two squares. Nor is $1001 = 7 \times 11 \times 13$, since $7 + 1$ or $11 + 1$ is divisible by 4.

FOURTEEN FRUITFUL FRACTIONS

Is there a simple formula that gives all the primes? Most mathematicians would say no, though some of them have given rather artificial ones that do the job (usually based on Wilson's test).

Here's our own best effort along these lines, although it isn't actually a formula.

$\frac{17}{91}$	$\frac{78}{85}$	$\frac{19}{51}$	$\frac{23}{38}$	$\frac{29}{33}$	$\frac{77}{29}$	$\frac{95}{23}$	$\frac{77}{19}$	$\frac{1}{17}$	$\frac{11}{13}$	$\frac{13}{11}$	$\frac{15}{14}$	$\frac{15}{2}$	$\frac{55}{1}$
A	B	C	D	E	F	G	H	I	J	K	L	M	N

FIGURE 5.3 *Fourteen fruitful fractions for producing primes.*

Start with the number 2 and then repeatedly multiply by the first of our fourteen fruitful fractions (Figure 5.3) that yields a whole number answer. The letters over the arrows in Figure 5.4 indicate which of these fractions is being used:

$$2 \xrightarrow{\times M} 15 \xrightarrow{\times N} 825 \xrightarrow{E} 725 \xrightarrow{F} 1925 \xrightarrow{K} 2275 \xrightarrow{A} 425 \xrightarrow{B} 390 \xrightarrow{J} 330$$

$$\xrightarrow{E} 290 \xrightarrow{F} 770 \xrightarrow{K} 910 \xrightarrow{A} 170 \xrightarrow{B} 156 \xrightarrow{J} 132$$

$$\xrightarrow{E} 116 \xrightarrow{F} 308 \xrightarrow{K} 364 \xrightarrow{A} 68 \xrightarrow{I} 4 = 2^2$$

FIGURE 5.4 *The first nineteen steps yield 2^2, so 2 is the first prime.*

Every now and then you will see a new power of 2, for example, after the first 19 steps of Figure 5.4. The exponents of these are just the prime numbers 2, 3, 5, 7, . . . , in order of magnitude. Here the fractions *conceal* a version of the sieving process, but we'll see in the next chapter that fractions really can help us to understand the prime numbers.

NOTE

The following amount is now known about the prime factorizations of the 9th to the 13th Fermat numbers:

$F_9 = 2424833.7455602825647884208337395736200454918783366342657.p_{99}$

$F_{10} = 45592577.6487031809.4659775785220018543264560743076778192897.p_{252}$

$F_{11} = 319489.974849.167988556341760475137.3560841906445833920513.p_{564}$

$F_{12} = 114689.26017793.63766529.190274191361.1256132134125569.p_{1187}$

$F_{13} = 2710954639361.2663848877152141313.$
$3603109844542291969.319546020820551643220672513.p_{2391}$

where the numbers written out in full are primes, and p_N or c_N denotes an N-digit prime or composite number.

See Richard P. Brent. Factorization of the Tenth and Eleventh Fermat numbers, *Math. Comput.*, in press.

REFERENCES

C.L. Baker and F.J. Gruenberger. *The First Six Million Prime Numbers.* Microcard Foundation, Madison, WI, 1953.

P.T. Bateman, J.L. Selfridge and S.S. Wagstaff. The new Mersenne conjecture. *Amer. Math. Monthly,* **96** (1989): 125–128; *MR* **90c**: 11009.

Richard P. Brent and John M. Pollard. Factorization of the eighth Fermat number. *Math.Comput.* **36** (1981): 627-630; *MR* **83h**: 10014.

Richard P. Brent, G.L. Cohen & H.J.J. te Riele. Improved techniques for lower bounds for odd perfect numbers, *Math. Comput.*, **57** (1991) 857-868; *MR* **92c**: 11004.

John Brillhart, D.H. Lehmer, J.L. Selfridge, Bryant Tuckerman, and S.S. Wagstaff. Factorizations of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11, 12$ up to high powers. *Contemp. Math.*, **22**, Amer. Math. Soc., Providence, RI, 1983, 1988; *MR* **84k**: 10005, **90d**: 11009.

John Brillhart, Peter L. Montgomery, and Robert D. Silverman. Tables of Fibonacci and Lucas factorizations. *Math. Comput.*, **50** (1988): 251-260 and S1-S15.

Richard E. Crandall, J. Doenias, C. Norrie, and Jeff Young. The twenty-second Fermat number is composite. *Math. Comput.*, **64** (1995): 863-868.

William John Ellison, and M. Mendes-France. *Les Nombres Premiers.* Hermann, Paris, 1975; translated William and Fern Ellison, *Prime Numbers,* Wiley, New York, 1985.

Martin Gardner. *The Sixth Book of Mathematical Games.* W.H. Freeman, San Francisco, 1963; Chapter 9, Patterns and Primes.

Martin Gardner. The remarkable lore of prime numbers. *Scientific Amer.*, **210** no. 3 (Mar. 1964): 120-128.

Richard K. Guy. Conway's prime producing machine. *Math. Mag.* **56** (1983): 26-33.

Richard K. Guy. *Unsolved Problems in Number Theory,* 2nd ed. Springer-Verlag, New York, 1994; Chapter A, Prime Numbers.

A.M. Legendre. *Essai sur la Théorie des Nombres,* Duprat, Paris, 1808.

R. Sherman Lehman. On the difference $\pi(x) - \mathrm{li}(x)$. *Acta Arith.* **11** (1966): 397-410; *MR* **34** #2546.

J.E. Littlewood. Sur la distribution des nombres premiers. *C. R. Acad. Sci. Paris,* **158** (1914): 1869-1872.

Paulo Ribenboim. *The Book of Prime Number Records.* Springer-Verlag, New York, 1988.

Paulo Ribenboim. *The Little Book of Big Primes.* Springer-Verlag, New York, 1991.

G.F.B. Riemann. Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse, *Monatsber. Königl. Preuss. Akad. Wiss. Berlin,* 1859, 671.

Hans Riesel. *Prime Numbers and Computer Methods for Factorization.* Birkhaüser, Boston, 1985.

Michael Rubinstein and Peter Sarnak. Chebyshev's bias, *Experimental Math.*, **3** (1994) 173-197.

Jeff Young and Duncan Buell. The twentieth Fermat number is composite. *Math. Comput.* **50** (1988): 261-263.

6

Further Fruitfulness of Fractions

So far we have concentrated on whole numbers. But there are lots of other numbers, like 2/3, 4/7, . . . , that also behave in very interesting ways. We'll just call these **fractions**, although mathematicians usually refer to them as **rational numbers**. This rather curious name arises from the fact that a rational number is defined to be the *ratio* of two whole numbers.

We expect that you are familiar with the more mundane uses of fractions, so the real theme of this chapter is how fractions can be used to throw light on some subtle properties of whole numbers.

Each fraction has many forms: $\frac{4}{6} = \frac{2}{3} = \frac{6}{9} = \frac{202}{303} = \cdots.$

The **golden rule** for fractions is that you may multiply the **numerator** and the **denominator** by the same number without affecting the value of the fraction. A fraction with no such common factor is in its **lowest terms** ($^2/_3$ in the above example). The golden rule enables you to add, subtract, multiply, and divide fractions:

$$\frac{2}{3} + \frac{1}{4} = \frac{8}{12} + \frac{3}{12} = \frac{11}{12}; \quad \frac{2}{3} \times \frac{1}{4} = \frac{4}{6} \times \frac{1}{4} = \frac{1}{6};$$

$$\frac{2}{3} \div \frac{1}{4} = \frac{^2/_3}{^1/_4} = \frac{^8/_3}{1} = \frac{8}{3}.$$

Of course, fractions arise all over mathematics and the sciences. Indeed, it was this ubiquity that convinced the Pythagorean brotherhood that numbers ruled the world. One of the discoveries that impressed them most was a rather unlikely application of fractions to music: the notes of two similar vibrating strings sound harmonious just when the ratio of their lengths is a simple fraction. We'll return to this in Chapter 8.

FAREY FRACTIONS AND FORD CIRCLES

The British geologist Farey[1] suggested arranging all the proper fractions with (lowest) denominator up to some point, in order of magnitude. For example, the sixth **Farey series** is

$$\frac{0}{1}\quad\frac{1}{6}\quad\frac{1}{5}\quad\frac{1}{4}\quad\frac{1}{3}\quad\frac{2}{5}\quad\frac{1}{2}\quad\frac{3}{5}\quad\frac{2}{3}\quad\frac{3}{4}\quad\frac{4}{5}\quad\frac{5}{6}\quad\frac{1}{1}.$$

Such series have many nice arithmetical properties. For instance, if $\frac{a}{c}$ and $\frac{b}{d}$ are consecutive fractions in the series, then the "cross products," $a\times d$ and $b\times c$, are consecutive integers.

Thus $\frac{3}{5}$ and $\frac{2}{3}$ give the consecutive numbers $3\times3 = 9$ and $5\times2 = 10$. Higher Farey series are obtained by inserting certain fractions: the first fraction to be inserted between $\frac{a}{c}$ and $\frac{b}{d}$ is always the **mediant** fraction $\frac{a+b}{c+d}$. Thus, to get the seventh Farey series from the sixth, we insert

[1]G.H. Hardy remarked (Notes to Ch. III of *An Introduction to the Theory of Numbers*) that Farey did not publish anything on the subject until 1816, when he stated the theorem that "if $\frac{a}{c}$, $\frac{e}{f}$, $\frac{b}{d}$ are three successive terms of a Farey series, then $\frac{e}{f} = \frac{(a+b)}{(c+d)}$ in a note in the *Philos. Mag.* He gave no proof, and it is unlikely that he had found one, since he seems to have been at best an indifferent mathematician. Cauchy, however, saw Farey's statement and supplied the proof (*Exer. Math.*, i, 114–116). Mathematicians generally have followed Cauchy's example in attributing the results to Farey, and the series will no doubt continue to bear his name.

"Farey has a biography of twenty lines in the *Dict. National Biog.*, where he is described as a geologist. As a geologist he is forgotten, and his biographer does not mention the one thing in his life which survives."

$$\frac{0+1}{1+6} \quad \frac{1+1}{4+3} \quad \frac{2+1}{5+2} \quad \frac{1+3}{2+5} \quad \frac{2+3}{3+4} \quad \frac{5+1}{6+1}\ .$$

Warning: forming the mediant is *not* the way to add fractions, unless you're calculating batting averages!

Lester R. Ford found a nice picture for the Farey series. Above each rational number, $\frac{p}{q}$, on the real number line we draw a circle of diameter $1/q^2$, as in Figures 6.1, 6.2, and 6.3. It turns out that these Ford circles never overlap, but they often touch. In fact the circles at

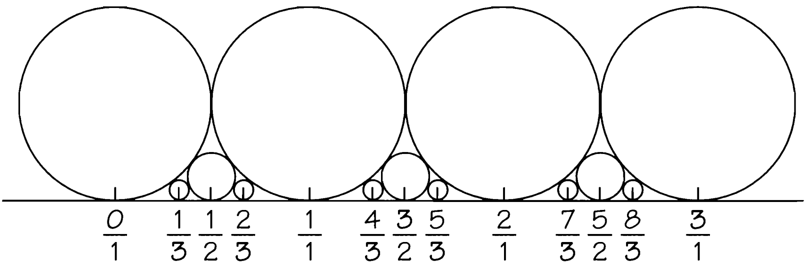


FIGURE 6.1 *Ford circles corresponding to wholes, halves, and thirds.*

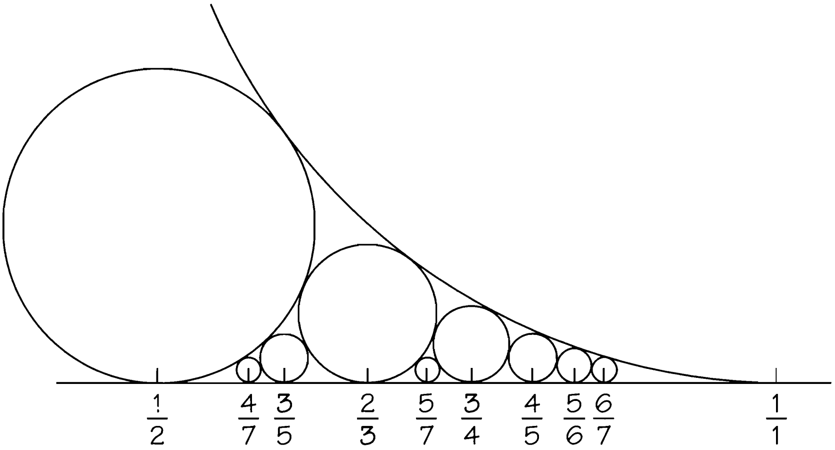


FIGURE 6.2 *Enlargement of Figure 6.1, showing half of the Farey series of order 7.*

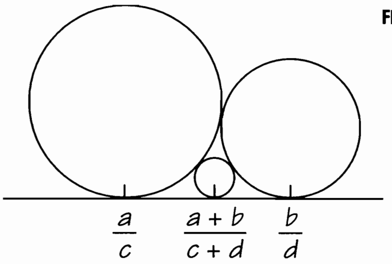


FIGURE 6.3 *Two Ford circles and their mediant.*

a/c and b/d touch just when ad and bc are consecutive whole numbers, and then the largest circle between them is that for the mediant fraction, $(a+b)/(c+d)$.

EULER'S TOTIENT NUMBERS

While we're on the subject, let's ask for how many of the fractions

$$\frac{0}{n}\ \frac{1}{n}\ \frac{2}{n}\ \cdots\ \frac{n-1}{n}$$

is n the least possible denominator? Thus, for $n = 12$, these fractions simplify to

$$\frac{0}{1}\ \frac{1}{12}\ \frac{1}{6}\ \frac{1}{4}\ \frac{1}{3}\ \frac{5}{12}\ \frac{1}{2}\ \frac{7}{12}\ \frac{2}{3}\ \frac{3}{4}\ \frac{5}{6}\ \frac{11}{12}$$

so just the four fractions 1/12, 5/12, 7/12, and 11/12 really need to be written as twelfths. Here's a little table for small denominators.

denominator	fractions	number
1	$\frac{0}{1}$	1
2	$\frac{1}{2}$	1
3	$\frac{1}{3}, \frac{2}{3}$	2
4	$\frac{1}{4}, \frac{3}{4}$	2
5	$\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}$	4
6	$\frac{1}{6}, \frac{5}{6}$	2
7	$\frac{1}{7}, \frac{2}{7}, \frac{3}{7}, \frac{4}{7}, \frac{5}{7}, \frac{6}{7}$	6
8	$\frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}$	4
9	$\frac{1}{9}, \frac{2}{9}, \frac{4}{9}, \frac{5}{9}, \frac{7}{9}, \frac{8}{9}$	6
10	$\frac{1}{10}, \frac{3}{10}, \frac{7}{10}, \frac{9}{10}$	4

For a general n, the number of these fractions is called Euler's **totient number**, $\phi(n)$. So our table shows that

$$\phi(1)=\phi(2)=1, \quad \phi(3)=\phi(4)=\phi(6)=2,$$

$$\phi(5)=\phi(8)=\phi(10)=4, \quad \phi(7) = \phi(9) = 6.$$

It seems that every totient number that happens, happens at least twice, but nobody has yet managed to prove this. If there's one that doesn't, it must have more than 10,000 digits!

What is the hundredth totient number, $\phi(100)$? A fraction with denominator 100 will simplify only if the numerator is divisible by 2 or 5. Half the cases are odd, and 4 out of 5 of the remainder don't divide by 5, so

$$\phi(100) = 100 \times \frac{1}{2} \times \frac{4}{5} = 40$$

is the number of noncancelling proper fractions with denominator 100.

In a similar way, the nth totient number, $\phi(n)$, is

$$n \times \left(1 - \frac{1}{p}\right) \times \left(1 - \frac{1}{q}\right) \times \left(1 - \frac{1}{r}\right) \times \cdots,$$

where p, q, r, . . . are the different prime divisors of n.

Let's arrange the 12 fractions 0/12, 1/12, . . . , 11/12 according to their denominator after simplification:

The fractions:	Their number
$\frac{1}{12}$ $\frac{5}{12}$ $\frac{7}{12}$ $\frac{11}{12}$	$\phi(12) = 4$
$\frac{1}{6}$ $\frac{5}{6}$	$\phi(6) = 2$
$\frac{1}{4}$ $\frac{3}{4}$	$\phi(4) = 2$
$\frac{1}{3}$ $\frac{2}{3}$	$\phi(3) = 2$
$\frac{1}{2}$	$\phi(2) = 1$
$\frac{0}{1}$	$\phi(1) = 1$

Total: $\phi(12) + \phi(6) + \phi(4) + \phi(3) + \phi(2) + \phi(1) = 12$

In fact,

Any whole number is the total of the totients of its divisors.

How totients tot up.

This result gives another way of working out the totient numbers. To see how it works, we'll find $\phi(12)$ again:

We start with			$\phi(1) = 1$
then	$\phi(1)+\phi(2) =$	2 gives	$\phi(2) = 1$
then	$\phi(1)+\phi(3) =$	3 gives	$\phi(3) = 2$
then	$\phi(1)+\phi(2)+\phi(4) =$	4 gives	$\phi(4) = 2$
then	$\phi(1)+\phi(2)+\phi(3)+\phi(6) =$	6 gives	$\phi(6) = 2$
then	$\phi(1)+\phi(2)+\cdots+\phi(12) =$	12 gives	$\phi(12) = 4$

HOW LONG IS A FAREY SERIES?

The answer is obviously that the nth Farey series has length

$$1 + \phi(1) + \phi(2) + \phi(3) + \cdots + \phi(n-1) + \phi(n)$$

(the initial 1 comes from the fact that we count both 0/1 and 1/1).

There's no simple formula for this particular sum of totient numbers, but it is known that the answer is about

$$3\left(\frac{n}{\pi}\right)^2 \approx 0.3039635509 \times n^2$$

and that the approximation gets proportionally better as n gets larger. For example, the tenth Farey series has length 33, as compared with $3 \times 100/\pi^2 \approx 30.4$, and the one hundredth has length 3045, as compared with $(3 \times 100^2)/\pi^2 \approx 3039.6$.

FRACTIONS CYCLE INTO DECIMALS

Everyone who has played with numbers must have noticed the relations between fractions with the same denominator:

$$\frac{1}{7} = .1428571428571428\ldots$$

$$\frac{2}{7} = .2857142857142857\ldots$$

$$\frac{3}{7} = .4285714285714285\ldots$$

$$\frac{4}{7} = .5714285714285714\ldots$$

$$\frac{5}{7} = .7142857142857142\ldots$$

$$\frac{6}{7} = .8571428571428571\ldots$$

One cycle of repeating digits works for all six fractions as shown in Figure 6.4(a). However, when we come to thirteenths,

$$\frac{1}{13} = .0769230769230769\ldots$$

$$\frac{2}{13} = .1538461538461538\ldots$$

$$\frac{3}{13} = .2307692307692307\ldots$$

$$\ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots\ \ldots$$

we find that there are two different cycles (Figures 6.4(b) & (c)).

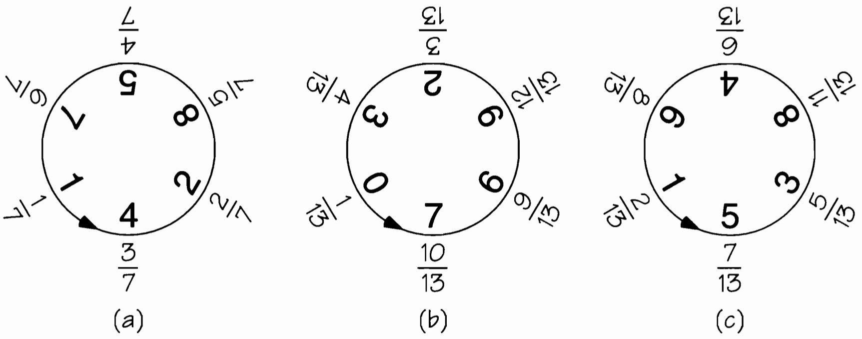


FIGURE 6.4 *Several fractions served by each cycle.*

Everybody knows that there are also two different cycles for denominator 3:

$$\frac{1}{3} = .333333\ldots \quad \frac{2}{3} = .666666\ldots,$$

and you possibly know the five different cycles for 11:

$$\frac{1}{11} = .09090\ldots \quad \frac{2}{11} = .18181\ldots \quad \frac{3}{11} = .27272\ldots \quad \frac{4}{11} = .36363\ldots \quad \frac{5}{11} = .45454\ldots$$

and that $\frac{10}{11}$, $\frac{9}{11}$, $\frac{8}{11}$, $\frac{7}{11}$, $\frac{6}{11}$ are the same, each starting one digit later. Table 6.1 shows the cycles for all the other primes less than 100.

Why do the fractions for a given denominator collect into cycles like this? How long are the cycles? Some of the answers are easy, and some are still too hard for us.

It's easy to see how the cycles happen: if we multiply the expansion of $\frac{1}{7}$ by 10, we get

$$1.428571428\ldots = 10\left(\frac{1}{7}\right) = \frac{10}{7} = 1\,\frac{3}{7}$$

and so indeed

$$.428571428\ldots = \frac{3}{7},$$

and

$$14.285714285\ldots = \frac{100}{7} = 14\,\frac{2}{7}$$

$$142.857142857\ldots = \frac{1000}{7} = 142\,\frac{6}{7}$$

$$1428.571428571\ldots = \frac{10000}{7} = 1428\,\frac{4}{7}$$

$$14285.714285714\ldots = \frac{100000}{7} = 14285\,\frac{5}{7}$$

$$142857.142857142\ldots = \frac{1000000}{7} = 142857\,\frac{1}{7}$$

You see that it is because

$$10 \equiv 3 \bmod 7 \quad 10^2 \equiv 2 \bmod 7 \quad 10^3 \equiv 6 \bmod 7$$
$$10^4 \equiv 4 \bmod 7 \quad 10^5 \equiv 5 \bmod 7 \quad 10^6 \equiv 1 \bmod 7$$

that $10/7$ has the same fractional part as $\frac{3}{7}$, $\frac{100}{7}$ has the same fractional part as $2/7$, and so forth.

In other words, you get the numerators in the cyclic order

$$1, 3, 2, 6, 4, 5, 1, 3, 2, \ldots$$

just by repeatedly multiplying by 10 modulo 7.

The reason that the fractions with denominator 13 come in more than one cycle is that, modulo 13, the powers of 10 repeat with period 6:

$$10^0 = 1,\ 10^1 = 10,\ 10^2 = 9,\ 10^3 = 12,\ 10^4 = 3,\ 10^5 = 4,\ 10^6 = 1,$$

and so the first cycle doesn't contain all the fractions.

Obviously, for denominator p,

> The length of the first cycle
> is the smallest number l
> with $10^l \equiv 1$ modulo p.

17:	0588235294117647
19:	052631578947368421
23:	0434782608695652173913
29:	0344827586206896551724137931
31:	032258064516129 096774193548387
37:	027 054 081 135 162 189 243 297 378 459 486 567
41:	02439 04878 07317 09756 12195 14634 26829 36585
43:	023255813953488372093 046511627906976744186
47:	0212765957446808510638297872340425531914893617
53:	0188679245283 0377358490566 0754716981132 0943396226415
59:	0169491525423728813559322033- -898305084745762711864406779661
61:	016393442622950819672131147540- -983606557377049180327868852459
67:	014925373134328358208955223880597 029850746268656716417910447761194
71:	01408450704225352112676056338028169 09859154929577464788732394366197183
73:	01369863 02739726 04109589 05479452 06849315 08219178 12328767 15068493 16438356
79:	0126582278481 0253164556962 0379746835443 0506329113924 0759493670886 1518987341772
83:	01204819277108433734939759036144578313253 02409638554216867469879518072289156626506
89:	01123595505617977528089887640449438202247191 03370786516853932584269662921348314606741573
97:	010309278350515463917525773195876288659793814432- -989690721649484536082474226804123711340206185567

TABLE 6.1 *Cycles for fractions with prime denominators (in* ***bold****).*

The cycles in Table 6.1, for a given denominator, are all of the same length. For example, denominator 73 gives nine 8-cycles:

$$\frac{1}{73} = .01369863\ldots \quad \frac{2}{73} = .02739726\ldots \quad \frac{3}{73} = .04109589\ldots$$

$$\frac{4}{73} = .05479452\ldots \quad \frac{5}{73} = .06849315\ldots \quad \frac{6}{73} = .08219178\ldots$$

$$\frac{9}{73} = .12328767\ldots \quad \frac{11}{73} = .15068493\ldots \quad \frac{12}{73} = .16438356\ldots$$

Why is this? Well, adding up copies of a repeating decimal can't make the cycle length longer (although it might make it shorter) so, for example, the length for $\frac{20}{73}$ is no longer than for $\frac{1}{73}$ since

$$\frac{1}{73} + \frac{1}{73} + \cdots + \frac{1}{73} = \frac{20}{73}.$$

But also

$$\frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} + \frac{20}{73} = \frac{220}{73} = 3\,\frac{1}{73}.$$

since $11 \times 20 = (3 \times 73) + 1$, and so the cycle length for $1/73$ is no longer than that for $20/73$. So they must be the same.

How did we find the number 11, which "undid" 20, mod 73? In Chapter 3 we saw that all the numbers $\frac{1}{1}, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{72}$ exist, mod 73, and we told you how to work them out. If you do this for $\frac{1}{20}$, mod 73, you'll find that it's 11. So since there's always an "undoing" multiplier. . .

For a prime denominator (other than 2 or 5), *all* the cycles have the same length.

You can see from Table 6.2 that there are quite a lot of primes for which the decimal period of $1/p$ has its full length, $p-1$. We'll call these the **long primes** (to base 10). Of course, in every case in Table 6.2,

$$\text{number} \times \text{length} = p-1;$$

since, together, the cycles cope with all the $p-1$ numerators 1, 2, . . . , $p - 1$.

prime denominator p	3	7	11	13	17	19	23	29	31	37	41	43	47	53	59
number of cycles	2	1	5	2	1	1	1	1	2	12	8	2	1	4	1
length of each cycle	1	6	2	6	16	18	22	28	15	3	5	21	46	13	58

prime	61	67	71	73	79	83	89	97	101	103	107	109	113	127	131	137
number	1	2	2	9	6	2	2	1	25	3	2	1	1	3	1	17
length	60	33	35	8	13	41	44	96	4	34	53	108	112	42	130	8

prime	139	149	151	157	163	167	173	179	181	191	193	197	199	211
number	3	1	2	2	2	1	2	1	1	2	1	2	2	7
length	46	148	75	78	81	166	86	178	180	95	192	98	99	30

prime	223	227	229	233	239	241	251	257	263	269	271	277	281	283
number	1	2	1	1	34	8	5	1	1	1	54	4	10	2
length	222	113	228	232	7	30	50	256	262	268	5	69	28	141

TABLE 6.2 *Cycle structures for prime denominators p, $3 \le p \le 283$.*

Both the number and the length of the cycles must exactly divide $p - 1$. But we already know what the length is:

> The smallest l for which
> $10^l \equiv 1, \bmod p,$

and this exactly divides $p-1$. For instance, for $p = 73$ the cycles have length 8, since

$$10^8 \equiv 1 \pmod{73}$$

and indeed, 8 does divide 72. If you take the ninth power of both sides of this, you'll see that

$$10^{72} \equiv 1 \pmod{73}$$

In just the same way,

> $10^{p-1} \equiv 1, \bmod p,$

for every prime number p other than 2 or 5. And there's nothing special about the base 10: we can use any other base, b,

FERMAT'S "LITTLE" THEOREM:

Provided p doesn't divide b, $b^{p-1} \equiv 1$, mod p.

REPEATED SHUFFLING

Of course, the usual reason for shuffling cards is to mix them up, and the more you shuffle them, the more mixed up you hope they'll be. However, if you repeat exactly the same shuffle the right number of times, you'll find that your cards return to their original order. We'll find this number for three kinds of shuffles: the in and out riffle shuffle and Monge's shuffle.

RIFFLE SHUFFLES

These are often used by magicians and card manipulators (Figure 6.5).

You cut a deck of $2n$ cards into two half-decks, n cards in each, and then slickly interleave them so that the cards in the final deck come alternately from your left and right hands (the **in shuffle**, Figure 6.5(a)), or your right and left hands (the **out shuffle**, Figure 6.5(b)). Notice that in the out shuffle, the outer cards stay where they are: you're only shuffling the inner $2n - 2$ cards.

In the **in shuffle** you'll see that the cards we've numbered

$$1\ 2\ 3\ 4\ 5\ 6\ 7\ 8$$

end up in the positions originally occupied by numbers

$$2\ 4\ 6\ 8\ 1\ 3\ 5\ 7.$$

The general rule is that card number k ends in the place that was originally occupied by the card whose number was $2k$, modulo $2n+1$. After s such shuffles, card number k will be in the place that was originally occupied by the card whose number was $2^s k$, mod $2n+1$, so s shuffles will restore the original order just when

$$2^s \equiv 1, \text{ mod } 2n + 1.$$

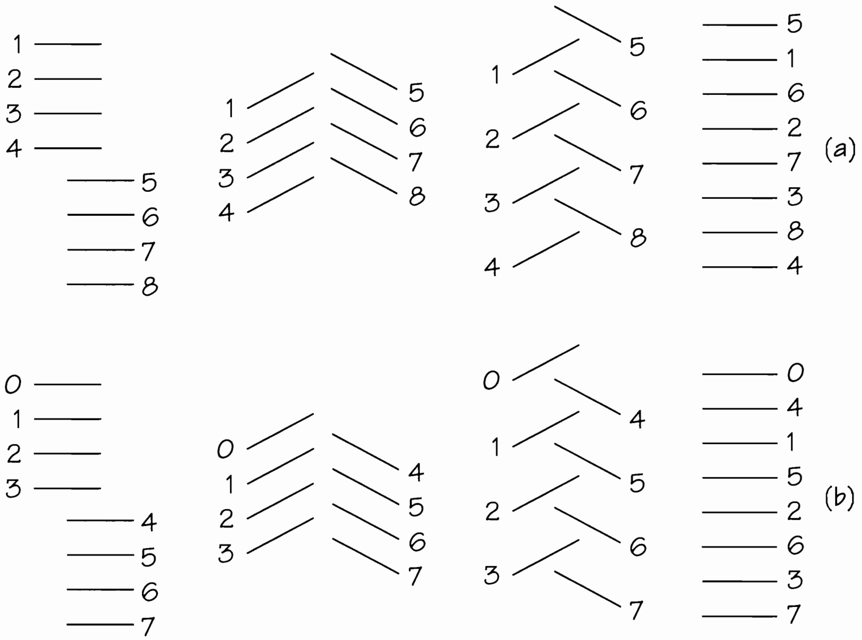


FIGURE 6.5 *Riffle shuffles. (a) The in shuffle. (b) The out shuffle.*

By Fermat's Little Theorem, when $2n + 1$ is a prime number, p, the required number of shuffles always divides $p - 1 = 2n$, the number of cards:

> If you in-shuffle $2n$ cards $2n$ times, and $2n+1$ is prime, then cards will come back to their original order.

It happens that 53 is a long prime to base 2, so you need exactly 52 in shuffles to restore the original order to a deck of 52 cards.

For the **out shuffle** we number the cards from 0 as in Figure 6.5(b) and get similar results: s shuffles will restore the order just when $2^s \equiv 1$, mod $2n-1$, and so:

> If you out-shuffle $2n$ cards $2n-2$ times, where $2n-1$ is prime, the cards will come back to their original order.

Since $2^8 = 256$ is congruent to 1 mod 51, just 8 out shuffles suffice to restore an ordinary deck of 52 cards. Our friend Persi Diaconis is one of the few dextrous people who have achieved eight consecutive perfect out shuffles.

MONGE'S SHUFFLE

Gaspard Monge investigated the following kind of shuffle. Take cards from the top of the deck in your left hand alternately to the bottom and top of the deck in your right hand, as in Figure 6.6.

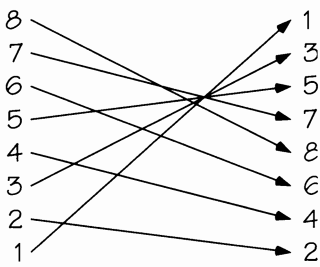


FIGURE 6.6 *Monge's shuffle.*

This time the place originally occupied by card k is the one whose number is congruent to $\pm 2k$, mod $4n+1$. The number of shuffles required to restore the original order is the smallest s for which

$$2^s \equiv \pm 1, \bmod 4n+1.$$

If $4n+1$ is prime, this number divides $2n$:

> If $4n+1$ is prime, then $2n$ Monge's shuffles of a $2n$ card deck restores the original order.

So you'll see that card shuffles are really investigating the cycle behavior of fractions to base 2.

HOW DOES THE CYCLE LENGTH CHANGE WITH THE BASE?

The cycle length of $1/p$ does depend on the base of notation, b (but because it's the smallest n with $b^n \equiv 1 \bmod p$, it actually depends only on the value of $b \bmod p$). Let's look at 1/13 in different bases (Table 6.3).

base 2	$\frac{1}{13}$ = .000100111011 . . .	cycle length	12
base 3	$\frac{1}{13}$ = .002002 . . .	cycle length	3
base 4	$\frac{1}{13}$ = .010323 . . .	cycle length	6
base 5	$\frac{1}{13}$ = .01430143 . . .	cycle length	4
base 6	$\frac{1}{13}$ = .024340531215 . . .	cycle length	12
base 7	$\frac{1}{13}$ = .035245631421 . . .	cycle length	12
base 8	$\frac{1}{13}$ = .04730473 . . .	cycle length	4
base 9	$\frac{1}{13}$ = .062062 . . .	cycle length	3
base 10	$\frac{1}{13}$ = .076923 . . .	cycle length	6
base 11	$\frac{1}{13}$ = .093425A17685 . . .	cycle length	12
base 12	$\frac{1}{13}$ = .0B0B0B . . .	cycle length	2
base 13	$\frac{1}{13}$ = .1	(terminating)	
base 14	$\frac{1}{13}$ = .111111 . . .	cycle length	1

TABLE 6.3 $\frac{1}{13}$ *in different bases. (Here A = 10, B = 11.)*

So the length is

12	for the 4 classes	$b \equiv$	2,6,7,11	mod 13
6	for the 2 classes		4,10	mod 13
4	for the 2 classes		5,8	mod 13
3	for the 2 classes		3,9	mod 13
2	for the 1 class		12	mod 13
1	for the 1 class		1	mod 13

Compare this with the behavior of the fractions $\frac{0}{12}, \frac{1}{12}, \frac{2}{12}, \ldots, \frac{11}{12}$. The lowest denominator is

12 for the 4 fractions		$\frac{1}{12}$				$\frac{5}{12}$		$\frac{7}{12}$				$\frac{11}{12}$
6 for the 2 fractions			$\frac{1}{6}$								$\frac{5}{6}$	
4 for the 2 fractions				$\frac{1}{4}$						$\frac{3}{4}$		
3 for the 2 fractions					$\frac{1}{3}$				$\frac{2}{3}$			
2 for the 1 fraction							$\frac{1}{2}$					
1 for the 1 fraction	$\frac{0}{1}$											

In general, as Euler discovered,

EULER'S TOTIENT RULE:

> The number of bases, mod p,
> in which $\frac{1}{p}$ has cycle length l
> is just the same as the number of fractions
> $$\frac{0}{p-1}, \frac{1}{p-1}, \ldots, \frac{p-2}{p-1}$$
> that have least denominator l.

To see why this is so, we must use a little algebra, namely the fact that an equation can have no more roots (solutions) than its degree. This is proved using only the four rules of arithmetic, so it still works modulo any prime (it *wouldn't* work modulo a *nonprime*, for instance, $x^2-1 = 0$ of degree 2 has *four* solutions: 1, 3, 5, 7 modulo 8).

Now by Fermat's test, $x^{12} - 1$ has its full complement of 12 roots mod 13, so that modulo 13, $x^{12} - 1$ is

$$(x-1)(x-2)(x-3)(x-4)(x-5)(x-6)(x-7)(x-8)(x-9)(x-10)(x-11)(x-12).$$

We can factor $x^{12} - 1$ in various ways; for example,

$$x^{12}-1 = (x^4-1)(x^8+x^4+1).$$

Here the two factors can have *at most* four and eight solutions and so must have *exactly* four and eight solutions, respectively. In this way we see that

$x-1$ has exactly 1 root mod 13,
x^2-1 has exactly 2 roots mod 13,
x^3-1 has exactly 3 roots mod 13,
x^4-1 has exactly 4 roots mod 13,
x^6-1 has exactly 6 roots mod 13,
$x^{12}-1$ has exactly 12 roots mod 13.

Or, in other words,

exactly 1 base, mod 13, yields cycle length 1,
exactly 2 bases, mod 13, yield length dividing 2,
exactly 3 bases, mod 13, yield length dividing 3,
exactly 4 bases, mod 13, yield length dividing 4,
exactly 6 bases, mod 13, yield length dividing 6,
exactly 12 bases, mod 13, yield length dividing 12.

From these facts you can work out exactly how many bases, mod 13, yield each possible length: it's just the same as the way the fractions $\frac{0}{12}, \frac{1}{12}, \frac{2}{12}, \ldots, \frac{11}{12}$ behave. This behavior explains Euler's totient rule.

WILSON'S THEOREM

Sir John Wilson (1741-1793) observed that when p is a prime, the factorial numbers, $(p - 1)!$, always leave the remainder $p - 1$ on division by p. We explain this as follows. We saw above that

$$(x - 1)(x - 2) \cdots (x - (p - 1)) \equiv x^{p-1} - 1 \bmod p.$$

We obtain Wilson's theorem by putting $x = p$ in this.

WILSON'S THEOREM:

For all primes p, $(p-1)! \equiv -1 \bmod p$

LONG PRIMES

The long primes are those for which the period of $1/p$ has the full length $p-1$. To base 10 these are

$$7, 17, 19, 23, 29, 47, 59, 61, 97, 109, 113, 131, 149, 167, \ldots.$$

It seems that about 37 percent of the primes are long in base 10. The illustrious Emil Artin suggested that this number is really

ARTIN'S NUMBER:

$$\frac{1}{2}\times\frac{5}{6}\times\frac{19}{20}\times\frac{41}{42}\times\frac{109}{110}\times\frac{155}{156}\times\frac{271}{272}\times\cdots = 0.3739558136\ldots = C$$

where there is a factor $(p^2-p-1)/(p^2-p)$ for each prime number p.

Although 7 is long in base 10, it *isn't* long in base 2, because the powers of 2 repeat with period 3, modulo 7:

$$2^0 = 1,\ 2^1 = 2,\ 2^2 = 4,\ 2^3 \equiv 1,\ 2^4 \equiv 2, \ldots.$$

On the other hand, 13 is long in base 2, but not in base 10.

By Euler's totient rule, there are always bases in which p is long, since there are certainly fractions with lowest denominator $p-1$. Indeed,

p is long in just $\phi(p-1)$ bases, modulo p.

HOW MANY LONG PRIMES DO WE THINK THERE ARE TO VARIOUS BASES?

There seems to be about the same proportion, C, of long primes in base 2 as in base 10, but for some other bases we apparently get other fractional multiples of C, according to Artin's guess, modified by Dick Lehmer (Table 6.4). No one has proved this or even shown that there is *any* base in which there are infinitely many long primes, but some deep work of Christopher Hooley makes it seem very likely that it is true.

Don't let the fact that we've concentrated on fractions with prime denominators fool you into thinking that those are the only interesting cases. There are plenty of pretty patterns with other denominators.

For example $1/81 = .012345679012345679012\ldots$ and if you take any number of the form

$$n/91 = .abcdefabcdef\ldots,$$

then the decimal obtained by reversing the six digits of the period, $.fedcbafedcba\ldots$, is some other number of the form $n'/91$. For instance,

$$\frac{13}{91} = \frac{1}{7} = .\dot{1}4285\dot{7} \quad \text{while} \quad .\dot{7}5824\dot{1} = \frac{69}{91},$$

$$\frac{7}{91} = \frac{1}{13} = .\dot{0}7692\dot{3} \quad \text{while} \quad .\dot{3}2967\dot{0} = \frac{30}{91},$$

$$\frac{1}{91} = .\dot{0}1098\dot{9} \quad \text{while} \quad .\dot{9}8901\dot{0} = \frac{90}{91}.$$

The remaining cycles are $\frac{5}{91} = .\dot{0}5494\dot{5}$ and

$$\frac{2}{91} = .\dot{0}2197\dot{8} \quad \text{and} \quad .\dot{8}7912\dot{0} = \frac{80}{91},$$

$$\frac{4}{91} = .\dot{0}4395\dot{6} \quad \text{and} \quad .\dot{6}5934\dot{0} = \frac{60}{91},$$

$$\frac{2}{13} = .\dot{1}5384\dot{6} \quad \text{and} \quad .\dot{6}4835\dot{1} = \frac{59}{91},$$

$$\frac{24}{91} = .\dot{2}6373\dot{6} \quad \text{and} \quad .\dot{6}3736\dot{2} = \frac{58}{91}.$$

bases, mod p	± 2	± 3	± 4	± 5	± 6	± 7	± 8
+	C	C	0	20C/19	C	C	3C/5
−	C	6C/5	C	C	C	42C/41	3C/5
bases	± 9	± 10	± 11	± 12	± 13	± 14	± 15
+	0	C	C	C	156C/155	C	C
−	C	C	110C/109	6C/5	C	C	94C/95
bases	± 16	± 17	± 18	± 19	± 20	± 21	± 22
+	0	272C/271	C	C	20C/19	204C/205	C
−	C	C	C	342C/341	C	C	C
bases	± 23	± 24	± 25	± 26	± 27	± 28	± 29
+	C	C	0	C	3C/5	C	812C/811
−	506C/505	C	C	C	6C/5	42C/41	C
bases	± 30	± 31	± 32	± 33	± 34	± 35	± 36
+	C	C	15C/19	544C/545	C	C	0
−	C	930C/929	15C/19	C	C	778C/779	C

TABLE 6.4 *Proportions of long primes, conjectured by Artin and Lehmer.*

PYTHAGOREAN FRACTIONS

As we'll see in Chapter 7, geometry problems don't always lead to rational numbers, but there are some interesting cases where they do. What shapes of rectangles have whole number sides and diagonals? We'll call the corresponding fractions b/l the **Pythagorean fractions** (Figure 6.7). Those of our readers who remember the 3, 4, 5 triangle will see that 3/4 and 4/3 are Pythagorean fractions.

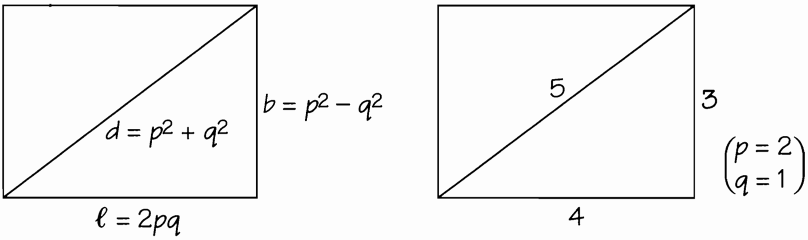


FIGURE 6.7 *Pythagorean rectangles.*

The famous Greek arithmetician Diophantus of Alexandria showed that the Pythagorean fractions are precisely the numbers

$$\frac{p^2 - q^2}{2pq}$$

with p and q whole numbers.

Here are all the right-angled triangles with whole number sides and legs <100 (omitting those where b, l, and d have a common factor):

1. 3,4,5	**7.** 12,35,37	**13.** 33,56,65
2. 5,12,13	**8.** 13,84,85	**14.** 36,77,85
3. 7,24,25	**9.** 16,63,65	**15.** 39,80,89
4. 8,15,17	**10.** 20,21,29	**16.** 48,55,73
5. 9,40,41	**11.** 20,99,101	**17.** 60,91,109
6. 11,60,61	**12.** 28,45,53	**18.** 65,72,97

FIGURE 6.8 *Primitive Pythagorean rectangles with sides <100.*

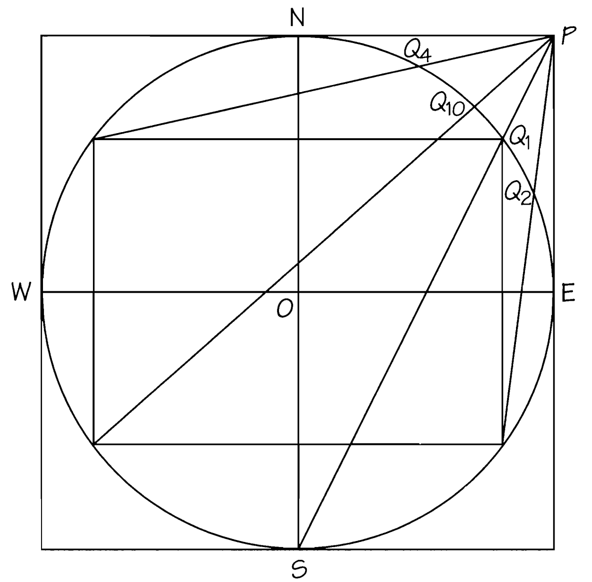


FIGURE 6.9 *Vogeler's Pythagorean fraction finder.*

Roger Vogeler has shown that the process sketched in Figure 6.9 gives each Pythagorean fraction x/y just once. A circle is inscribed in a square. Join a corner of the square, P, to where it touches the circle at S or W. The other place, Q_1, where this line cuts the circle is one corner of a (3,4,5)-shaped rectangle. If we join the other corners to P, we find the points Q_2, Q_4, and Q_{10}, on the circle which are the corners of (5,12,13)-, (8,15,17)-, and (20,21,29)-shaped rectangles, the second, fourth, and tenth of those in Figure 6.8. If you join P to the corners of each new rectangle, you will discover further rectangles, and so on forever.

A Babylonian Table of Pythagorean Fractions

These "Pythagorean fractions" were known long before Pythagoras. Figure 6.10a shows a Babylonian clay tablet that was thought to record some commercial transactions before Otto Neugebauer pointed out its connection with Pythagorean triples.

We taught you to read Babylonian cuneiform in Chapter 1, so you'll be able to check the translation given in Figure 6.10b. The tablet is obviously broken. There are various speculations as to the exact scope (and use!) of the original table, but our own reconstruction is given in Figure 6.11. According to this, the full table could have recorded all of the shapes

$$\text{length} : \text{breadth} : \text{diagonal}$$

proportional to

$$2pq : p^2 - q^2 : p^2 + q^2$$

with p and q "regular" numbers (i.e., that divide powers of 60) and q less than 60.

To close the chapter, we'll show you how to find fractions that are good approximations to other numbers.

(a)

(b)

15	119	169	1
58 14 50 6 15	3367	4825*	2
41 15 33 45	4601	6649	3
10 29 32 52 16	12,709	18,541	4
48 54 1 40	65	97	5
47 6 41 40	319	481	6
43 11 56 28 26 40	2291	3541	7
41 33 45* 14* 3 45	799	1249	8
38 33 36 36	481*	769	9
35 10 2 28 27 24 26 40	4961	8161	10
33 45	(3/4)	(1 1/4)	11
29 21 54 2 15	1679	2929	12
27 0 3 45	161*	289	13
25 48 51 35 6 40	1771	3229	14
23 13 46 40	28*	53	15

FIGURE 6.10 *Plimpton 322 (reproduced with permission of curators of George A. Plimpton Collection, Rare Book and Manuscript Library, Columbia University).*

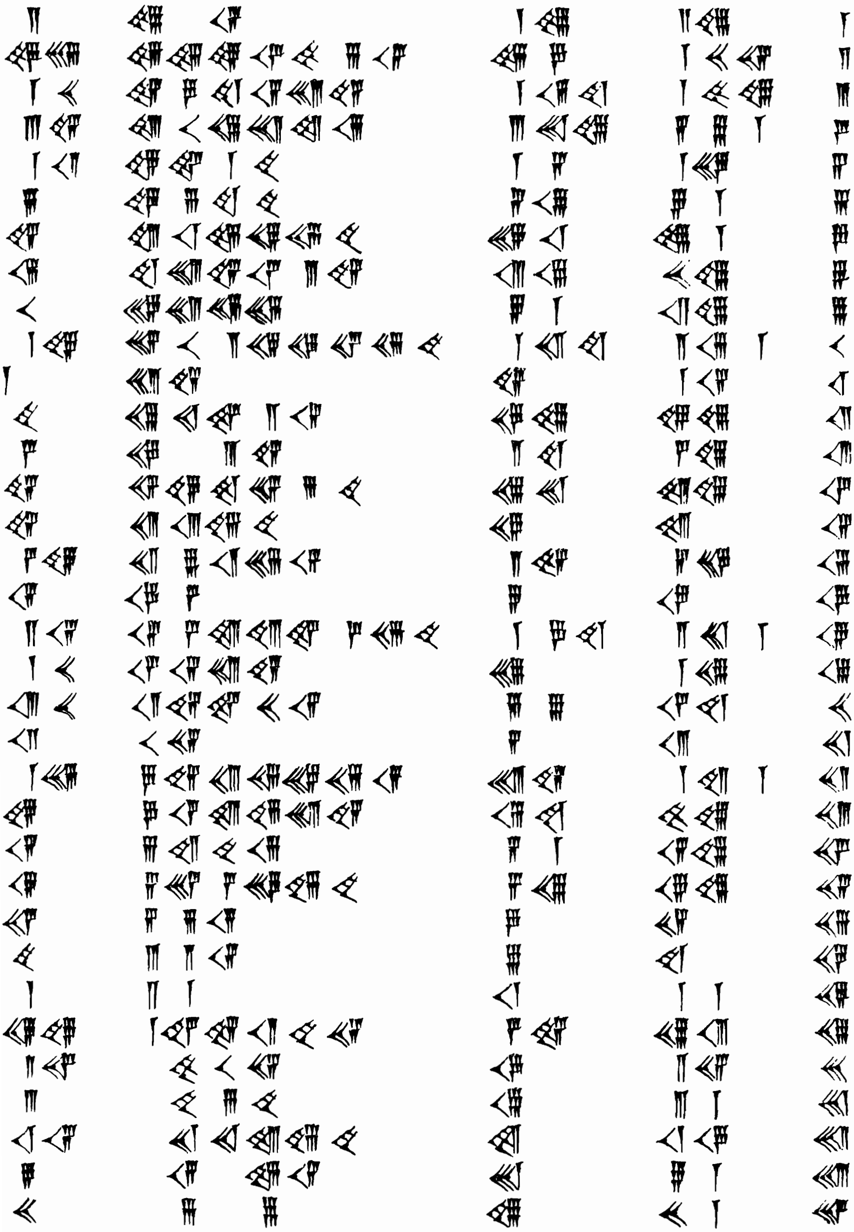


FIGURE 6.11 *Plimpton 322 restored and completed? (See box on p. 176.)*

We have a new suggestion for the error on line 2. We think that 3,12 + 1 (= 193) was written in error for 1,20,25 (= 25 × 193). We also believe that on line 11, 45 and 1,15 should be thought of as $\frac{3}{4}$ and $1\frac{1}{4}$, respectively, since the 3,4,5 case was regarded as basic and the length 4 was taken as the unit.

Other errors are

on line 8 59 (= 45 + 14) was written in error for 45,14,
on line 9 9,1 (= 541) was written in error for 8,1 (= 481),
on line 13 7,12,1 (= 161^2) was written in error for 2,41 (= 161),
on line 15 56 and 53 were written in place of 28 and 53 (or 56 and 106).

CONTINUED FRACTIONS

In our struggle to understand the world, we often find ourselves replacing the messy numbers around us by rough approximations to them, saying that an inch is $2\frac{1}{2}$ centimeters, or that a liter is $2\frac{1}{5}$ pints, or $1\frac{3}{4}$ pints, depending on which side of the Atlantic we are.

The ancients were faced with several such questions in astronomical contexts and found, for instance, but not with quite that accuracy, that the seasons recur every 365.242199 days: we'll call it a **year**, while the period of the moon's phases is 29.530588 days: we'll call it a **month**. In fact the numbers 365.242199 and 29.530588 decrease each century by 1 or 2 in the last decimal place because tidal friction is slowing the earth's rotation and making the day longer. However, their ratio, which is all we wish to use, is sensibly constant.

The Athenian astronomer Meton (ca. 432 B.C.) discovered that 235 months were very nearly equal to 19 years. This is the **Metonic cycle**, still used to determine the Jewish calendar and also the date of Easter.

Figure 6.12 gives the error in comparing various numbers of months with various numbers of years. The first two lines merely record the number of days in the year and month. Each subsequent line is obtained by adding a certain multiple of its predecessor to the one before that, the multiplier being chosen to be the first one that reduces the error to a new minimum. For example, we get the sixth line by adding twice the fifth to the fourth, since this gives a new record low error of

$$7.78\ldots - 2(3.09\ldots) = 1.59\ldots$$

(had we only added *one* copy, we would have had 7.78 . . . − 3.09 . . . = 4.68 . . . , *not* a new record).

So we see that the error of 0.0864 of a day (2 hours 4.4 minutes) in Meton's approximation is not bettered until we compare 4131 months with 334 years. The successive fractions

$$\frac{12}{1} \quad \frac{25}{2} \quad \frac{37}{3} \quad \frac{99}{8} \quad \frac{136}{11} \quad \frac{235}{19} \quad \frac{4131}{334}$$

may be written

$$12,\ 12+\frac{1}{2},\quad 12+\cfrac{1}{2+\cfrac{1}{1}},\quad 12+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2}}},\quad 12+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1}}}},$$

$$12+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{1}}}}},\quad 12+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{2+\cfrac{1}{1+\cfrac{1}{\cfrac{1+}{17}}}}}}.$$

	error	multiplier
0 months = 1 year	−365.242199 days	
1 month = 0 year	+29.530588 days	(12)
12 months = 1 year	−10.875143 days	(2)
25 months = 2 years	+7.780302 days	(1)
37 months = 3 years	−3.094841 days	(2)
99 months = 8 years	+1.590620 days	(1)
136 months = 11 years	−1.504221 days	(1)
235 months = 19 years	+0.086399 days	(17)
4131 months = 334 years	−0.035438 days	(2?)

FIGURE 6.12 *Better and better approximations for the month-to-year ratio.*

Expressions like these are commonly called **continued fractions** and written in an abbreviated notation:

$$12 + \frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{17+}\,\frac{1}{2}.$$

The numbers 12, 2, 1, 2, 1, 1, 17, 2 are called **partial quotients**, and the fractions $^{12}/_{1}$, $^{25}/_{2}$, $^{37}/_{3}$, . . . **convergents**.

Each rational number corresponds to precisely two such continued fractions, one in which the last partial quotient is 1, and one in which it isn't:

$$\frac{4131}{334} = 12 + \frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{16+}\,\frac{1}{1}$$

$$= 12 + \frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{17}.$$

It is obvious from the way we found them that the successive fractions

$$\frac{p}{q} = \frac{12}{1} \quad \frac{25}{2} \quad \frac{37}{3} \quad \frac{99}{8} \quad \frac{136}{11} \quad \frac{235}{19} \quad \frac{4131}{334}$$

are those that achieve new record minima for the difference between p months and q years.

DESIGNING GEAR TRAINS

Engineers use this in designing trains of gears. Figure 6.13 shows a simple gear train that might be used in a planetarium to simulate the relative motion of the sun and moon around the earth. It approximates 12.368267 . . . by $^{235}/_{19}$.

We meet continued fractions again in our next chapter, which starts with another piece of Babylonian mathematics.

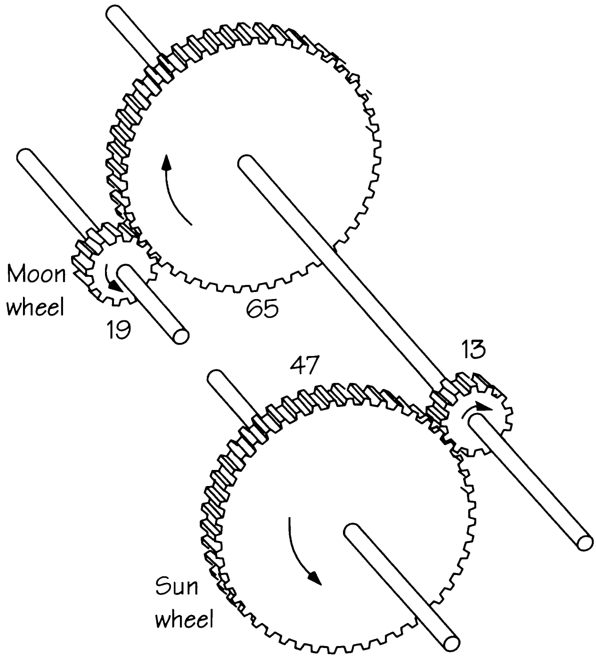


FIGURE 6.13 *Gear train simulating relative motion of sun and moon.*

REFERENCES

David Aldous and Persi Diaconis. Shuffling cards and stopping times. *Amer. Math. Monthly,* 93 (1986) no. 5, 333–348. (Has an extensive bibliography.)

Persi Diaconis, R.L. Graham and William M. Kantor. The mathematics of perfect shuffles. *Adv. Appl. Math.* 4 (1983): 175–196.

C. Hooley. On Artin's conjecture. *J. reine angew. Math.* **225** (1967): 209–220; *M.R.* 34 #7445.

C.G.J. Jacobi. Canon Arithmeticus, Berlin, 1839; Nach Berechnungen von Wilhelm Patz, in verbesserter und erweiterter Form neu herausgegeben von H. Brandt, Akademie-Verlag, Berlin, 1956.

Otto Neugebauer. *The Exact Sciences in Antiquity.* Princeton Univ. Press, Princeton, NJ, 1952.

A.E. Western and J.C.P. Miller. Indices and primitive roots. *Royal Society Math. Table,* 9, Cambridge Univ. Press, 1968.

Samuel Yates. *Prime Period Lengths* (published privately).

7

Geometric Problems and Algebraic Numbers

Historically, geometry has often been a source of new numbers (Figure 7.1). Figure 7.1 is our redrawing of tablet number 7289 in the Yale Babylonian Collection.

You must be getting quite good at reading Babylonic cuneiform, so you can probably make out the base 60 numbers on the diagonal of the square (𒁹 = 1, 𒌋 = 10). If we use the notation usually asso-

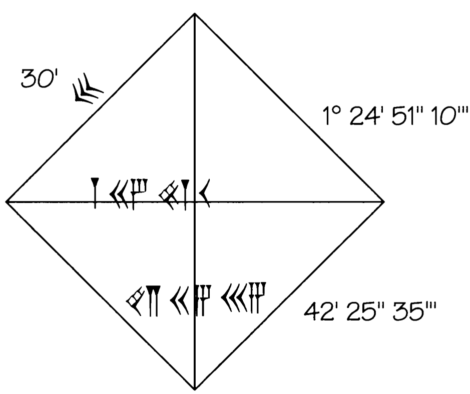


FIGURE 7.1 *YBC 7289.*

ciated with another bequest of the Babylonians, it is 1°24′51″10‴, meaning

$$1 + \frac{24}{60} + \frac{51}{60^2} + \frac{10}{60^3} = 1.41421296296296\ldots.$$

The tablet multiplies this value of $\sqrt{2}$ by the side (30′ = ½°) of the square, to obtain its diagonal.

How did the Babylonians find this remarkable approximation? Certainly not by measuring! If the side of the square is a mile, then, mixing "metraphors," the error in measuring the diagonal would be less than a millimeter! (When was man first able to measure with this accuracy? Probably only in this century.) Plainly, they used calculation.

However this value of $\sqrt{2}$ was calculated, it seems certain that the Babylonians knew that the calculation was not complete (although it is a very good place to stop the calculation, because the next sexagesimal "place" is quite small).

Does the calculation ever stop? Or become periodic? One of the most astounding discoveries attributed to Pythagoras implies that both answers are NO.

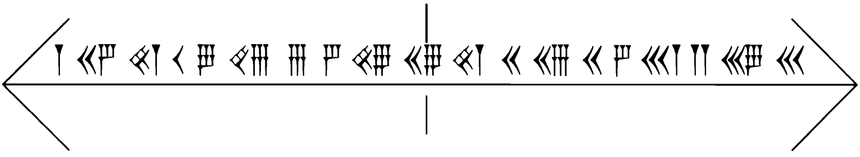


FIGURE 7.2 *The Babylonian calculation continued.*

The sequence of sexagesimal "digits" in Figure 7.2 should continue indefinitely, and neither is $\sqrt{2}$ a recurring decimal:

$$\sqrt{2} = 1.4142135623\ 7309504880\ 1688724209\ 6980785696\ldots.$$

It's surprising that such a deep fact was established so early in human history.

Actually, we don't really know that it was Pythagoras himself who discovered this fact, since the Pythagoreans were a mystical brotherhood who habitually ascribed all their discoveries to the master.

(There may also have been, as we hinted above, some Babylonian influence.) We do know that the Pythagoreans and the later Greek geometers were rightly very impressed by it. The standard story related by Proclus is that they sacrificed a hundred oxen in its honor.

The Pythagorean result is

$\sqrt{2}$ cannot be expressed as a fraction $\frac{a}{b}$ where a, b are whole numbers.

Pythagoras' Big Theorem

The discussion in Chapter 4 then implies that its decimal expansion does not terminate or become periodic. The same holds in any other scale of notation, for example, the Babylonian sexagesimal scale.

In the standard mathematical language we refer to ratios or **rational numbers**, rather than fractions, and we assert that

$\sqrt{2}$ ***is an irrational number***,

historically the first known example of such a number's existence.

We can see the irrationality of $\sqrt{2}$ by folding a square of paper (Figure 7.3)! From Pythagoras's theorem the length of the diagonal is

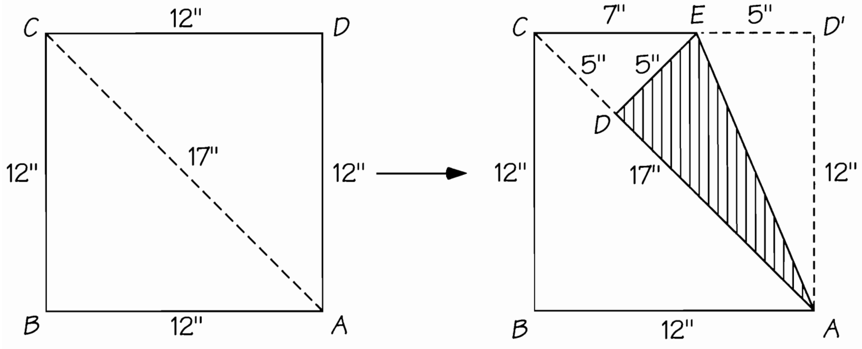


FIGURE 7.3 *Unfolding the mystery of* $\sqrt{2}$.

$\sqrt{2}$ times the length of the side. If the length of the side is 12 inches, the length of the diagonal is

$$12\sqrt{2} = 16.97056 \ldots \text{ inches},$$

or nearly 17 inches. So $\sqrt{2}$ is very nearly $^{17}/_{12}$. However, we shall show that $\sqrt{2}$ cannot be equal to any such fraction. If it *had* been $^{17}/_{12}$, the diagonal of a 12-inch square would be exactly 17 inches. In Figure 7.3 fold the side AD from its original position AD' onto the diagonal AC. Then CDE is the same shape as ABC (so that $CD = DE$) and still has whole number sides, namely

$$CD = 5, \quad \text{obtained as} \quad 17 - 12,$$

$$CE = 7, \quad \text{obtained as} \quad 12 - 5,$$

so the simplest fraction for $\sqrt{2}$ couldn't be $^{17}/_{12}$, for it would already have been $^{7}/_{5}$. The same argument, starting with a square of diagonal a and side b, shows that $\sqrt{2}$ can't equal any rational number a/b in its lowest terms.

Similar proofs in Figures 7.4 and 7.5 show that the ratios (diagonal/side) for a regular pentagon and (short diagonal/side) for a regular hexagon are irrational numbers. In Figure 7.4, BEA is similar to ABD, but $AB = AE = d$ and $BE = c - d$. In Figure 7.5, CDE is similar to ABC, but $CD = DE = e - f$ and $CE = 2f - (e - f) = 3f - e$.

The ratio (diagonal/side) for a regular pentagon is called the **golden number** $\tau = 1.61803398 \ldots$. Its exact value is $(1 + \sqrt{5})/2$ and its wonderful properties have intrigued many people from Greek times through today. Fra Luca Pacioli wrote an entire book on the subject, called *De Divina Proportione*, printed in 1509. Many people think that the **golden rectangle**, the ratio of whose sides is the golden number, is the most aesthetically pleasing.

For the hexagon, the ratio e/f equals $\sqrt{3} = 1.732050807 \ldots$ in Figure 7.5. Our geometrical proof that $\sqrt{3}$ is irrational can be generalized arithmetically.

Of course, some square roots of whole numbers *are* rational. For instance, $\sqrt{9} = 3$. The correct assertion is that if the square root of a whole number is not itself a whole number, then it is not a fraction either. There's an easy arithmetical proof. If N is a whole number for which $\sqrt{N}$ isn't a whole number, suppose B/A is the simplest fraction

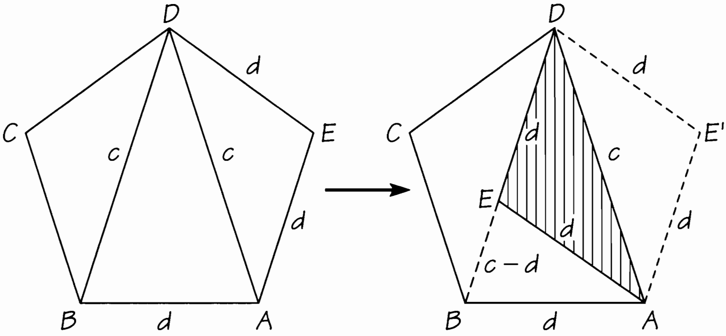


FIGURE 7.4 *Folding a regular pentagon.*

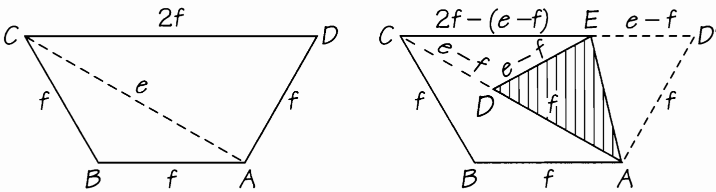


FIGURE 7.5 *Folding half a regular hexagon.*

for $\sqrt{N}$. Then B/A would also equal NA/B. The fractional parts of B/A and NA/B have the form a/A and b/B, where a,b are positive numbers smaller than A,B. But if two numbers are equal, their fractional parts are also equal:

$$\frac{a}{A} = \frac{b}{B}$$

and so

$$\frac{b}{a} = \frac{B}{A} = \sqrt{N}.$$

This gives a simpler form for $\sqrt{N}$, contrary to our assumption.

For instance, if $\sqrt{2}$ were $17/12$ it would also be $24/17$, and the fractional parts of these would be equal.

$$\frac{5}{12} \text{ would equal } \frac{7}{17}, \text{ and } \frac{7}{5} \text{ would equal } \frac{17}{12} = \sqrt{2},$$

giving a simpler fraction for $\sqrt{2}$.

So if you thought that all numbers were fractions, you were wrong! There are lots of irrational numbers: indeed, in Chapter 9 we'll see that there are many, many more irrational numbers than rational ones.

CONTINUED FRACTIONS FOR IRRATIONAL NUMBERS

Whether a number is rational or not, you can find a continued fraction for it in the way we did in the last chapter. This process ends only for a rational number: the irrational numbers lead to infinite continued fractions.

Lagrange proved that continued fractions are *periodic* just for algebraic numbers of degree 2; for example:

$$\sqrt{2} = 1 + \frac{1}{2+}\,\frac{1}{2+}\,\frac{1}{2+}\,\frac{1}{2+}\,\frac{1}{2+}\,\frac{1}{2+}\cdots,$$

$$\tau = \frac{1+\sqrt{5}}{2} = 1 + \frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\cdots,$$

$$\sqrt{3} = 1 + \frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\cdots,$$

$$\sqrt{5} = 2 + \frac{1}{4+}\,\frac{1}{4+}\,\frac{1}{4+}\,\frac{1}{4+}\,\frac{1}{4+}\,\frac{1}{4+}\cdots.$$

Very few other numbers display any recognizable pattern, but Napier's number does:

$$e = 2 + \frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{6+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{8+}\cdots,$$

and so does its cube. However,

$$\pi = 3 + \frac{1}{7+}\,\frac{1}{15+}\,\frac{1}{1+}\,\frac{1}{292+}\cdots$$

is quite chaotic.

LAGRANGE NUMBERS, MARKOV NUMBERS AND FREIMAN'S NUMBER

Notice that you get a good approximation to a number if you chop off the tail of its continued fraction just before a large partial quotient. For example, just before 15 or just before 292 in π,

$$\pi \approx 3 + \frac{1}{7} = \frac{22}{7} \text{ or } \pi \approx 3 + \frac{1}{7+}\,\frac{1}{15+}\,\frac{1}{1} = \frac{355}{113}$$

Lagrange proved a theorem that says that you can find infinitely many approximations $\frac{p}{q}$ to any real number r that satisfy

$$\left| r - \frac{p}{q} \right| \leq \frac{1}{\sqrt{5}\, q^2}$$

The worst numbers to approximate are associated with the golden number τ whose partial quotients are, as we've just seen, all 1. If you rule out such numbers, then Lagrange improves his result to

$$\left| r - \frac{p}{q} \right| \leq \frac{1}{\sqrt{8}\, q^2}$$

The next most difficult numbers are associated with $\sqrt{2}$ and if you leave these out as well, then the theorem improves to

$$\left| r - \frac{p}{q} \right| \leq \frac{1}{\sqrt{221/25}\, q^2}$$

and so on. These constants, $\sqrt{5}$, $\sqrt{8}$, and $\sqrt{221/25}$, . . . , are the **Lagrange numbers.** They are of shape

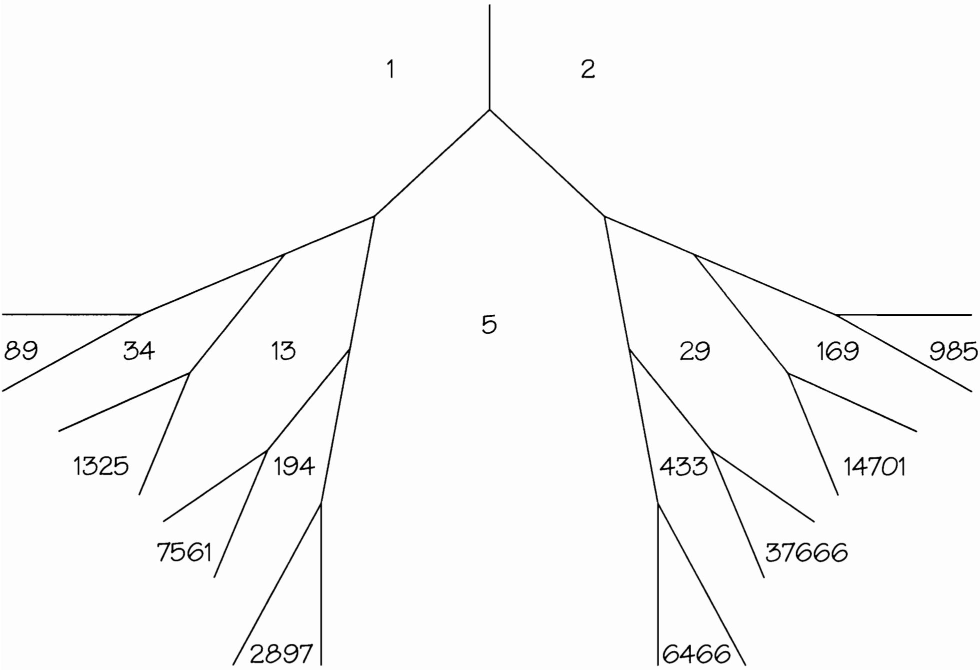
$$\sqrt{9 - \frac{4}{m^2}}$$

where m is a **Markov number.** The Markov numbers occur in the solutions of the diophantine equation

$$x^2 + y^2 + z^2 = 3xyz$$

It is clear that $(x,y,z) = (1,1,1)$ is a solution, and so is $(1,1,2)$. As the equation is quadratic in x,y,z, each different number in a solution leads to a new number. We can arrange the subsequent solutions

(1,2,5), (1,5,13), (2,5,29), . . . , by numbering the regions round an infinite trivalent tree; the three numbers round a vertex form a solution:



It is a famous unsolved problem as to whether there are two different regions with the same label. The regions adjacent to the 1-region have alternate Fibonacci labels, 1, 2, 5, 13, 34, 89, . . . , and those adjacent to the 2-region have alternate Pell labels, 1, 5, 29, 169, 985, ... Notice that the sum of the labels at the two ends of an edge is equal to three times the product of the labels on either side of that edge: $13 + 89 = 3 \times 1 \times 34$, $5 + 7561 = 3 \times 13 \times 194$, and $29 + 6466 = 3 \times 5 \times 433$, for example.

The Lagrange numbers, that can be the 'degree of approximability' of some number, form the **Lagrange spectrum.** The only numbers in the spectrum that are less than 3 are the ones formed from the Markov numbers, but there are many more. In fact G. A. Freiman has shown that the last gap ends at **Freiman's number:**

$$\frac{2221564096 + 283748\sqrt{462}}{491993569}$$

which is approximately 4.5278295661 6087914088 2695988070 4696469298 3363276972 8374065061 . . . and all real numbers larger than this are in the **Markov spectrum.**

ALGEBRAIC NUMBERS

The three irrational numbers that came from the ratios diagonal : side for the square, pentagon, and hexagon satisfy algebraic equations. Thus

$x = a/b$ for the square satisfies $x^2 = 2$ or $x^2-2 = 0$;

$y = c/d$ for the pentagon satisfies $y^2 = y+1$ or $y^2-y-1 = 0$;

$z = e/f$ for the hexagon satisfies $z^2 = 3$ or $z^2-3 = 0$,

so they are called **algebraic numbers**.

In fact any number x that is a solution of an equation

$$ax^n + bx^{n-1} + cx^{n-2} + \cdots + kx + l = 0 \qquad (a \neq 0)$$

where $a, b, c, \ldots, k, l$ are integers, is called an algebraic number, whose **degree** is the smallest possible number n in such an equation. The equation of smallest degree for an algebraic number is essentially unique. Other equations that x satisfies are all found by multiplying this simplest one by various factors.

The simplest such equation satisfied by an algebraic number α may also be satisfied by some other numbers, $\beta, \gamma, \ldots$. In fact, it will always have n distinct solutions, $\alpha, \beta, \gamma, \ldots$ (although some of these may be complex; see the next chapter):

$$ax^n+bx^{n-1}+cx^{n-2}+\cdots+kx+l = a(x-\alpha)(x-\beta)(x-\gamma)\cdots.$$

The numbers $\beta, \gamma, \ldots$ are called the **conjugates** of α. If α satisfies *any* other algebraic equation

$$Ax^N + Bx^{N-1} + Cx^{N-2} + \cdots + Kx + L = 0$$

(where $A, B, C, \ldots, K, L$ are integers), then its conjugates will satisfy this equation (because it is a multiple of the simplest one).

For instance, the simplest equation for the golden number

$$\tau = (1 + \sqrt{5})/2$$

is $x^2 - x - 1 = 0$, which has another root, $\sigma = (1 - \sqrt{5})/2$. Now τ also satisfies the equation $x^5 = 5x + 3$, so its conjugate, σ, does also.

The irrational numbers we have found so far are all **algebraic numbers of degree two** (or **quadratic surds**). Our next two examples are of degree three.

THREE GREEK PROBLEMS

Delos is an island in the Greek archipelago, once famous as the reputed birthplace of Apollo and Artemis. The story is told that when a plague was raging at Athens, the inhabitants sent an emissary to ask the oracle of Apollo at Delos what to do. The oracle replied that the plague would cease if the altar to Apollo were exactly doubled in size.

The altar was a geometrical cube of edge length one cubit. So the Athenians hastily prepared a new altar that was a cube of edge length two cubits. But the plague continued, since, as the oracle explained, the new altar was eight times the size of the old one. They tried again, putting two one-cubit cubes side by side, but the plague raged on, because the altar was no longer a cube.

It seems that Apollo would only be pleased by a cube whose edge length was $\sqrt[3]{2}$. The Greek geometers knew ruler and compass constructions for lines of length $\sqrt{2}$, $\sqrt{3}$, etc., and also for $\sqrt[4]{2}$, but none of them could provide such a construction for $\sqrt[3]{2}$. We now know that there isn't one.

Another problem that teased the Greeks, but doesn't seem to have an accompanying legend, is the trisection of the angle. The Greek geometers knew how to bisect a line (Figure 7.6(a)), also how to trisect it (Figure 7.6(b)), or divide it into any number of equal parts. They also knew how to bisect an angle (Figure 7.6(c)), and by repeating this they could quadrisect it, or further divide it into 2^n equal parts, but they couldn't find a ruler and compare construction for *trisecting* an arbitrary angle. We now know that there isn't one.

The third and most notorious of the problems that the Greeks

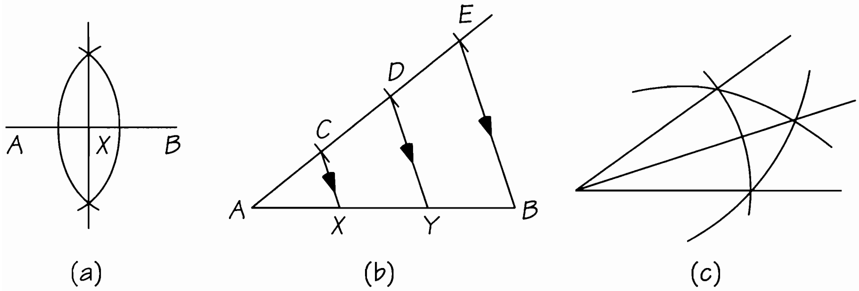


FIGURE 7.6 *(a) and (c) are easy. To trisect* AB, *mark off equal lengths* AC, CD, DE *on another line, and draw parallels* CX, DY *to* EB.

couldn't solve is "squaring the circle": given a circle, construct a square with the same area. "Squaring the circle" was used as a phrase for an impossible problem long before it was actually proved to be impossible by F. Lindemann in 1882. An equivalent problem is to construct a line whose length is π times that of a given one.

In the first two of the Greek problems, the numbers involved are cubic algebraic numbers, but in this third one, π is transcendental (see Chapter 9).

Before we discuss these problems further, we'd better precisely lay down the rules.

RULER AND COMPASS CONSTRUCTIONS

Euclid seems to have regarded geometry as some sort of game. His rules for this game are: (i) given two points, you can use your ruler to join them and extend the line as far as you like, or (ii) you can use your compass to draw a circle, centered at one of them and passing through the other. Further points, found as the places where straight lines and circles meet, may be used to find other points, and so on. Start from two points, O, X, distance 1 apart, what other points of the plane can we reach with these rules? To answer this we use coordinates. Take $O = (0, 0)$, $X = (1, 0)$; Figure 7.7 shows how to reach the points $A = (6, 0)$ and $B = (2\frac{1}{2}, 0)$.

It's not hard to show that you can get to any point like $(\frac{1}{27}, -5\frac{1}{2})$ whose coordinates are both *rational* numbers. But Euclid's first con-

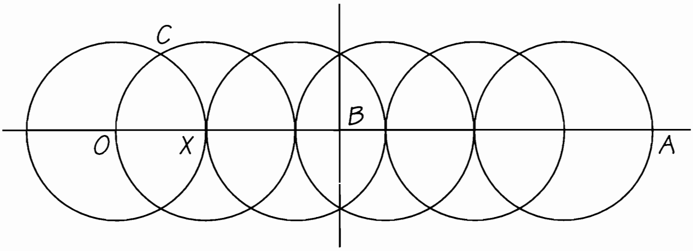


FIGURE 7.7 *Finding integer and fractional points.*

struction, an equilateral triangle, involves ***irrational*** coordinates. In Figure 7.7 the point C has coordinates $(\frac{1}{2}, \frac{1}{2}\sqrt{3})$.

HOW GEOMETRIC PROBLEMS LEAD TO ALGEBRAIC NUMBERS

Most of the numbers that arise in geometric problems are algebraic numbers. Why is this? Pythagoras's theorem shows that square roots arise. You'll note that to find the point where two straight lines meet, you solve two *linear* equations, but the *two* points where two circles meet, or where a circle meets a straight line, are obtained by solving a *quadratic* equation.

Working in the other direction, we can see that any number that's the root of a quadratic equation can be found geometrically. Here's one way to do this. To find the roots x_1, x_2 of the typical quadratic equation $x^2 - ax + b = 0$, we construct the points Y and Z with coordinates $(0, 1)$ and (a, b) (Figure 7.8). Then the circle with diameter YZ cuts the horizontal axis in X_1, X_2 with coordinates $(x_1, 0)$, $(x_2, 0)$. You can therefore give a Euclidean construction for any number that is obtained from a chain of quadratic equations.

We call numbers obtained by repeatedly solving quadratic equations **Euclidean numbers**.

For instance,

$$q = \frac{1}{8}\left\{-1 + \sqrt{17} + \sqrt{34 - 2\sqrt{17}}\right.$$

$$\left. + \sqrt{68 + 12\sqrt{17} - 16\sqrt{34 + 2\sqrt{17}} + 2(-1 + \sqrt{17})\sqrt{34 - 2\sqrt{17}}}\right\}$$

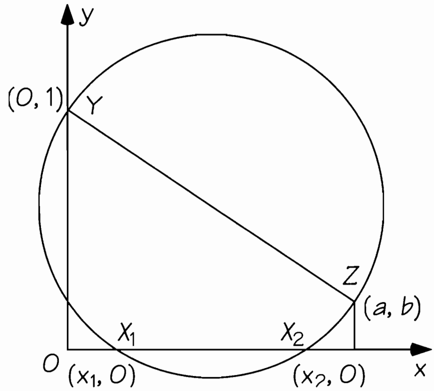


FIGURE 7.8 *Solving a quadratic equation geometrically.*

is a Euclidean number: this is found, after messy algebra, to be the larger root of the quadratic equation $x^2 - ax + b = 0$, where a, b are the larger roots of the equations

$$x^2 - \alpha x - 1 = 0,$$
$$x^2 - \beta x - 1 = 0$$

and α and β are $(-1 \pm \sqrt{17})/2$, the roots of $x^2 + x - 4 = 0$.

This example arises in Gauss's famous construction of the regular polygon with 17 sides (see Chapter 8). The solutions of

$$x^2 - \alpha x - 1 = 0,\ x^2 - \beta x - 1 = 0$$

satisfy $(x^2 - \alpha x - 1)(x^2 - \beta x - 1) = 0$, namely, $x^4 + x^3 - 6x^2 - x + 1 = 0$, which has the four roots

$$a, c = \frac{\alpha \pm \sqrt{\alpha^2 + 4}}{2} = \frac{-1 + \sqrt{17} \pm \sqrt{34 - 2\sqrt{17}}}{4},$$

$$b, d = \frac{\beta \pm \sqrt{\beta^2 + 4}}{2} = \frac{-1 - \sqrt{17} \pm \sqrt{34 + 2\sqrt{17}}}{4}.$$

A point with coordinates (x, y) is geometrically obtainable by ruler and compass according to Euclid's rules only if x and y are both Euclidean numbers. This is what makes those Greek problems impossible. The problem of duplicating the cube requires the **Delian number**, $\sqrt[3]{2}$, but

the Delian number, $\sqrt[3]{2}$, is not a Euclidean number.

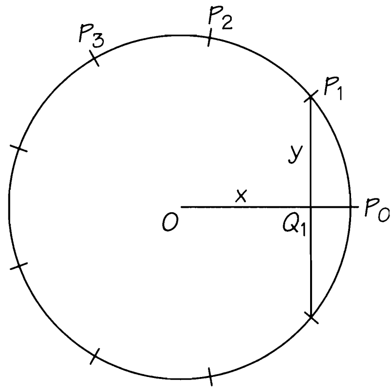


FIGURE 7.9 *An enneagon can be constructed just if the length* x *can.*

If you could trisect an angle with ruler and compasses, you could easily construct, for instance, a regular polygon with nine sides in a circle of radius 2 (Figure 7.9). The point P_1 has coordinates (x, y), where x satisfies the equation $x^3 - 3x + 1 = 0$. But

this x is not a Euclidean number.

Finally, if you could square the circle, you could construct the point with coordinates $(\pi, 0)$, but

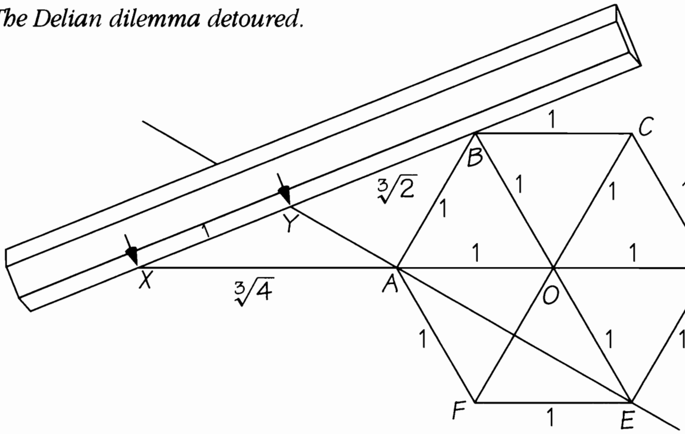
Ludolph's number, π, is not a Euclidean number.

This third problem is completely different from the other two, which led to numbers satisfying the *algebraic* equations $x^3 - 2 = 0$, $x^3 - 3x + 1 = 0$. In fact, the reason we know that the circle can't be squared is that Lindemann proved in 1882 that π, far from being a Euclidean number, is not even algebraic.

BENDING THE RULES

Euclid's rules don't allow you to have any marks on your ruler. If you *do* have marks, X, Y, on your ruler, distance 1 apart, you can use them in a rather cheating way to solve the first two of the famous Greek problems (Figure 7.10). Place your ruler so that it passes through the corner B of a regular hexagon of side 1, with its two

FIGURE 7.10 *The Delian dilemma detoured.*



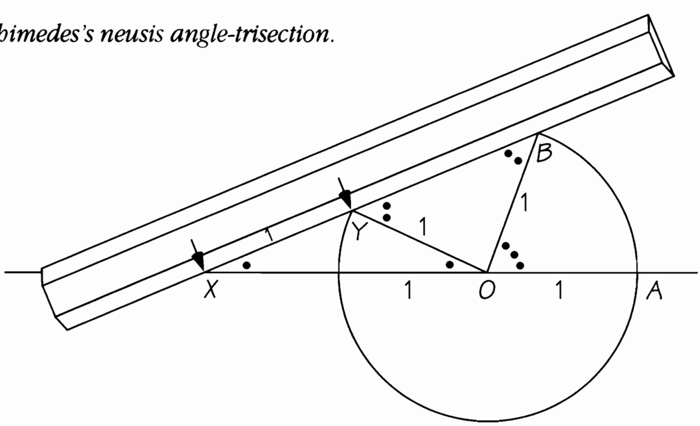
marks, X, Y, on the line AO (X not at O) and AE. Then $BY = \sqrt[3]{2}$ and $AX = \sqrt[3]{4}$.

The Greeks were well aware of such cheating constructions and called them ***neusis*** or "verging" constructions. Perhaps the nicest is Archimedes's neusis trisection of the angle.

To trisect the angle AOB in a circle of radius 1, place your ruler through B with its marks X, Y on OA (but not at O) and on the circle, as in Figure 7.11. Then angle OXY is one-third of angle AOB.

How can we tell from the geometry what degree of algebraic equation will arise? It's possible to make an educated guess just by

FIGURE 7.11 *Archimedes's neusis angle-trisection.*



A 0° 90° 180° 270° 360° 450° 540° 630° 720° 810° 900° 990° 1080° ···
B 0° 30° 60° 90° 120° 150° 180° 210° 240° 270° 300° 330° 360°, ···

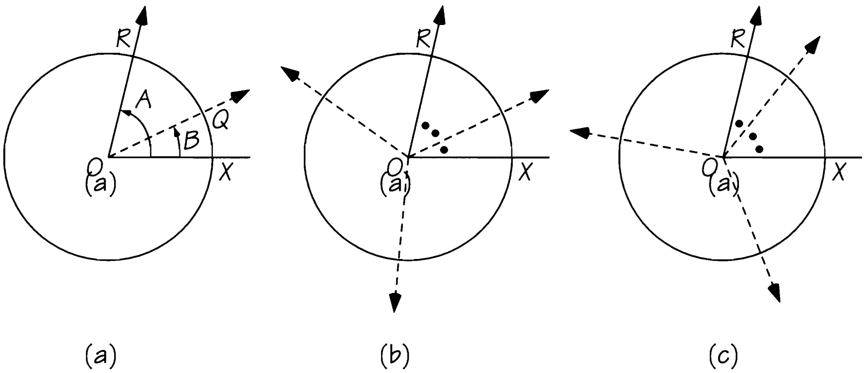


FIGURE 7.12 *The three trisectors of the angle* XÔR, *and the* reverse *trisectors.*

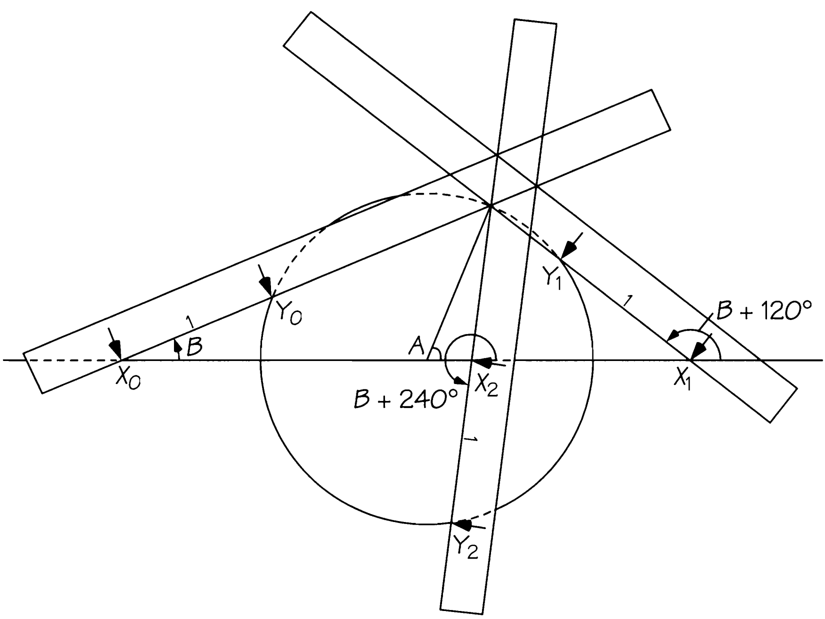


FIGURE 7.13 *Three ways to follow Archimedes's instructions.*

counting the number of solutions properly. For instance, let's imagine trisecting an angle A, with B the trisecting angle. Look at Figure 7.12(a) in which angle XOR is A and angle $XOQ = B = \frac{1}{3}A$, and see what happens as A passes through multiples of 90°.

A	0°	90°	180°	270°	360°	450°	540°	630°	720°	810°	900°	990°	1080°	...
B	0°	90°	60°	90°	120°	150°	180°	210°	240°	270°	300°	330°	360°	...

Since lines are unaffected by complete revolutions (multiples of 360°), if B is a trisecting angle for A, so are $B + 120°$ and $B + 240°$. We see that

> There are really *three* trisectors for any given angle.

Figure 7.12(b) shows the three trisectors (dashed) of angle $X\hat{O}R$. Don't confuse these with the trisectors of the reversed angle $R\hat{O}X$, which are three *other* lines, called the reversed trisectors (Figure 7.12(c)). Figure 7.13 shows how Archimedes's construction finds all three trisecting angles for a given one.

Any algebraic equation corresponding to this geometrical problem won't distinguish between these three angles and so can be expected to lead to a degree 3 equation.

CONSTRUCTING REGULAR POLYGONS

If p is prime, the construction of a regular p-sided polygon in a circle, with one vertex at P_0, is solved as soon as you can construct *any* of the $p - 1$ vertices $P_1, P_2, \ldots, P_{p-1}$, so it should be expected to lead to an equation of degree $p - 1$, and it is natural to suspect that

> For p prime, a regular p-sided polygon can be constructed with ruler and compass just if p is one of the Fermat primes 3, 5, 17, 257, 65537, . . .

since these are the only primes such that $p-1$ is a power of 2.

As we've already remarked, Gauss showed that all of these *are* constructible. Similarly, the construction of a regular polygon with an arbitrary number, n, of sides is easily completed as soon as one can construct one of the vertices P_r for which the fraction r/n is already in its lowest terms (the other corners, P_s, lie on regular polygons with fewer sides). Since the number of such fractions is Euler's totient number, $\phi(n)$ (see Chapter 6), we should expect an equation of this degree. Indeed it happens that

An n-sided regular polygon
is constructible with ruler and compass
only if $\phi(n)$ is a power of 2.

With other instruments, other regular polygons become constructible. For a regular heptagon the algebraic equation has degree $6 = 7-1$, and its roots can be found by solving a cubic, then a quadratic equation. But an angle trisector, if you have one, can be used to solve any cubic equation whose roots are all real. You can therefore use it, alongside your ruler and compass, to construct the regular heptagon (Figure 7.19).

Figures 7.17 to 7.22 give our constructions for regular polygons with 3, 5, 7, 9, 13, and 17 sides. Some of these numbers are not Fermat primes, so sometimes we use an angle-trisector.

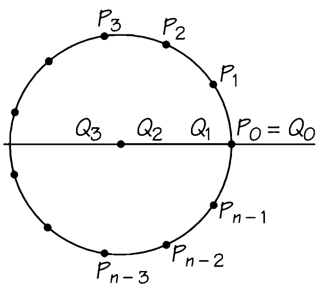


FIGURE 7.14

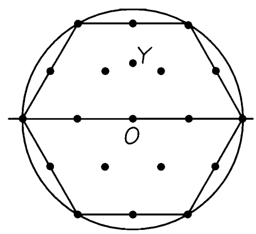


FIGURE 7.15

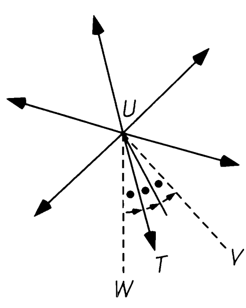


FIGURE 7.16

The polygon will always contain the rightmost point, P_0, of the circle (Figure 7.14), and it's enough to find the points Q_1, Q_2, . . . where the chords P_1P_{n-1}, P_2P_{n-2}, . . . cut the **axis** (diameter through P_0). Many of the initial points of our constructions come from the hexagonal lattice of Figure 7.15, whose shortest distance is half the radius of the circle. We'll regard such points as well enough defined when we say that "they're in the lattice." One other point is Y, the midpoint of the vertical radius. The constructions usually involve the

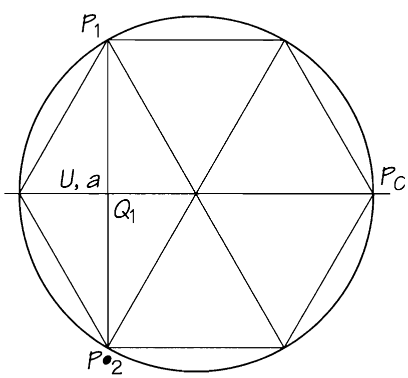


FIGURE 7.17 n = *3*. Q_1, P_1, P_2 *are in the lattice.*

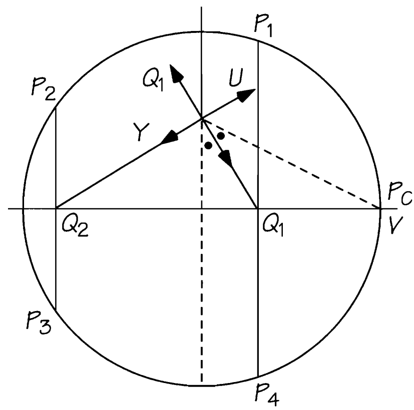


FIGURE 7.18 n = *5*. U = Y; V = P_0 *is in the lattice;* bisectors *cut axis in* Q_1, Q_2.

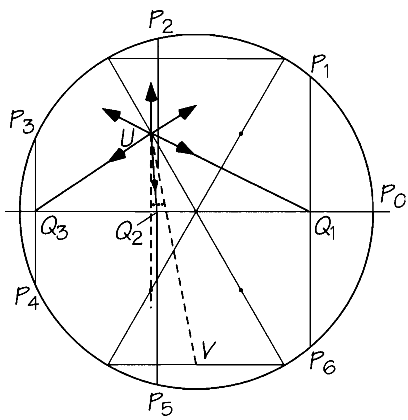


FIGURE 7.19 n = *7*. U, V *are in the lattice;* trisectors *cut axis in* Q_2, Q_3, Q_1.

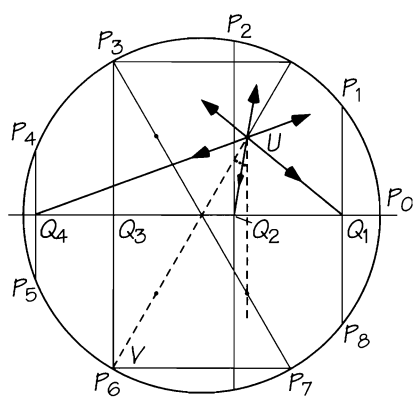


FIGURE 7.20 n = *9*. U, V, = P_6, P_3 *and* Q_3 *are in the lattice;* trisectors *cut axis in* Q_2, Q_1, Q_4.

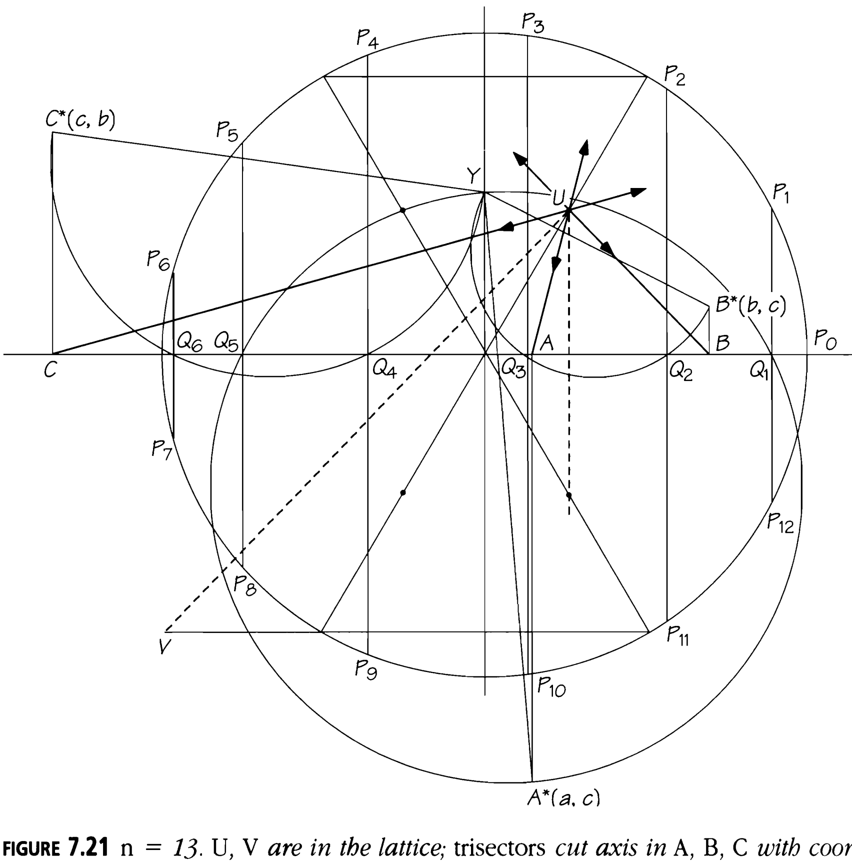


FIGURE 7.21 n = *13*. U, V *are in the lattice;* trisectors *cut axis in* A, B, C *with coordinates* (a, 0), (b, 0), (c, 0). *It is easy to construct points* A*, B*, C* *with coordinates* (a, c), (b, a), (c, b), *by laying off appropriate distances on the vertical lines. Then the three circles with diameters* YA*, YB*, YC* *cut the axis in* Q_2, Q_3; Q_1, Q_5; Q_4, Q_6, *respectively.*

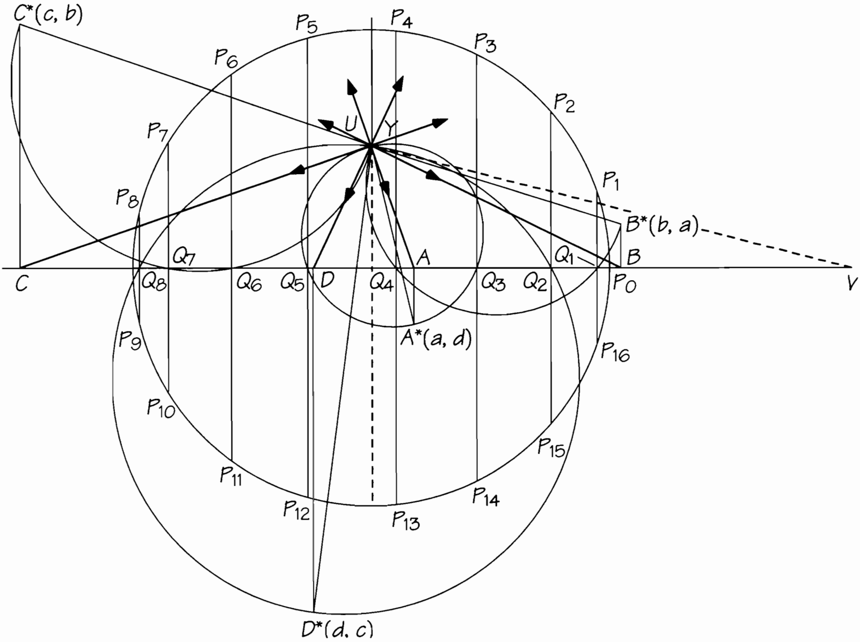


FIGURE 7.22 n = *17.* U = Y*;* V *is in the lattice;* quadrisectors *cut axis in* A, B, C, D *with coordinates* (a, 0), (b, 0), (c, 0), (d, 0), *and the points* A*, B*, C*, D* *have coordinates* (a, d), (b, a), (c, b), (d, c) *much as in* n = *13. The circles with diameters* YA*, YB*, YC*, YD* *cut the axis in* Q_1, Q_4; Q_3, Q_5; Q_2, Q_8; Q_6, Q_7, *respectively.*

bisection, trisection, or quadrisection of an angle. This will always be the angle WUV from the downward vertical, UW, through U, to UV, as in Figure 7.16, which illustrates trisection, $W\hat{U}T = \frac{1}{3}W\hat{U}V$.

We've had enough geometry for a time: let's see where else algebraic numbers occur.

Algebraic Numbers in Arithmetic Problems

Of course, algebraic numbers can also arise from arithmetical problems. We've already met the Fibonacci numbers, f_n, and the Lucas numbers, l_n.

n =	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	...
f_n =	0	1	1	2	3	5	8	13	21	34	55	89	144	233	377	610	987	...
l_n =	2	1	3	4	7	11	18	29	47	76	123	199	322	521	843	1364	2207	...

Each Fibonacci or Lucas number is the sum of the previous two. Can we find any other sequences with this property? Yes.

$$1 \;\; x \;\; x^2 \;\; x^3 \;\; x^4 \;\; x^5 \;\; x^6 \;\; x^7 \;\; x^8 \;\; x^9 \;\ldots$$

will do, if $x^2 = x + 1$, that is, if x is one of the algebraic numbers

$$\tau = \frac{1+\sqrt{5}}{2} \text{ or } \sigma = \frac{1-\sqrt{5}}{2},$$

so we have two rather special solutions:

$$1 \;\; \tau \;\; \tau^2 \;\; \tau^3 \;\; \tau^4 \;\; \tau^5 \;\; \tau^6 \;\; \tau^7 \;\; \tau^8 \;\; \tau^9 \;\ldots,$$
$$1 \;\; \sigma \;\; \sigma^2 \;\; \sigma^3 \;\; \sigma^4 \;\; \sigma^5 \;\; \sigma^6 \;\; \sigma^7 \;\; \sigma^8 \;\; \sigma^9 \;\ldots.$$

By combining multiples of these, we can adjust the first two terms to be anything we like. In particular, we can get

$$f_n = \frac{\tau^n - \sigma^n}{\tau - \sigma} = \frac{1}{\sqrt{5}}\left\{\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right\},$$

$$l_n = \tau^n + \sigma^n = \left\{\left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n\right\}$$

(which don't look at all like whole numbers!). For positive n, the nth Fibonacci number is the nearest whole number to $\tau^n/\sqrt{5}$, and the nth Lucas number is the nearest whole number to τ^n. As n increases, the ratios

$$f_{n+1}/f_n \text{ and } l_{n+1}/l_n \text{ approach } \tau = 1.6180339887498948482\ldots,$$

while l_n/f_n approaches $\sqrt{5} = 2.2360679774997896964\ldots.$

n	f_{n+1}/f_n	l_{n+1}/l_n	l_n/f_n
1	1	3	1
2	2	1.333 . . .	3
3	1.5	1.75	2
4	1.666 . . .	1.5714 . . .	2.333 . . .
5	1.6	1.6363 . . .	2.2
6	1.625	1.6111 . . .	2.25
7	1.6153 . . .	1.6206 . . .	2.2307 . . .
8	1.6190 . . .	1.6170 . . .	2.2380 . . .
9	1.6176 . . .	1.6184 . . .	2.2352 . . .
10	1.6181 . . .	1.6178 . . .	2.2363 . . .

Again, if we look at the whole multiples of $\sqrt{2}$, we can't expect to find a whole number, but we can select successively closer near misses:

$$1 \times \sqrt{2} = 1 + 0.414213\ldots,$$
$$2 \times \sqrt{2} = 3 - 0.171572\ldots,$$

(but not $3\sqrt{2} = 4.2426\ldots$ nor $4\sqrt{2} = 5.6568\ldots$ because the errors $+0.2426\ldots$ and $-0.3431\ldots$ are bigger than our previous best, $-0.1715\ldots$)

$$5 \times \sqrt{2} = 7 + 0.0710672\ldots,$$
$$12 \times \sqrt{2} = 13 - 0.029437\ldots,$$
$$29 \times \sqrt{2} = 41 + 0.012193\ldots,$$
$$70 \times \sqrt{2} = 99 - 0.005050\ldots,$$
$$169 \times \sqrt{2} = 239 + 0.002092\ldots,$$

and obtain the successively better approximations to $\sqrt{2}$,

$$\left(\frac{1}{0}\right) \frac{1}{1} \frac{3}{2} \frac{7}{5} \frac{17}{12} \frac{41}{29} \frac{99}{70} \frac{239}{169} \frac{577}{408}, \ldots.$$

These fractions have many fascinating properties. We saw as we discovered them that they are alternately less and greater than $\sqrt{2}$ itself. The numerators and denominators give the solutions of the equations $x^2 - 2y^2 = \pm 1$ and can be obtained by doubling the previous one and adding the one before that. This is because the continued fraction for $\sqrt{2}$ is

$$1 + \frac{1}{2+} \frac{1}{2+} \frac{1}{2+} \frac{1}{2+} \ldots.$$

The numerator and denominator of the nth approximation are

$$\frac{1}{2}\{(1+\sqrt{2})^n + (1-\sqrt{2})^n\} \text{ and } \frac{1}{2\sqrt{2}}\{(1+\sqrt{2})^n - (1-\sqrt{2})^n\}$$

An equation of the shape $x^2 - dy^2 = 1$ is called a **Pell equation** and can always be solved by a formula of this kind.

We can now find which triangular numbers are squares. The condition that the Nth triangle is also the Mth square is

$$\frac{1}{2} N(N + 1) = M^2$$

or, equivalently,

$$x^2 - 2y^2 = 1$$

in terms of $x = 2N + 1, y = 2M$, so we get the numbers M and N by halving the denominator and rounding off half the numerator of each of the above fractions when the denominator is even:

$$(T_0 = 0^2)\ T_1 = 1^2\ T_8 = 6^2\ T_{49} = 35^2\ T_{288} = 204^2\ T_{1681} = 1189^2 \ldots.$$

The astute reader will have noticed that the subscripts are alternately squares and doubles of squares.

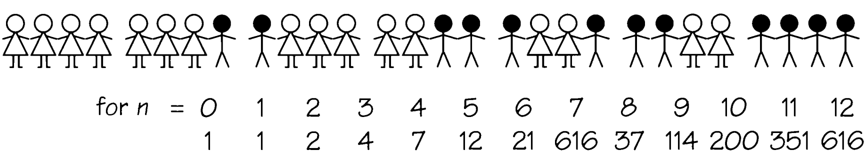
Similar results hold for other sequences in which any given term is found by adding fixed multiples of the *d* previous terms. For instance, if you add the three previous terms you'll get

$$0 \quad 1 \quad 1 \quad 2 \quad 4 \quad 7 \quad 13 \quad 24 \quad 44 \quad 81 \quad 145 \ldots,$$

which can be expressed by powers of the roots of the degree three equation $x^3 = x^2 + x + 1$.

ALGEBRAIC NUMBERS FOR GIRLS AND BOYS

Lots of combinatorial problems lead to sequences of this sort. How many patterns of n children in a row are there if every girl is next to at least one other girl? For example, with four children there are seven patterns:



In general, if $P(n)$ is the number of patterns of n children, then

$$P(n) = 2P(n-1) - P(n-2) + P(n-3),$$

and the ratio $P(n+1)/P(n)$ approaches the degree 3 algebraic number 1.754877666247 . . . , a solution of the equation $x^3 = 2x^2 - x + 1$.

If the girls must appear in groups of at least three, then the numbers of patterns are

$$1 \quad 1 \quad 1 \quad 2 \quad 4 \quad 7 \quad 11 \quad 17 \quad 27 \quad 44 \quad 72 \ldots,$$

and the relevant equation, $x^4 - 2x^3 + x^2 - 1 = 0$, factors into two quadratic equations, and we have four algebraic numbers, each of degree two. We've met two of them, τ and σ, and we'll meet the other two as "Eisenstein units" in the next chapter.

For girls in groups of four or more, the numbers and the equation are of degree 5: $x^5 - 2x^4 + x^3 - 1 = 0$. The important root is $x = 1.528946354519\ldots$.

So we see that easy mathematical problems sometimes lead to algebraic numbers, usually quadratic or cubic. We end with four topics, the last of which is childishly simple but leads to a surprisingly complicated algebraic number.

CALABI'S TRIANGLE

There are obviously three equal largest squares that you can fit into an equilateral triangle (Figure 7.23). Eugenio Calabi discovered the surprising fact that there is a unique *other* shape of triangle in which there are three equally large biggest squares. In Calabi's triangle (Figure 7.24), the ratio of the largest side to the other two is an algebraic number $x = 1.55138752455\ldots$, a solution of $2x^3 - 2x^2 - 3x + 2 = 0$.

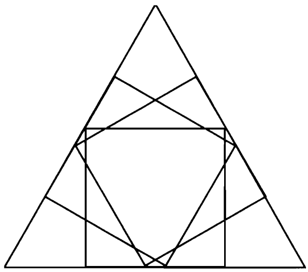


FIGURE 7.23 *Equilateral triangle.*

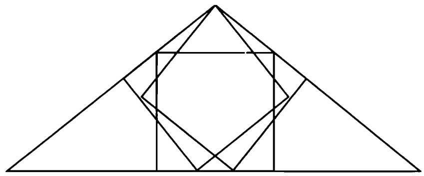


FIGURE 7.24 *Calabi's triangle.*

GRAHAM'S BIGGEST LITTLE HEXAGON

What's the largest area hexagon you can have if no two of the corners are more than unit distance apart? Ron Graham has shown that the answer looks like Figure 7.25, whose area $A = 0.674981\ldots$ is an algebraic number of degree 10 satisfying the equation

$$4096A^{10} - 8192A^{9} - 3008A^{8} - 30{,}848A^{7} + 21{,}056A^{6} + 146{,}496A^{5}$$
$$- 221{,}360A^{4} + 1232A^{3} + 144{,}464A^{2} - 78{,}488A + 11{,}993 = 0$$

and is larger than the area of the regular hexagon, $3\sqrt{3}/8 = 0.649519\ldots.$

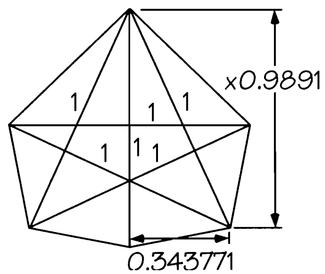


FIGURE 7.25 *Graham's biggest little hexagon.*

PERIODIC POINTS

In Chapter 6 we asked how many shuffles of a particular type are needed to return the cards to their particular places. Mathematicians have also considered shuffling infinite collections. Even for very simple "shuffles," such problems can lead to complicated algebraic numbers. For instance, under the "shuffle" $x \to x^2-2$ of the numbers between -2 and 2, the six particular numbers

$$1.532088886\ldots \to 0.347296355\ldots \to -1.879385241\ldots \to 1.532088886\ldots;$$

$$1.2469796037\ldots \to -0.4450418679\ldots \to -18019377358\ldots \to 1.246979603\ldots$$

each have period 3. They are the roots of

$$x^3 - 3x + 1 = 0 \quad \text{and} \quad x^3 + x^2 - 2x - 1 = 0.$$

Now Sharkovsky has a wonderful theorem that if, under a continuous map ("smooth shuffle") of the numbers in an interval, there is a number having a particular period, then there are other numbers having all the periods that occur later in Sharkovsky's shuffling "sequence." This "sequence" is really an infinite sequence of sequences followed by a backward sequence:

3, 5, 7, 9, . . . , 6, 10, 14, 18, . . . , 12, 20, 28, 36, . . . ,
24, 40, 56, 72, ... 32, 16, 8, 4, 2, 1.

SHARKOVSKY'S SHUFFLING "SEQUENCE."

So, since 3 is the first term of the Sharkovsky shuffling "sequence," we know that there must be a number with any other period that you like under the map that takes x to x^2-2. For example,

$$2 \to 2 \to \ldots \text{ and } -1 \to -1 \to \ldots$$

have period 1, and $(\sqrt{5} - 1)/2 =$

$$0.618033988749\ldots \to -1.618033988746\ldots \to 0.618033988749\ldots$$

has period 2. This is an example of James Yorke's slogan, "Period three implies chaos."

As c decreases from 1 to 0, the "shuffle" that takes x to cx^2-2 drops more and more terms from the start of Sharkovsky's "sequence."

The period of a particular point doubles when a parameter in the function passes through certain critical values. Feigenbaum has found that these "bifurcation points" are usually spaced approximately like geometric series with limiting ratio

$$F_1 = 4.6692016090\ldots.$$

Almost certainly Feigenbaum's number is not an algebraic number, but a transcendental one, such as we will study in the next chapter.

THE LOOK AND SAY SEQUENCE

About how many digits has the nth term in the sequence

1 11 21 1211 111221 312211 13112221 1113213211 31131211131221 . . . ?

Have you guessed the general rule yet? The first term consists of one "one," so the second is "one one." This consists of two "ones," so the third is "two one." This in turn is made from one "two," and one "one" and so gives "one two one one," and so on.

It can be shown that the number of digits in the nth term is roughly proportional to

$$(1.303577269034296391257099112152551890730702504659 4\ldots)^n,$$

where the simplest algebraic equation defining the number in parentheses is

$$\begin{aligned}
&x^{71} - x^{69} - 2x^{68} - x^{67} + 2x^{66} + 2x^{65} + x^{64} - x^{63} - x^{62} - x^{61} - x^{60} - x^{59}\\
&+2x^{58} + 5x^{57} + 3x^{56} - 2x^{55} - 10x^{54} - 3x^{53} - 2x^{52} + 6x^{51}\\
&+6x^{50} + x^{49} + 9x^{48} - 3x^{47} - 7x^{46} - 8x^{45} - 8x^{44} + 10x^{43}\\
&+6x^{42} + 8x^{41} - 5x^{40} - 12x^{39} + 7x^{38} - 7x^{37} + 7x^{36} + x^{35}\\
&-3x^{34} + 10x^{33} + x^{32} - 6x^{31} - 2x^{30} - 10x^{29} - 3x^{28} + 2x^{27}\\
&+9x^{26} - 3x^{25} + 14x^{24} - 8x^{23} - 7x^{21} + 9x^{20} + 3x^{19} - 4x^{18}\\
&-10x^{17} - 7x^{16} + 12x^{15} + 7x^{14} + 2x^{13} - 12x^{12} - 4x^{11} - 2x^{10}\\
&+5x^{9} + x^{7} - 7x^{6} + 7x^{5} - 4x^{4} + 12x^{3} - 6x^{2} + 3x - 6 = 0.
\end{aligned}$$

REFERENCES

Asger Aaboe. *Episodes from the Early History of Mathematics,* New Math. Library, vol. 13, Math. Assoc. Amer., 1964, esp. pp. 81-92.

Richard B. Austin and Richard K. Guy. Binary sequences without isolated ones. *Fibonacci Quart.,* **16** (1978): 84-86; *MR* **57** #5778.

Petr Beckmann. *A History of* π, The Golem Press, Boulder CO, 4th edition, 1977.

John H. Conway. The weird and wonderful chemistry of audioactive decay. *Eureka,* **1985**: 5-18.

Richard Dedekind. *Stetigkeit und Irrationalzahlen,* Braunschweig, 1872.

Howard Eves. *An Introduction to the History of Mathematics.* Holt, Rinehart & Winston, New York, 4th ed., 1976; esp. p. 201, problem study 7.11(c).

G. A. Freiman. *Diophantine Approximations and the Geometry of Numbers* (in Russian), *Kalinin. Gosudarstv. Univ.*, Kalinin, 1975.

Ronald L. Graham. The largest small hexagon. *J. Combin. Theory Ser. A*, **18** (1975): 165–170; *MR* **50** #12803.

F. Lindemann. Ueber die Zahl π *Math. Ann.*, **20** (1882): 213–225.

8

Imagining Imaginary Numbers

Historically, complex numbers first arose from the solution of quadratic equations. You can solve the equation

$$x^2 - x - 6 = 0$$

either by factoring it as $(x-3)(x+2) = 0$, or by using the well-known formula, or by writing it as

$$\left(x - \frac{1}{2}\right)^2 = 6\,\frac{1}{4} = \left(2\,\frac{1}{2}\right)^2,$$

and you get the answers $\frac{1}{2} \pm 2\frac{1}{2}$, namely, 3 or -2; sensible answers and there's no problem.

But for the equation

$$x^2 - 2x + 2 = 0,$$

we arrive at

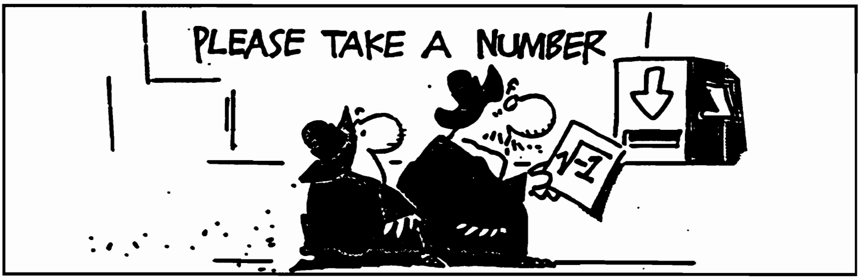
$$(x-1)^2 = -1,$$

and since the squares of real numbers are positive (or possibly zero), it looks as though there aren't any roots. But if we allow ourselves to invent a number called "the square root of minus one," then we could get

$$x - 1 = \pm\sqrt{-1},$$

$$x = 1 + \sqrt{-1} \text{ or } 1 - \sqrt{-1}.$$

Why do these seem nonsensical?



IMAGINARY NUMBERS ARE REAL! COMPLEX NUMBERS ARE SIMPLE!

The early investigators were very puzzled about the dubious nature of these new numbers, and all sorts of insults were hurled at them. In spite of this, some brave souls just followed the formulas wherever they led and seemed not to make mistakes.

Although Gauss, Argand, and Wessel cleared the matter up nearly 200 years ago, the mystery persists in some people's minds even today. Indeed, we still call the square roots of negative numbers "imaginary," although, as we'll show you soon, they're just as "real" as real numbers. They turn out to be invaluable in many applications of mathematics to engineering, physics, and almost every other science. Moreover, these numbers obey all the rules which you already know for "real" numbers.

We pause to point out that the so-called real numbers are not as real as you might think and do not have much relevance to physical reality. Does $\sqrt{2}$ correspond to anything in the physical world? The definition as the ratio of the diagonal of a square to its side is part of theoretical Euclidean geometry and supposes an infinite precision of measurement that is physically impossible.

The geometry that the physicists have used since Einstein is not Euclidean, and the diagonal of a large enough square might differ appreciably from $\sqrt{2}$ times its side. Even if the large-scale geometry of physics *were* Euclidean, after the quantum theory it becomes unlikely that at the atomic level it would be Euclidean, or even meaningful. It has been questioned whether space can be divided any more finely than the Planck length of $1.05457266 \times 10^{-34}$ or time below the Planck time of 10^{-43} seconds.

No! The square root of two, like other infinite precision real numbers such as *e* and π, is not really real in the physical sense! They are all figments of the mathematician's mind: concepts of abstract mathematics that only approximately correspond to things in the real world.

In fact, astonished reactions have greeted the introduction of each new kind of number.

When negative numbers were introduced, they were deemed impossible. What does it mean to speak of -3 apples? Of course -3 is not a "real" number! But *now* it seems quite sensible to speak of negative temperatures and negative bank balances.

Consternation also reigned when Pythagoras first discovered incommensurable numbers. The word "irrational" is etymologically the same as "unreasonable," while "surd" is similarly related to "absurd."

No doubt when fractions were first introduced, people wondered what was meant by 4/5 of a person, and we still find it amusing when we're told that the average family contains 2.3 children (Figure 8.1).

Even the whole numbers themselves are artificial constructs: a brace of pheasants is real enough, but the number 2 is a theoretical

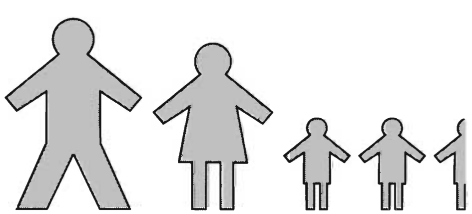


FIGURE 8.1 *The average family.*

abstraction. We saw in Chapter 1 that some primitive languages have no names for abstract numbers.

REALIZING COMPLEX NUMBERS

What a real number really is, is a distance together with a direction given by its sign, as measured in terms of some given unit.

Thus, in Figure 8.2, the number x is the ratio OX/OU, measured with respect to a fixed length OU, and x would be negative if X were to the left of O.

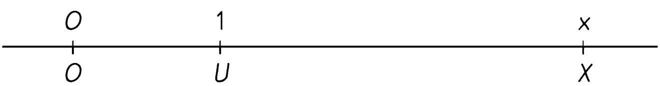
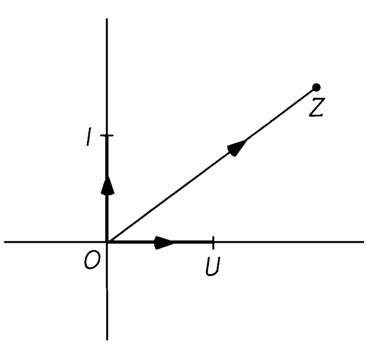


FIGURE 8.2 *What a real number really is.*

For real numbers, the only directions are left and right along a fixed line. In the same spirit we can define *complex numbers* as distances along arbitrary directions in a fixed plane. A **complex number** is a point of the plane in which there is a distinguished directed line segment OU, of length one unit. In particular, we define $i = \sqrt{-1}$ to be the point I, 1 unit from O in a direction 90° from OU (Figure 8.3). O, U, and I are called 0, 1, and i.

For the new numbers we could define addition and multiplication however we like, but we'll do it so that they fit with the rules we already know for real numbers, where adding 2 corresponds to sliding 2 units to the right on the real number line, and multiplying by 3 means expanding the scale by 3, keeping O fixed.

FIGURE 8.3 *What a complex number really is.*



SHIFTS AND TWIRLS: ADDING AND MULTIPLYING BY GEOMETRY

We'll use the word **shift** for a movement in which the entire plane is translated without any rotation or distortion. By a **twirl** we mean a rotation combined with an expansion or contraction, with the point *O* remaining fixed.

To **add** a fixed number *K* to each point of the plane, apply the shift that takes *O* to *K* (Figure 8.4(a)). To **multiply** by *K*, perform the twirl that takes *U* (= 1) to *K* (Figure 8.4(b)).

Subtraction and **division** do not require separate consideration. To *subtract* *c*, use the unique shift that takes *c* to *O*(= 0). To *divide* by *c*, use the unique twirl that takes *c* to 1.

WHY THE RULES WORK

The geometrical definitions really do have nice arithmetic properties. A shift followed by a shift is another shift. For instance, the shift "$+b$"

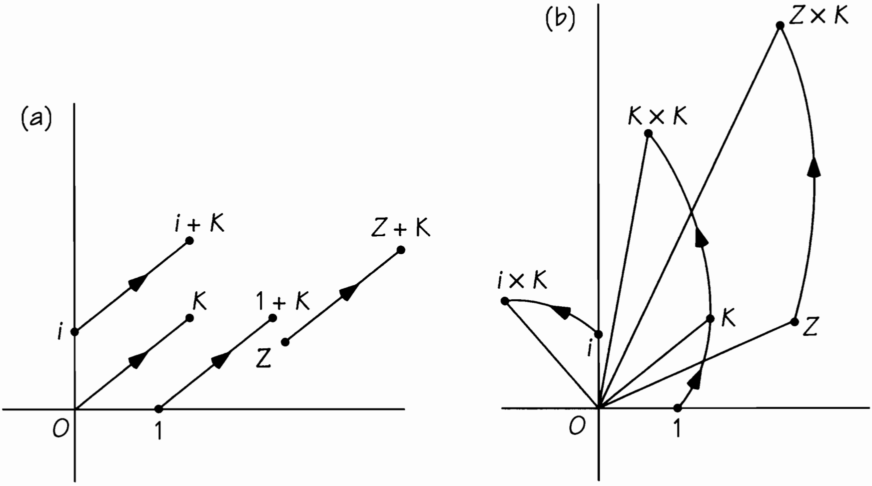


FIGURE 8.4 (*a*) *The shift "+K" that adds the number K.* (*b*) *The twirl "×K" that multiplies by K.*

followed by the shift "$+c$" takes 0 to $b + c$ and so must be the shift "$+(b + c)$." This establishes the **associative law of addition**:

$$(a + b) + c = a + (b + c)$$

and if we replace shifts by twirls we similarly find the **associative law of multiplication**:

$$(ab)c = a(bc).$$

In fact, all the algebraic rules have easy geometrical explanations.

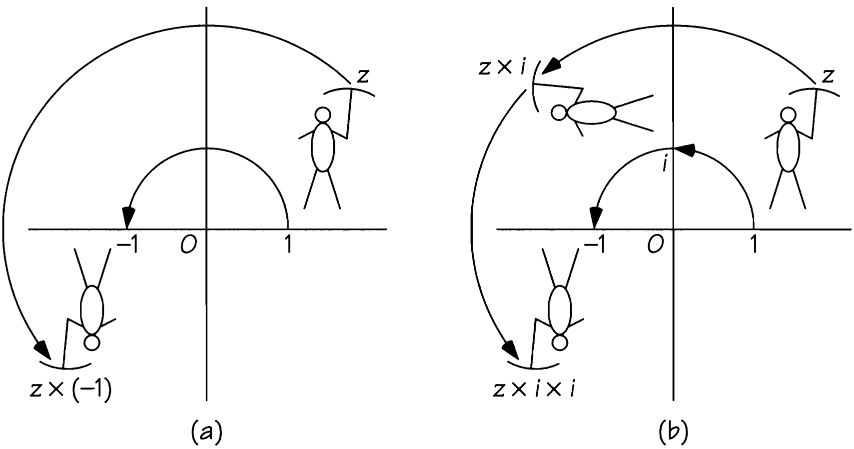


FIGURE 8.5 *(a) The twirl "×(−1)." (b) Two successive twirls "×i."*

The twirl "$\times(-1)$" is just a half-turn about 0 (Figure 8.5(a)), but the twirl "$\times i$" is a quarter-turn about 0, and two quarter-turns make a half-turn (Figure 8.5(b)), so if you follow "$\times i$" by another "$\times i$" you get the same effect as "$\times(-1)$". So indeed

> **If you multiply a complex number by i,**
> **and then by i again, you've multiplied it by -1,**
> $z\times i\times i = z\times(-1) = -z \qquad i\times i = -1$

Our geometrical definitions have produced a system of numbers that have a square root of -1 and satisfy the usual algebraic rules.

GAUSS'S WHOLE NUMBERS

Which complex numbers are most analogous to the ordinary whole numbers? The great Gauss suggested that we call a complex number $a + bi$ an integer when a and b are among the ordinary integers, $\ldots -2, -1, 0, 1, 2, \ldots$. Today we call these numbers the **Gaussian integers**. They form a square lattice in the complex plane (Figure 8.6).

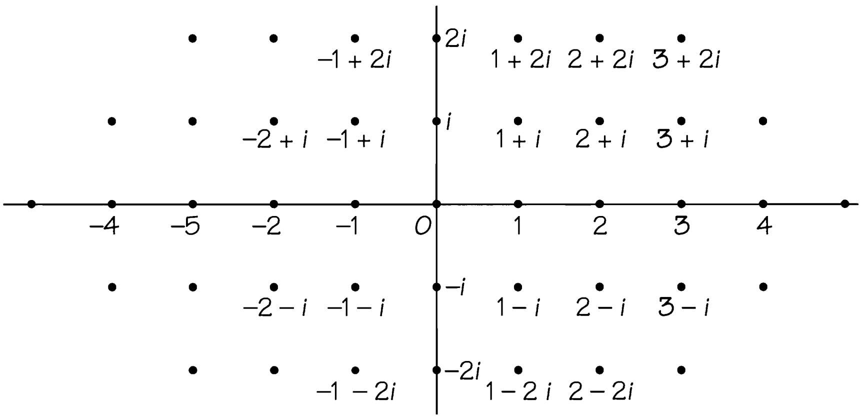


FIGURE 8.6 *The Gaussian integers.*

Gauss discovered the wonderful fact that his complex integers can be uniquely factored into prime Gaussian integers. Any ordinary positive or negative integer is uniquely the product of a power of -1 and powers of positive prime numbers. Analogously, Gauss found that

> Every nonzero Gaussian integer is uniquely expressible as a product of a power of i and powers of "positive" Gaussian primes.

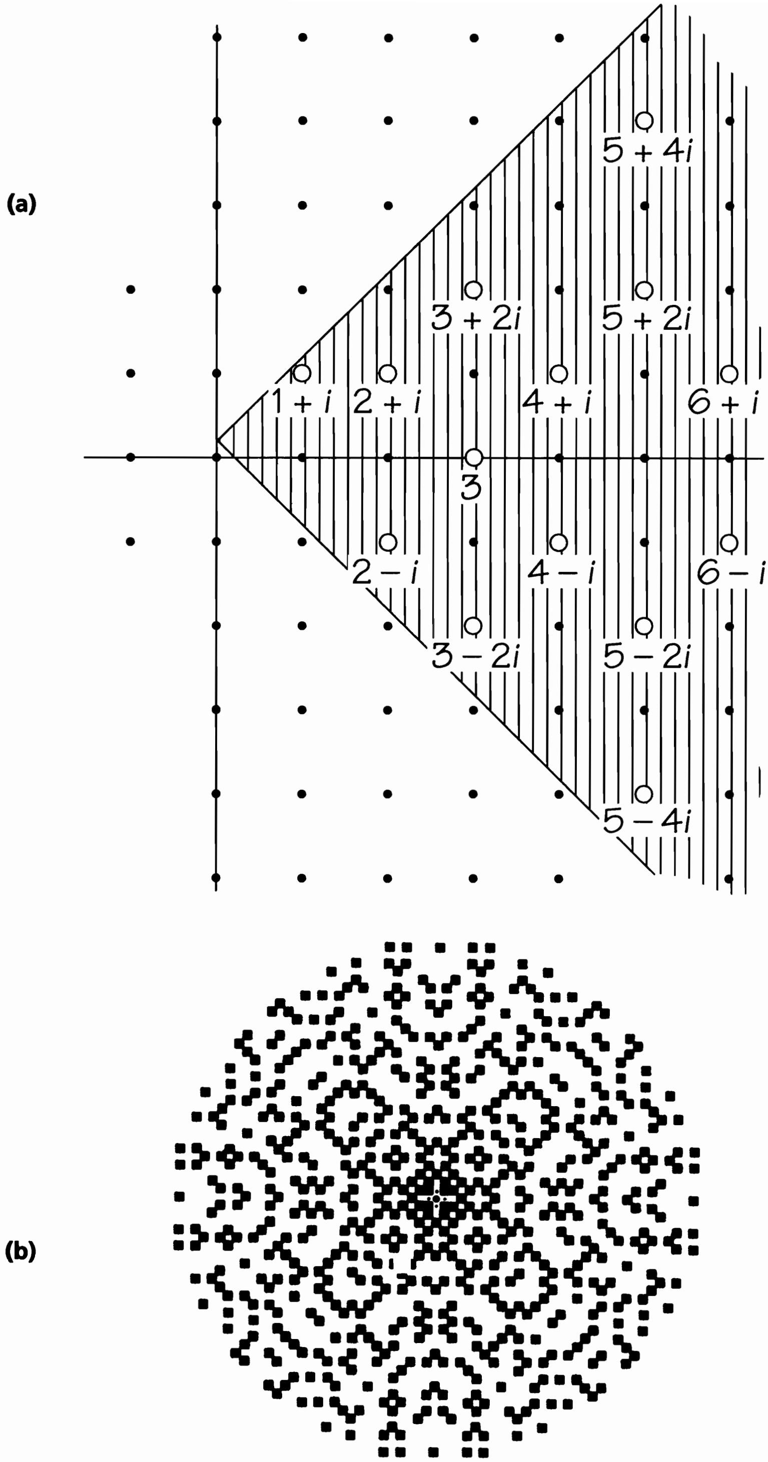


FIGURE 8.7 *(a) The "positive" Gaussian numbers. (b) The Gaussian primes.*

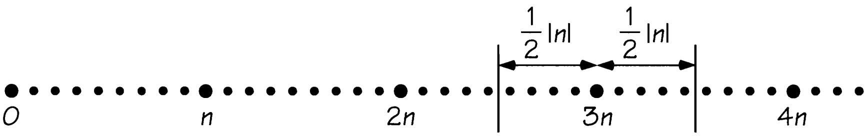


FIGURE 8.8 *Getting within $\frac{1}{2}|n|$ when you're dividing by n.*

The "positive" Gaussian numbers are those in the shaded region in Figure 8.7(a).

The first few (positive) Gaussian primes, circled in Figure 8.7(a), are $1+i$, $2+i$, $2-i$, 3, $3+2i$, $3-2i$, $4+i$, $4-i$, $5+2i$, $5-2i$, $6+i$, $6-i$, $5+4i$, $5-4i$, 7, $7+2i$, $7-2i$, . . . , and you can read off more Gaussian primes from Figure 8.7(b), where we've included their multiples by powers of i, as well as the "positive" ones.

When we proved the unique factorization theorem for ordinary integers in Chapter 5, we only used the fact that you could divide one number by another and get a smaller remainder. In fact, if you allow *negative* remainders, you can get within $\frac{1}{2}|n|$ of a multiple of the number n you're dividing by (Figure 8.8, where $|n|$ is the distance from 0 to n).

Figure 8.9 shows how every Gaussian integer is within $|n|/\sqrt{2}$ of

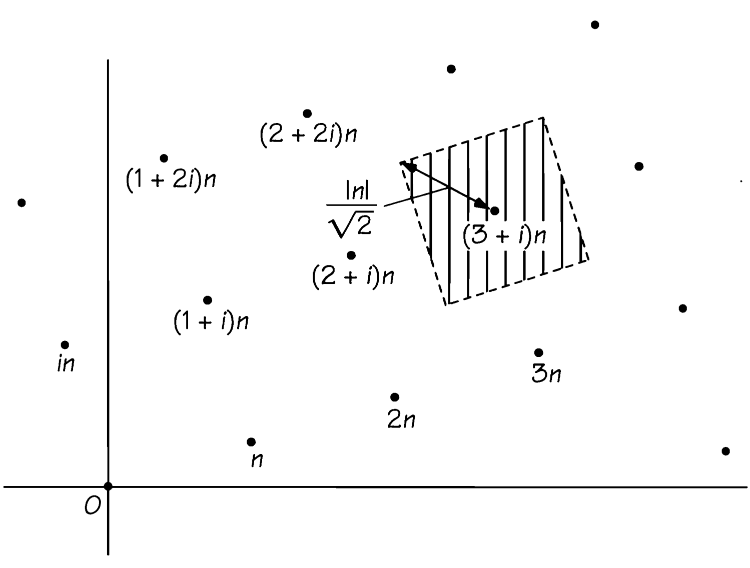


FIGURE 8.9 *The size of the remainder when you divide by a Gaussian integer.*

a multiple of a Gaussian integer n. Since $1/\sqrt{2}$ is less than 1, the remainder is smaller in size than the divisor and our proof of unique factorization will work as well for the Gaussian integers.

Let's see why 13 *isn't* prime as a *Gaussian* number. From Wilson's test for primality we know that 13 divides

$$12! + 1 = 1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 \times 9 \times 10 \times 11 \times 12 + 1.$$

Working modulo 13, this is congruent to

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times (-6) \times (-5) \times (-4) \times (-3) \times (-2) \times (-1) + 1,$$

which is $(6!)^2+1 = (6!)^2-i^2 = (6!+i)(6!-i)$, so, as a Gaussian integer, 13 can't be prime since it divides the product of $6!+i$ and $6!-i$ without dividing either factor (of course the Gaussian multiples of 13 are of shape $13a+13bi$). If $a+bi$ is a "positive" prime divisor of 13, then so is $a-bi$, and 13 is divisible by $(a+bi)(a-bi) = a^2+b^2$, so 13 *must be* of such a form.

A similar argument shows that an ordinary prime number, p, factorizes in the Gaussian sense just when $p + 1$ is *not* a multiple of 4, and so establishes a famous assertion of Fermat:

A prime number p may be expressed as the sum of 2 squares just if $p + 1$ is not a multiple of 4.

Fermat's two-square theorem.

In the *Gaussian* world,

$2 = (1 + i)(1 - i) = 1^2 + 1^2$,
3 is still prime,
$5 = (2 + i)(2 - i) = 2^2 + 1^2$,
7 and 11 are still primes,
$13 = (3 + 2i)(3 - 2i) = 3^2 + 2^2$.

The expression of a prime as the sum of two squares, when possible, is unique.

Gauss's disciple, Eisenstein, suggested an alternative system of complex "whole numbers." The **Eisenstein integers** are the numbers $a+b\omega$, where $\omega = (-1+i\sqrt{3})/2$ is one of the roots of $x^3 = 1$, and the others are 1 and $\omega^2 = (-1-i\sqrt{3})/2$. The Eisenstein integers form a triangular lattice (Figure 8.10(a)). Figure 8.10(b) is similar to Figure 8.9 and shows that every Eisenstein integer is within a distance $|n|/\sqrt{3}$ of some multiple of a given Eisenstein integer n.

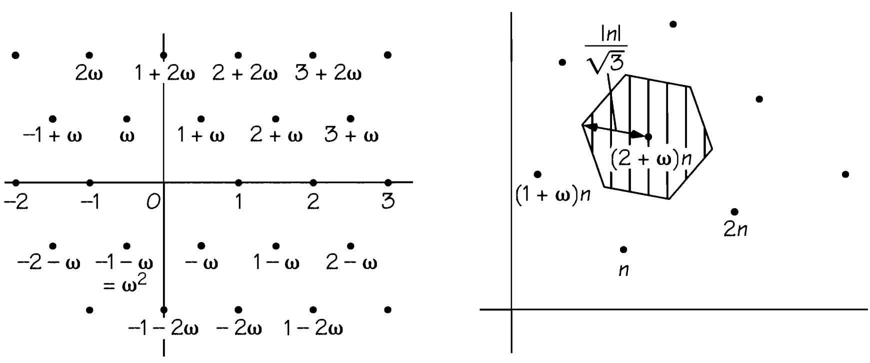
So the Eisenstein integers also have unique factorization. This time there are six **Eisenstein units**, ± 1, $\pm\omega$, $\pm\omega^2$, and the precise statement is

> Any nonzero Eisenstein integer is uniquely the product of powers of -1, ω, and the "positive" Eisenstein primes,

where the "positive" numbers are those in the 60° sector in Figure 8.11(a), where we've again circled the primes. Figure 8.11(b) shows the beautiful symmetry of these primes when you multiply the "positive" ones by the Eisenstein units.

The analog of Fermat's sum of two squares theorem is that an ordinary prime number p can be written in the form

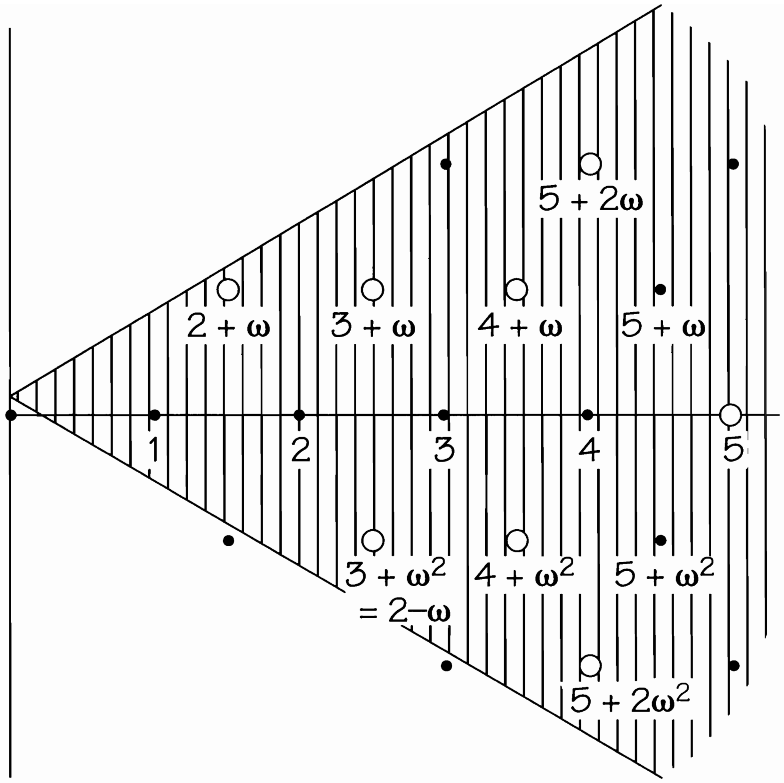
$$a^2 - ab + b^2 = (a + b\omega)(a + b\omega^2)$$



(a) (b)

FIGURE 8.10 *(a) The triangular lattice of Eisenstein integers. (b) The size of the remainder when you divide by an Eisenstein integer.*

(a)



(b)

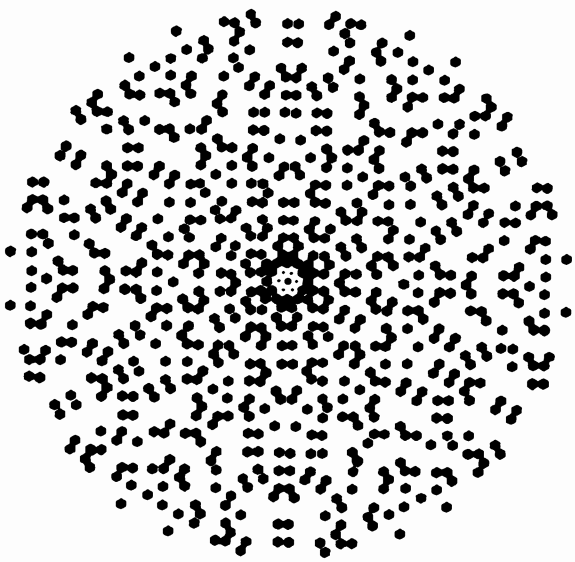


FIGURE 8.11 *(a) The "positive" Eisenstein numbers. (b) The Eisenstein primes.*

just if 3 does not divide $p + 1$. These are precisely the ordinary primes that factor in the world of Eisenstein integers:

$$\begin{aligned} 3 &= (2+\omega)(2+\omega^2) = -\omega(2+\omega)^2, \\ 7 &= (3+\omega)(2+\omega^2), \\ 13 &= (4+\omega)(4+\omega^2), \\ 19 &= (5+2\omega)(5+2\omega^2), \\ 31 &= (6+\omega)(6+\omega^2), \end{aligned}$$

and so on: they are just the primes that can be expressed in the form $3m^2 + n^2$. To see this, use one of these three forms of $a^2 - ab + b^2$:

$$3\left(\frac{a}{2}\right)^2 + \left(\frac{a}{2} - b\right)^2 \quad 3\left(\frac{b}{2}\right)^2 + \left(\frac{b}{2} - a\right)^2 \quad 3\left(\frac{a-b}{2}\right)^2 + \left(\frac{a+b}{2}\right)^2$$

according as it is a, b or $a + b$ that is even.

Is There Always Unique Factorization?

It was not obvious that ordinary integers have unique factorization into primes, and our proofs for the Gaussian and Eisenstein integers (which involve $\sqrt{-1}$ and $\sqrt{-3}$) depended on the geometrical facts about squares and equilateral triangles.

Indeed, for the numbers $a + b\sqrt{-5}$, it isn't even true! In this system of numbers, 6 factorizes in two different ways:

$$6 = 2\times 3 = (1+\sqrt{-5})(1-\sqrt{-5}),$$

and none of the numbers 2, 3, $1+\sqrt{-5}$, $1-\sqrt{-5}$ will factor any further.

When doing divisions in this system of numbers, it's not always possible to get a remainder with smaller size than the divisor. Figure 8.12 shows why.

The distance of the corners of the shaded rectangle from the nearest integer multiple of $|d|$ is $\sqrt{\frac{3}{2}}|d|$, so there will be integers near those corners that are distance more than $|d|$ away from the nearest integer multiple.

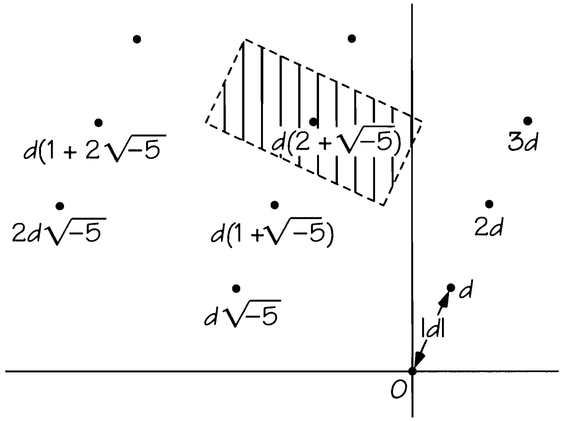


FIGURE 8.12 *Integer multiples of d in the number system $a + b\sqrt{-5}$.*

THE NINE MAGIC DISCRIMINANTS

For exactly which negative numbers $-d$ does $\sqrt{-d}$ lead to a number system that has unique factorization into primes? The answer is now known; $-d$ must be one of the "Heegner numbers"

$$-1,\ -2,\ -3,\ -7,\ -11,\ -19,\ -43,\ -67,\ -163.$$

(In all except the first two cases we must allow, as integers, numbers $a + b\sqrt{-d}$ in which a and b are *halves* of integers, as we did for the Eisenstein integers.)

For a long time mathematicians were aware of these nine but were in the tantalizing position of knowing that there could be at most one more. The problem of deciding whether this outsider really existed was a notorious one, called the "tenth discriminant problem."

In 1936 Heilbronn and Linfoot showed that such a tenth d was bigger than 10^9. In 1952 Heegner published a proof that the list of nine was complete, but experts had some doubt about its validity. In 1966–67 two young mathematicians, Harold Stark in the United States and Alan Baker in Great Britain, independently obtained proofs, and the world was convinced. The story didn't really end here, because a year or two later Stark made a careful and detailed examination of

Heegner's proof and found that the critics had been unfair: the proof was essentially correct.

These numbers have many interesting properties. Euler discovered that the formula

$$n^2 - n + 41$$

gives the prime numbers

41 43 47 53 61 71 83 97 113 131 151 173 197 223 251
281 313 347 383 421 461 503 547 593 641 691 743 797 853
911 971 1033 1097 1163 1231 1301 1373 1447 1523 1601

when we set $n = 1, 2, 3, \ldots, 40$. What's the explanation?

The equation $x^2 - x + 41 = 0$ has solutions $x = \frac{1}{2}(1 \pm \sqrt{-163})$, and it can be shown that for a number $k > 1$, the formula

$$n^2 - n + k$$

represents primes for the consecutive numbers $n = 1, 2, \ldots, k - 1$ as long as $1 - 4k$ is one of the Heegner numbers. Now that we know them all, this leaves only the cases $k = 2, 3, 5, 11, 17$ and the one we've seen, 41.

Values for $n = 1, 2, \ldots, k - 1$

n^2-n+2	2
n^2-n+3	3 5
n^2-n+5	5 7 11 17
n^2-n+11	11 13 17 23 31 41 53 67 81 101
n^2-n+17	17 19 23 29 37 47 59 73 89 107 127 149 173 199 227 257

Another remarkable fact is that the numbers

$$e^{\pi\sqrt{43}} = 884736743.999777\ldots,$$

$$e^{\pi\sqrt{67}} = 147197952743.99999866\ldots,$$

$$e^{\pi\sqrt{163}} = 262537412640768743.99999999999925007\ldots$$

are suspiciously close to integers. This is no mere accident! (The last one was part of Martin Gardner's famous April Fool hoax in 1975.) It can in fact be shown that for these numbers $X = e^{\pi\sqrt{d}}$, the formula

$$X - 744 + \frac{196884}{X} - \frac{21493760}{X^2} + \cdots$$

is exactly an integer and indeed a perfect cube! For the above values X is so large that the later terms are extremely small, so X itself must be nearly an integer.

$$e^{\pi\sqrt{43}} = 960^3 + 744 - \text{(a bit)},$$

$$e^{\pi\sqrt{67}} = 5280^3 + 744 - \text{(a tiny bit)},$$

$$e^{\pi\sqrt{163}} = 640320^3 + 744 - \text{(a very tiny bit)!}$$

De Moivre's Circle-Cutting Numbers

Go to Mr. De Moivre; he knows these things better than I do.

Isaac Newton

Draw a regular polygon, centered at the origin in the complex plane, with one corner being the number 1 (Figure 8.13). What are the complex numbers corresponding to all the corners? These numbers were studied by the English mathematician Abraham De Moivre (1667–1754) long before it was realized that they had a geometrical meaning.

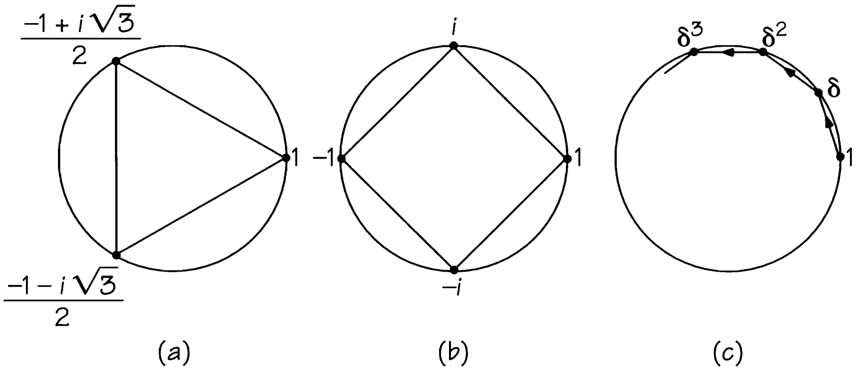


FIGURE 8.13 *De Moivre's cyclotomic numbers.*

When the polygon is a square (Figure 8.13(b)), the answer is easy. The numbers (reading counterclockwise) are 1, i, -1, $-i$. We call these the fourth-order **De Moivre numbers**. In a similar way, the third-order De Moivre numbers (Figure 8.13(a)) are 1, $(-1+i\sqrt{3})/2$, $(-1-i\sqrt{3})/2$, and the sixth-order ones are these and their negatives:

$$1,\ (1+i\sqrt{3})/2,\ (-1+i\sqrt{3})/2,\ -1,\ (-1-i\sqrt{3})/2,\ (1-i\sqrt{3})/2$$

(again reading counterclockwise). The fifth-order ones are a bit harder: You have to solve one quadratic equation on top of another.

What can we say about these numbers? They are all powers of one of them! The De Moivre number that immediately follows 1 in counterclockwise order we'll call δ_n (if the polygon has n corners), or just δ when n is obvious. Since the twirl that takes 1 to δ (Figure 8.13(c)) takes δ to δ^2, δ^2 to δ^3, and so on, we see that the full set of nth-order De Moivre numbers is

$$1,\ \delta,\ \delta^2,\ \ldots,\ \delta^{n-1},\ \text{where } \delta = \delta_n,$$

and we also see that $\delta^n = 1$. But we saw that $\delta_4 = i$, which satisfies $\delta^2 + 1 = 0$, so $\delta^n = 1$ is not always the simplest equation satisfied by δ_n. The following table gives the simplest equation for δ_n for $n = 1, \ldots, 12$.

n	Simplest equation	degree
1	$\delta - 1 = 0$	1
2	$\delta + 1 = 0$	1
3	$\delta^2 + \delta + 1 = 0$	2
4	$\delta^2 + 1 = 0$	2
5	$\delta^4 + \delta^3 + \delta^2 + \delta + 1 = 0$	4
6	$\delta^2 - \delta + 1$	2
7	$\delta^6 + \delta^5 + \delta^4 + \delta^3 + \delta^2 + \delta + 1 = 0$	6
8	$\delta^4 + 1 = 0$	4
9	$\delta^6 + \delta^3 + 1 = 0$	6
10	$\delta^4 - \delta^3 + \delta^2 - \delta + 1 = 0$	4
11	$\delta^{10} + \delta^9 + \delta^8 + \delta^7 + \delta^6 + \delta^5 + \delta^4 + \delta^3 + \delta^2 + \delta + 1 = 0$	10
12	$\delta^4 - \delta^2 + 1 = 0$	4

In fact, the nth-order De Moivre number, δ_n, is an algebraic number whose degree is the nth of Euler's totient numbers, $\phi(n)$. The solutions of this equation are all the **primitive** nth roots of unity. They satisfy $x^k = 1$ for $k = n$, but for no smaller value of k.

We said that δ_5 could be found by solving one quadratic on top of another. In fact,

$$\delta_5 = \frac{1}{4}\left(-1 + \sqrt{5} + i\sqrt{10 + 2\sqrt{5}}\right)$$

as you can see by writing the equation $\delta^4 + \delta^3 + \delta^2 + \delta + 1 = 0$ in the form $(\delta^2 + \frac{1}{2}\delta + 1)^2 = (\frac{1}{2}\delta\sqrt{5})^2$.

THE ONLY RATIONAL TRIANGLE

Suppose you have a triangle all of whose sides are rational numbers and all of whose angles are rational numbers of degrees. Then it must be equilateral! You might think that this was a very deep result, but in fact it's surprisingly easy to verify it, using De Moivre's numbers. Put your triangle in the complex plane (Figure 8.14). Then the angles indicated will be p $n$ths and q $n$ths of a revolution, where we take the smallest possible value of n. Since the three indicated shifts add up to zero, we see that the nth-order De Moivre number $\delta = \delta_n$ must satisfy

$$a + b\delta^p + c\delta^q = 0,$$

as, therefore, must all the algebraic conjugates of δ_n that have the form δ^k whenever k and n have no common factor.

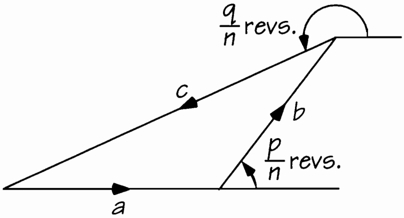


FIGURE 8.14 *An equilateral triangle?*

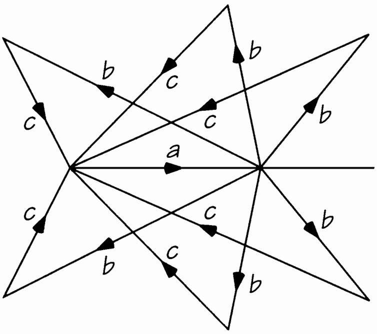


FIGURE 8.15 *At most two such triangles!*

So indeed we have $\phi(n)$ equations

$$a + b\delta^{kp} + c\delta^{kq}$$

and so $\phi(n)$ different triangles, such as in Figure 8.15, all of which have the same base and the same edge lengths, but different angles. This is nonsense, because there can be at most two such triangles: the original one (Figure 8.14) and its reflection in the base (the two rightmost in Figure 8.15). Now the only numbers with $\phi(n) \leq 2$ are $n = 1, 2, 3, 4$, or 6, so all angles are at least $\frac{1}{6}$ of a revolution (i.e., 60°), and your triangle must be equilateral, since its angles add up to 180°.

THE REGULAR HEPTADECAGON

In Chapter 7 (Figure 7.22) we constructed a regular 17-sided polygon inside a circle of radius 2. Its vertices, $P_0, P_1, \ldots, P_{16}$ are therefore the doubles of the 17th-order De Moivre numbers ($\delta = \delta_{17}$):

$$2,\ 2\delta,\ 2\delta^2,\ 2\delta^3,\ \ldots,\ 2\delta^{16}$$

and the points are

Q_1	Q_2	Q_3	Q_4	Q_5	Q_6	Q_7	Q_8
$\delta^1 + \delta^{16}$	$\delta^2 + \delta^{15}$	$\delta^3 + \delta^{14}$	$\delta^4 + \delta^{13}$	$\delta^5 + \delta^{12}$	$\delta^6 + \delta^{11}$	$\delta^7 + \delta^{10}$	$\delta^8 + \delta^9$

Now δ satisfies the equation

$$\delta^{16} + \delta^{15} + \delta^{14} + \delta^{13} + \delta^{12} + \delta^{11} + \delta^{10} + \delta^9 + \delta^8 + \delta^7 + \delta^6 + \delta^5 + \delta^4 + \delta^3 + \delta^2 + \delta + 1 = 0.$$

Let's write the powers of δ in the order

$$\delta^1\ \delta^3\ \delta^9\ \delta^{10}\ \delta^{13}\ \delta^5\ \delta^{15}\ \delta^{11}\ \delta^{16}\ \delta^{14}\ \delta^8\ \delta^7\ \delta^4\ \delta^{12}\ \delta^2\ \delta^6$$

where the exponents, modulo 17, are

$$1\ \ 3\ \ 3^2\ \ 3^3\ \ 3^4\ \ 3^5\ \ 3^6\ \ 3^7\ \ 3^8\ \ 3^9\ \ 3^{10}\ \ 3^{11}\ \ 3^{12}\ \ 3^{13}\ \ 3^{14}\ \ 3^{15}$$

By taking alternate terms we get the two numbers

$$\alpha = \delta^1 + \delta^9 + \delta^{13} + \delta^{15} + \delta^{16} + \delta^8 + \delta^4 + \delta^2,$$
$$\beta = \delta^3 + \delta^{10} + \delta^5 + \delta^{11} + \delta^{14} + \delta^7 + \delta^{12} + \delta^6,$$

so that $\alpha+\beta = -1$. If we multiply α by β we get the sum of 64 terms consisting of all the numbers $\delta, \delta^2, \ldots, \delta^{16}$ each repeated 4 times, so that $\alpha\beta = -4$, and $\alpha = (-1+\sqrt{17})/2$ and $\beta = (-1+\sqrt{17})/2$ are the roots of the quadratic equation $x^2+x-4 = 0$.

We now repeat the process, taking alternate terms from α and β, to find the four numbers

$$a = \delta^1 + \delta^{13} + \delta^{16} + \delta^4, \qquad b = \delta^3 + \delta^5 + \delta^{14} + \delta^{12},$$
$$c = \delta^9 + \delta^{15} + \delta^8 + \delta^2, \qquad d = \delta^{10} + \delta^{11} + \delta^7 + \delta^6$$

that satisfy $a+c = \alpha$, $b+d = \beta$, $ac = bd = -1$, so that a, c and b, d are the roots of the quadratics $x^2-\alpha x-1 = 0$ and $x^2-\beta x-1 = 0$.

In the geometrical construction for any particular vertex of the polygon, we solve four quadratic equations in succession. Our construction does the first two of these by angle bisection, the third by the method of Chapter 7, Figure 7.8, and the last by finding the two points where the vertical line hits the circle. In fact, in Chapter 7, Figure 22, we *quadrisect* a certain angle: the *bisectors* of that angle pass through the points $(2\alpha, 0)$, $(2\beta, 0)$, and its quadrisectors through the points $(a, 0)$, $(b, 0)$, $(c, 0)$, $(d, 0)$. The circles on YA^*, YB^*, YC^*, YD^* then solve the equations

$$x^2 - ax + b = 0,$$
$$x^2 - bx + c = 0,$$
$$x^2 - cx + d = 0,$$
$$x^2 - dx + a = 0,$$

cutting the axis in Q_1, Q_4; Q_3, Q_5; Q_8, Q_2; Q_7, Q_6, with coordinates

$$\delta+\delta^{16},\ \delta^{13}+\delta^4;\ \delta^3+\delta^{14},\ \delta^5+\delta^{12};\ \delta^9+\delta^8,\ \delta^{15}+\delta^2;\ \delta^{10}+\delta^7,\ \delta^{11}+\delta^6.$$

HYPERCOMPLEX NUMBERS

"Nor could I resist the
impulse—unphilosophical
as it may have been—to
cut with a knife on a stone
of Brougham bridge the
fundamental formula with
the symbols i, j, k:
$i^2 = j^2 = k^2 = ijk = -1$."

We have seen that complex numbers are much the same as couples of real numbers and that they have a lot to tell us about plane geometry. The famous Irish mathematician, William Rowan Hamilton (1805–1865) spent a considerable amount of time trying to find an analog with triples of real numbers, which he hoped would perform a similar service for solid geometry. Every morning, on coming down to breakfast, his young son would ask him, "Well, Papa, can you multiply triplets?" but for a long time he was forced to reply with a sad shake of his head, "No, I can only add and subtract them."

Hamilton finally discovered that the correct thing to do was to use not only three coordinates, but four, and he devised a new system of numbers that he called **quaternions**. The typical quaternion,

$$q = a + bi + cj + dk$$

has a (one-dimensional, real) **scalar** part, a, but its "imaginary" part is the three-dimensional **vector**

$$v = bi + cj + dk.$$

You add them in the obvious way, but multiply them using

$$\boxed{\begin{array}{ccc} & i^2 = j^2 = k^2 = -1 & \\ ij = k & jk = i & ki = j \\ ji = -k & kj = -i & ik = -j \end{array}}$$

Hamilton's rules.

For instance:

$$(2 + i)(3 + j) = 6 + 3i + 2j + k,$$

$$(3 + j)(2 + i) = 6 + 3i + 2j - k,$$

but

Hamilton's quaternions are *not* commutative!

However, quaternions *do* satisfy the associative and distributive laws.

Hamilton's quaternions are indeed useful in geometry! The rotation through angle 2θ about any unit (length 1) vector, $bi + cj + dk$, takes any other vector $xi + yi + zk$ to

$$q^{-1}(xi + yj + zk)q,$$

where $q = \cos\theta + (bi + cj + dk)\sin\theta$.

The appropriate measure of the "size" of a typical quaternion $a + bi + cj + dk$ is its **norm** $a^2 + b^2 + c^2 + d^2$, and one of the first things that Hamilton did after discovering his rules was to check that the norm of the product of two quaternions is just the product of their norms. This gives the famous four-square formula

$$(a^2+b^2+c^2+d^2)(\alpha^2+\beta^2+\gamma^2+\delta^2) =$$
$$(a\alpha-b\beta-c\gamma-d\delta)^2+(a\beta+b\alpha+c\delta-d\gamma)^2+$$
$$(a\gamma-b\delta+c\alpha+d\beta)^2+(a\delta+b\gamma-c\beta+d\alpha)^2$$

already sent by Euler in a letter to Goldbach, April 15, 1705, and used by Lagrange in his proof of Fermat's assertion that every integer is the sum of four perfect squares.

Quaternions are also useful for representing groups in pure mathematics and spin in particle physics.

THE QUATERNION MACHINE

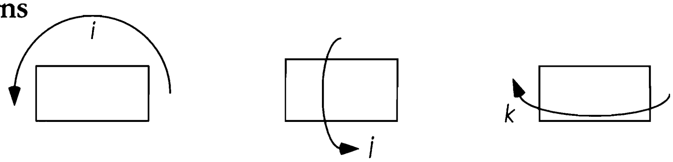
Figure 8.16 shows a simple little device that illustrates the multiplication of quaternions.

It consists of a rectangular card with 1, *i*, *j*, *k* written on it in various orientations, and hung from a rod by some stout webbing tape. The rod is perpendicular to our paper, so that the webbing is always twisted through at least 90° in our figures. Two small sign indicators are mounted halfway up the edges of the tape.

The rotations replacing *x* by

$i^{-1}xi$ $\quad j^{-1}xj$ $\quad k^{-1}xk$

are the half-turns



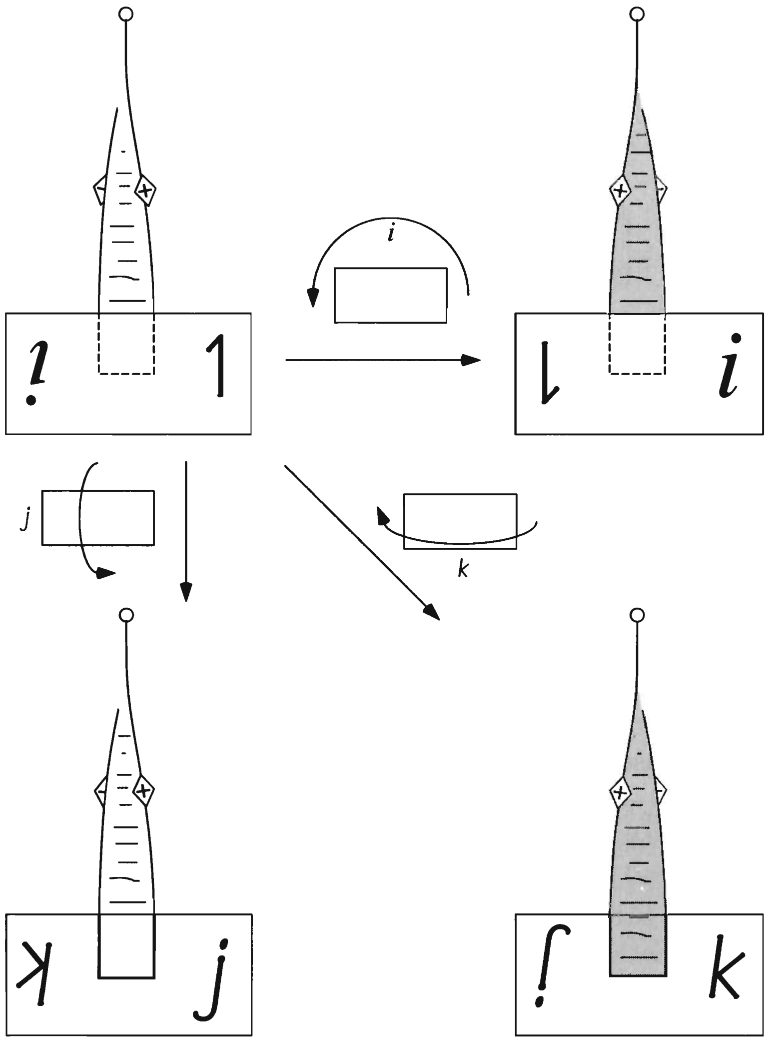


FIGURE 8.16 *Four positions for the quaternion machine.*

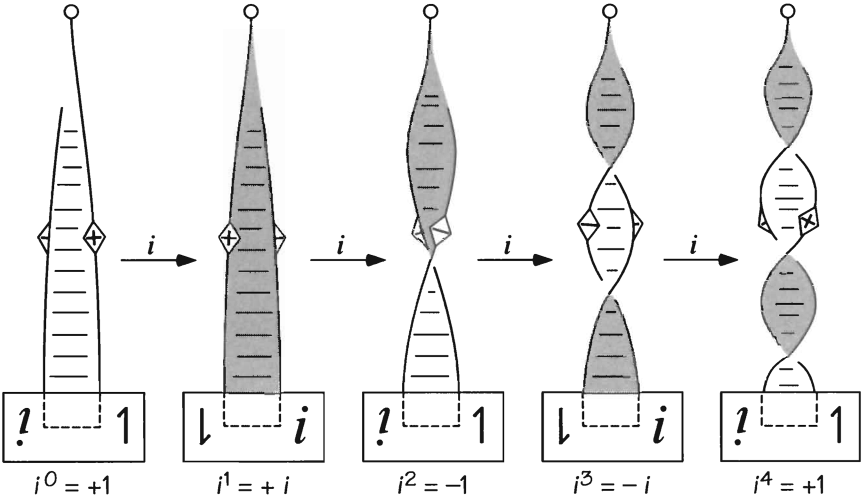


FIGURE 8.17 $i^4 = 1$.

If you do any sequence of these, then of course the resulting compound rotation is revealed by the final orientation of the card. But our sign indicator will also display the sign of the product of the quaternions.

After you've performed any number of multiplications by i, j, k in this way, you're allowed to move the card so as to simplify the result, ***provided that you carefully preserve its orientation throughout the motion.*** For example, the last configuration in Figure 8.17 can be untwisted into the first configuration, keeping the orientation (i inverted; 1—facing the observer) fixed.

Hamilton was anticipated in some respects by the very rich Spanish banker and mathematician, Rodrigues.

CAYLEY NUMBERS

"It is possible to form an analogous theory with 7 imaginary roots of −1."
Arthur Cayley, 1845

Arthur Cayley (1821–1895) discovered an eight-dimensional algebra of "numbers," called **octonions** or **Cayley numbers**, which have been used to explain certain special properties of seven-dimensional and eight-dimensional space. The typical Cayley number has the form

$$a + bi_0 + ci_1 + di_2 + ei_3 + fi_4 + gi_5 + hi_6,$$

where each of the triples

$$(i_0, i_1, i_3)\ (i_1, i_2, i_4)\ (i_2, i_3, i_5)\ (i_3, i_4, i_6)\ (i_4, i_5, i_0)\ (i_5, i_6, i_1)\ (i_6, i_0, i_2)$$

behaves like Hamilton's (i, j, k).

Any rotation of eight-dimensional space may be written in the form

$$x \text{ goes to } ((((((xc_1)c_2)c_3)c_4)c_5)c_6)c_7$$

for suitable Cayley numbers $c_1, c_2, c_3, c_4, c_5, c_6, c_7$.

BEWARE! Cayley numbers are *not* associative, so this *cannot* be written as

$$xc_1c_2c_3c_4c_5c_6c_7.$$

In 1898 Hurwitz proved that the algebras of real numbers, complex numbers, quaternions, and Cayley numbers are the *only* ones in which all the multiplication operators by unit vectors preserve distances. J.F. Adams proved in 1956 that *only* for $n = 1, 2, 4$, and 8 can n-dimensional vectors be turned into an algebra in which division (except by 0) is always possible.

REFERENCES

H.A. Heilbronn and E.H. Linfoot. On the imaginary quadratic corpora of class-number one. *Quart. J. Math. (Oxford),* **5** (1934): 293–301; *Zbl.* 10.337.

A. Hurwitz. Über die Komposition quadratischer Formen von beliebig vielen Variabeln, *Nachr. Gesell. Wiss. Göttingen* **1898** 309–316.

D.H. Lehmer. On imaginary quadratic fields whose class number is unity abst. #188 *Bull. Amer. Math. Soc.,* **39** (1933), 360.

Harold M. Stark. On complex quadratic fields with class number equal to one. *Trans. Amer. Math. Soc.,* **122** (1966): 112–119; *MR* **33** #4043.

9

Some Transcendental Numbers

In this chapter we'll meet some numbers that transcend the bounds of algebra. The most famous ones are Ludolph's number π, Napier's number e, Liouville's number l, and various logarithms.

The Number π

Also he made a molten sea of ten cubits from brim to brim, round in compass, and five cubits the height thereof; and a line of thirty cubits did compass it round about.

2 Chronicles, iv, 2

The area of a circle of radius 1 is roughly

$$3.1415926535\ldots.$$

The perimeter of a circle of diameter 1 is the same number. Just what is this number? The above quotation suggests that its author thought that π was 3, although better approximations were known to the Babylonians ($3\frac{1}{8}$) and Egyptians ($2^8/3^4 \approx 3.1605$ in the Rhind papyrus, <1650 B.C.).

My dear friend, that must be a delusion, what can a circle have to do with the number of people alive at a given time?

An actuary, quoted by Augustus de Morgan.

The mathematical history of π really begins with Archimedes. He proved that the above two definitions do give exactly the same answer, and he also showed that it's the same number that occurs in the slightly more complicated formulas

$$4\pi r^2 \qquad \text{and} \qquad \frac{4}{3}\pi r^3$$

for the surface area and volume of a sphere, respectively. He also proved that π is strictly less than $3\frac{1}{7}$, but strictly greater than $3\frac{10}{71}$. He did this by computing the areas of a regular polygon of 96 sides, and his ingenious computation makes use of clever approximations for square roots.

FORMULAS FOR π

Much of the history of π consists of assertions that its value was approximately or exactly equal to some number. For instance, the great Indian mathematician Brahmagupta gave the value $\sqrt{10} = 3.162\ldots$ in about 600 A.D., even though the Chinese astronomer Tsu Ch'ung-chih (born 430 A.D.) had already asserted that π lay between 3.1415926 and 3.1415927, although, inconsistently, he gave 355/113, which he called the "accurate" value.

In 1593 Vieta exhibited a formula for π,

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}} \times \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}} \times \sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2} + \frac{1}{2}\sqrt{\frac{1}{2}}}} \times \cdots.$$

He found this by using Archimedes's idea, but with polygons with 4, 8, 16, . . . , 2^n sides in a circle of radius 1 (Figure 9.1).

Ludolph van Ceulen spent a large part of his life computing π by similar means, and in Germany π is still sometimes called **Ludolph's number**.

He directed that it be inscribed on his gravestone, which is now unfortunately lost. In 1596 he computed π to 20 decimal places and

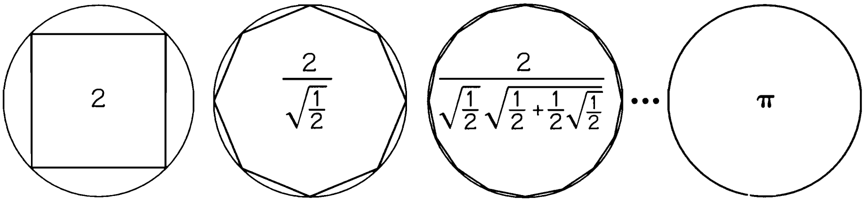


FIGURE 9.1 *Successive approximations to π.*

by the time of his death he had computed π to 35 decimal places:

$$3.14159\ 26535\ 89793\ 23846\ 26433\ 83279\ 50288\ldots.$$

Soon after Ludolph's death, many new formulas for π were discovered in great profusion, some of which can be used to compute π with great rapidity.

WHAT'S THE NATURE OF π?

It is possible today to calculate π to millions of decimal places, and it's natural to ask what kind of number it is. You know that it's very close to $\frac{22}{7}$, but not exactly equal. It's even closer to $\frac{355}{113}$ = 3.1415929203 . . . , a value discovered by the father of Adrian Metius in 1585, but even that number is not exact. Is it possible that π is exactly equal to some fraction? The answer is "no," as was shown by the great German mathematician Lambert in 1761. Lambert also showed that π wasn't even the square root of a fraction.

In fact, many mathematicians must have suspected that π was not the root of any algebraic equation with whole number coefficients, that is, that π is a **transcendental number**. The eventual proof of this by Lindemann in 1882 is one of the milestones of mathematics.

LIOUVILLE'S NUMBER

The first provably transcendental numbers were found by Liouville in 1844. Liouville's number

$$l = 0.110001\ 000000\ 000000\ 000001\ 000000\ 000000\ldots$$

is typical. The only nonzero digits are those in the 1st, 2nd, 6th, . . . , (n!)th . . . decimal places.

It *nearly* satisfies the equation

$$10x^6 - 75x^3 - 190x + 21 = 0,$$

but not exactly, since, for $x = l$, the left side works out to be

$$-0.00000\ 00059\ 48422\ 11323\ 4\ldots.$$

We'll show that *no* such equation can be exactly satisfied by l. Take the suspected equation and gather the terms of given sign on opposite sides, getting an equation such as

$$75x^3 + 190x = 10x^6 + 21$$

between two polynomials with positive integer coefficients. Call this $f(x) = g(x)$. For Liouville's number, such a polynomial has a value whose decimal expansion has long strings of zeros. For example, let's write

$$\begin{aligned} l_1 &= 0.1, \\ l_2 &= 0.11, \\ l_3 &= 0.110001, \\ l_4 &= 0.110001\ 000000\ 000000\ 000001, \end{aligned}$$

and generally l_n is the number obtained by stopping Liouville's number at the nth 1. Then, if $f(x)$ and $g(x)$ are the left and right sides of the above equation, the values of $f(l_n)$ and $g(l_n)$ are very good approximations to $f(l)$ and $g(l)$. In fact, for $n = 6, 7, 8, \ldots$, the decimal form of $f(l)$ consists of the $6 \times n!$ digits of $f(l_n)$, followed by a long string of zeros, and then some other digits; and in a similar way, the first $6 \times n!$ digits of $g(l)$ are those of $g(l_n)$. If l really *did* satisfy the equation

$$f(l) = g(l),$$

we would also have

$$f(l_6) = g(l_6), \quad f(l_7) = g(l_7), \quad f(l_8) = g(l_8), \quad \ldots$$

and so a sixth-degree polynomial with more than six roots. A similar contradiction arises from supposing that l satisfies any other algebraic equation with whole number coefficients. Therefore

Liouville's number is transcendental.

In fact, Liouville showed that any number that has an *extremely rapid* sequence of rational approximations is forced to be transcendental.

Now that we know that numbers *can* be transcendental, we return to the problem of π, and of speedy ways of calculating it. Until recently, most of these used

Gregory's Numbers

The man in Figure 9.2(a) is standing at the corner of a hairpin bend. The uphill road has slope 1 in 3. The downhill road has a 1 in 2 slope.

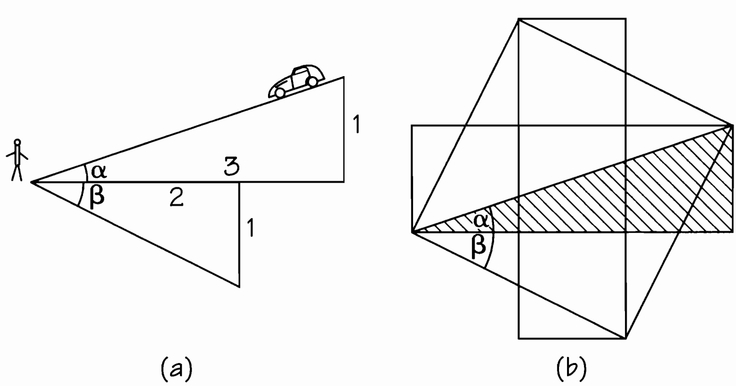


FIGURE 9.2 *Euler's relation between the Gregory numbers made easy.*

What's the total angle between them? You can find the answer very closely by using the arctan function on your pocket calculator. The 1 in 3 and 1 in 2 hills have angles as displayed below:

		degrees		radians
$\alpha = \arctan \frac{1}{3}$	=	18.43494882...	=	0.3217505544...
$\beta = \arctan \frac{1}{2}$	=	26.56505118...	=	0.4636476090...
Total $\alpha + \beta$	=	45.0	=	0.7853981634... $= \frac{\pi}{4}$

But Figure 9.2(b) makes it clear that the answer is in fact *exactly* 45°, which is the slope of a 1 in 1 hill. Let's use t_n for the angle in a 1 in n hill, expressed in radian measure. In this notation, we've just shown that

$$\boxed{t_1 = t_2 + t_3}$$

a formula known to Euler.

In 1671 the Scottish mathematician David Gregory found a nice formula for t_x:

$$t_x = \frac{1}{x} - \frac{1}{3x^3} + \frac{1}{5x^5} - \frac{1}{7x^7} + \frac{1}{9x^9} - \frac{1}{11x^{11}} + \cdots;$$

this has been used by many people in formulas for π.

We'll call t_x a **Gregory number** whenever x is any fraction or whole number. If you put $x = 1$ in Gregory's series, and use the fact that t_1, the angle of a 1 in 1 hill, is $\frac{\pi}{4}$ radians, you get **Leibniz's formula**

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + \cdots,$$

which, although very elegant, is much too slow for practical calculation. (You'd need tens of thousands of terms to get the answer correct to four decimal places.) We can speed things up using the formula $t_1 = t_2 + t_3$:

$$\frac{\pi}{4} = \frac{1}{2} - \frac{1}{3 \times 2^3} + \frac{1}{5 \times 2^5} - \frac{1}{7 \times 2^7} + \frac{1}{9 \times 2^9} - \frac{1}{11 \times 2^{11}} + \cdots$$

$$+ \frac{1}{3} - \frac{1}{3 \times 3^3} + \frac{1}{5 \times 3^5} - \frac{1}{7 \times 3^7} + \cdots .$$

The terms we've written here already suffice to give π to four decimal places.

Euler found several equations for π in terms of the t_n. In addition to $t_1 = t_2 + t_3$, he suggested the use of

$$t_1 = 2t_3 + t_7 \quad \text{and}$$

$$t_1 = 5t_7 + 2t_{18} - 2t_{57}.$$

He pointed out that if you use his formula

$$t_n = \frac{1}{\sqrt{m}}\left(1 + \frac{1}{2} \cdot \frac{1}{3m} + \frac{1 \times 3}{2 \times 4}\frac{1}{5m^2} + \frac{1 \times 3 \times 5}{2 \times 4 \times 6} \cdot \frac{1}{7m^3} + \cdots\right)$$

(where $m = n^2+1$) instead of Gregory's, then since

$$3^2+1 = 10, \quad 7^2+1 = 50, \quad 18^2+1 = 325, \quad 57^2+1 = 3250,$$

the calculations are quite easy in the decimal system; in fact, the calculations for t_{18} and t_{57} are almost identical.

In 1706 John Machin calculated π to 100 decimals using his formula

$$t_1 = 4t_5 - t_{239}$$

(notice that $239^2 + 1 = 2 \times 13^4$). Machin's formula is the best way to express π in terms of *two* of the t_n. In fact,

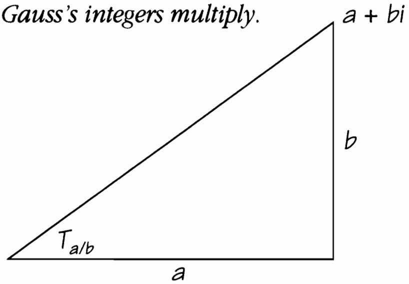
$$4t_1, \quad 4t_2 + 4t_3, \quad 8t_2 - 4t_7, \quad 8t_3 + 4t_7, \quad \text{and} \quad 16t_5 - 4t_{239}$$

are the *only* ways to express π as a combination of at most two of the t_n. This follows from Størmer's theory.

STØRMER INTRODUCES GAUSS TO GREGORY!

Carl Størmer discovered in 1896 that he could use Gauss's numbers, $a + bi$ (which we met in Chapter 8), to understand all the relations between the Gregory numbers, t_x, for arbitrary fractions or whole numbers x.

FIGURE 9.3 *Gregory's numbers add while Gauss's integers multiply.*



The angle $t_{a/b}$ of a hill of slope b in a is just the angle of the Gaussian number $a + bi$. These angles add the way Gauss's numbers multiply (Figure 9.3). For instance,

$$(2 + i)(3 + i) = 5 + 5i$$

so

$$t_2 + t_3 = t_{5/5} = t_1$$

(5 in 5 hills have the same slope as 1 in 1 hills).

We've seen in Chapter 8 that every Gaussian number can be written in essentially just one way as the product of Gauss's prime numbers:

$$1 \pm i, \quad 3, \quad 2 \pm i, \quad 7, \quad 11, \quad 3 \pm 2i, \quad 4 \pm i, \quad 19, \quad 23, \quad 5 \pm 2i, \quad 31, \ldots.$$

This tells us that any Gregory number $t_{a/b}$ is uniquely expressible in terms of the "prime" Gregory numbers

$$t_{1/1}, t_{2/1}, t_{3/2}, t_{4/1}, t_{5/2}, t_{6/1}, t_{5/4}, \ldots$$

for which $a^2 + b^2$ is a prime number and $a/b \geq 1$. (These come from the complex Gaussian primes, $a + bi$,

$$1+i, \quad 2+i, \quad 3+2i, \quad 4+i, \quad 5+2i, \quad 6+i, \quad 5+4i, \ldots.$$

The conjugate ones, $a - bi$, just give the negatives of these angles, and the *real* Gaussian primes, 3, 7, 11, . . . , all give angle zero.) For example, $5 + i$ factors as $(1 + i)(3 - 2i)$ and so t_5 is a "composite" Gregory number, $t_5 = t_1 - t_{3/2}$.

STØRMER'S NUMBERS

Størmer's numbers are the positive whole numbers n for which the largest prime factor, p, of $n^2 + 1$ is at least $2n$. Størmer found that every Gregory number t_x can be uniquely expressed as a sum in terms of the t_n for which n is a Størmer number.

To find Størmer's decomposition for $t_{a/b}$, you repeatedly multiply $a + bi$ by numbers $n \pm i$ for which n is a Størmer number and the sign is chosen so that you can cancel the corresponding prime number p (n is the smallest number for which $n^2 + 1$ is divisible by p).

We'll do t_{70}. We find $70^2 + 1 = 13 \times 29$, so our first Størmer number will be 12, since $12^2 + 1 = 5 \times 29$. Now

$$(70 + i)(12 - i) = 29(29 - 2i),$$

and we continue in the same way:

$$(29 - 2i)(5 - i) = 13(11 - 3i)$$

$$(11 - 3i)(5 - i) = 26(2 - i)$$

so that

$$t_{70} - t_{12} - t_5 - t_5 = -t_2$$

or

$$t_{70} = t_{12} + 2t_5 - t_2.$$

Figure 9.4 illustrates this process geometrically and Table 9.1 lists the first 30 Størmer numbers, n, with their corresponding primes, p.

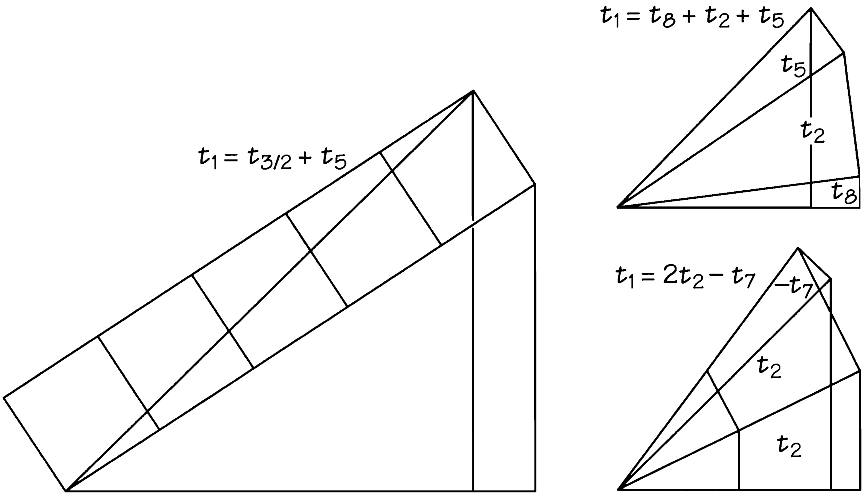


FIGURE 9.4 *Relations between Størmer numbers by multiplying Gauss's integers.*

n	p	n	p	n	p	n	p	n	p
1	2	10	101	19	181	26	617	35	613
2	5	11	61	20	401	27	73	36	1297
4	17	12	29	22	97	28	157	37	137
5	13	14	197	23	53	29	421	39	761
6	37	15	113	24	577	33	109	40	1601
9	41	16	257	25	313	34	89	42	353

TABLE 9.1 *The first 30 Størmer numbers with their corresponding primes.*

Table 9.2 expresses the Gregory numbers, t_n, for which n *isn't* a Størmer number in terms of those for which it is. The entries in this table are easily checked using Lewis Carroll's observation that

$$t_n = t_{n+c} + t_{n+d}$$

holds just if $cd = n^2 + 1$. For instance, $100^2 + 1 = 73 \times 137$, so $t_{100} = t_{173} + t_{237}$. You can use Table 9.2 to find other formulas for π, for instance, Størmer's *own* formula

$$\frac{\pi}{4} = t_1 = 12t_8 + 8t_{57} - 5t_{239}.$$

$t_3 = t_1 - t_2$
$t_7 = -t_1 + 2t_2$
$t_8 = t_1 - t_2 - t_5$
$t_{13} = t_1 - t_2 - t_4$
$t_{17} = -t_1 + 2t_2 - t_{12}$
$t_{18} = t_1 - 2t_2 + t_5$
$t_{21} = t_4 - t_5$
$t_{30} = t_1 - t_2 - t_4 - t_{23}$
$t_{31} = t_5 - t_6$
$t_{32} = -t_1 + 2t_2 - t_9$
$t_{38} = -t_2 + 2t_4$
$t_{41} = t_1 - 2t_2 + 2t_{12}$
$t_{43} = t_1 - 2t_2 + t_6$
$t_{46} = -t_1 + 2t_2 - t_{12} - t_{27}$
$t_{47} = t_2 - t_4 - t_5$
$t_{50} = t_9 - t_{11}$
$t_{55} = t_4 - t_5 - t_{34}$
$t_{57} = -2t_1 + 3t_2 + t_5$
$t_{68} = 2t_1 - 3t_2 - t_6$
$t_{70} = -t_2 + 2t_5 + t_{12}$
$t_{72} = -t_1 + t_2 + t_4 + t_{11}$
$t_{73} = t_1 - t_2 - t_5 - t_9$
$t_{75} = -t_1 + 2t_2 - t_{12} - t_{22}$
$t_{76} = t_{23} - t_{33}$
$t_{83} = t_1 - 2t_2 + t_5 - t_{23}$
$t_{91} = t_9 - t_{10}$
$t_{93} = t_1 - 2t_2 + t_6 - t_{80}$
$t_{98} = t_1 - t_2 - t_4 - t_{15}$
$t_{99} = t_1 - t_2 - 2t_5 + t_{12}$
$t_{100} = t_{27} - t_{37}$
$t_{105} = t_5 - t_6 - t_{44}$
$t_{111} = t_{10} - t_{11}$
$t_{112} = t_2 - t_4 - t_5 - t_{81}$
$t_{117} = -t_1 + t_2 + 2t_6$
$t_{119} = t_{22} - t_{27}$
$t_{122} = -2t_1 + 3t_2 + t_5 - t_{107}$
$t_{123} = -t_2 + t_4 + t_5 + t_{34}$
$t_{128} = t_1 - 2t_2 + t_{12} + t_{15}$
$t_{129} = t_{23} - t_{28}$
$t_{132} = -t_1 + 2t_2 - t_4 + t_9$
$t_{133} = t_{11} - t_{12}$
$t_{142} = -t_1 + 2t_2 - t_6 + t_{33}$
$t_{144} = t_4 - t_5 - t_{34} - t_{89}$
$t_{155} = t_1 - t_2 - t_5 - t_9 - t_{138}$
$t_{157} = -t_1 + t_2 + t_4 + t_{12}$
$t_{162} = -t_1 + 2t_2 - t_{12} - t_{19}$
$t_{172} = -t_1 + 2t_2 - t_{11} - t_{22}$
$t_{173} = t_1 - 2t_2 + t_9 + t_{27}$
$t_{174} = -t_2 + t_4 + t_5 + t_{37}$
$t_{177} = t_2 - t_4 - t_5 - t_{64}$
$t_{182} = -3t_1 + 5t_2 + t_{23}$
$t_{183} = t_1 - t_2 - t_4 - t_{14}$
$t_{185} = t_{28} - t_{33}$
$t_{187} = -2t_1 + 3t_2 + t_5 - t_{82}$
$t_{189} = t_1 - 2t_2 + t_5 - t_{23} - t_{148}$
$t_{191} = -t_1 + t_2 + t_4 + t_6 - t_{12}$
$t_{192} = -t_1 + 2t_2 - t_{10} - t_{27}$
$t_{193} = 2t_1 - 3t_2 - t_5 + t_{44}$
$t_{200} = -t_4 + t_5 + t_{19}$
$t_{203} = t_1 - t_2 - t_5 - t_9 - t_{114}$
$t_{211} = t_{14} - t_{15}$
$t_{212} = -t_1 + t_2 + t_5 + t_{10} + t_{34}$
$t_{213} = t_1 - 2t_2 + t_5 - t_{23} - t_{136}$
$t_{216} = t_1 - t_2 - t_5 - t_6 + t_{22}$
$t_{217} = t_2 - t_4 - t_5 - t_{60}$
$t_{228} = t_1 - 2t_2 + t_6 - t_{53}$
$t_{233} = t_1 - t_2 - t_5 - t_{11} - t_{34}$
$t_{237} = -t_1 + 2t_2 - t_9 - t_{37}$
$t_{239} = -t_1 + 4t_5$

TABLE 9.2 *Gregory numbers t_n for which n is not a Størmer number.*

Bernt Klostermark has produced many such formulas, such as

$$\frac{\pi}{4} = 44t_{57} + 24t_{1068} + 12t_{3001/3} - 5t_{239}.$$

Now that we know so much about π, we're ready for some other transcendental numbers. Of course, π is the best-known transcendental number, but there are many more, notably the values of the logarithm and exponential functions. It's rather surprising that Lindemann's proof that π is transcendental actually makes use of these notions.

LOGARITHMS

Electronic calculators and computers have almost made logarithms obsolete as a practical method of calculation, but they're still very much alive in mathematics as a whole. We hope you remember their important property that

$$\log(a \times b) = \log a + \log b.$$

THE AGE OF DISCOVERY

The discovery of logarithms was a great boon to seventeenth-century navigators, astronomers, and other calculators. Sailors, who now sailed across broad oceans, required accurate navigation. Astronomers made heavy calculations to test their theories, and in commerce numbers were steadily getting bigger. All these people welcomed the new tables of logarithms, which miraculously turned multiplication into addition.

COMMON LOGARITHMS

The common logarithms that you find in tables or on your pocket calculator are the powers to which 10 must be raised to yield the given numbers. Thus

number	⋯	$\frac{1}{100}$	$\frac{1}{10}$	1	10	100	1000	10000	⋯
logarithm	⋯	−2	−1	0	1	2	3	4	⋯

We think that the difficulty most people have in understanding logs is that most of them are irrational. The common logarithm of 2, for instance, is

$$0.301029996\ldots.$$

How do we know that this is irrational? The answer is very easy! If log 2 were the rational fraction p/q, then we should have

$$10^{p/q} = 2,$$

and so, raising each side to the qth power,

$$10^p = 2^q$$

for two positive integers p and q. This is obvious nonsense. The candidates 10, 100, . . . for the left-hand side all end in zero, while those on the right, 2, 4, 8, 16, 32, 64, . . . all end in 2, 4, 8, or 6.

It is even easier to see that log 3 is irrational, because the powers of 3 are all odd. This kind of argument is quite general and proves that $\log_b a$ is irrational whenever a and b are not perfect powers of the same integer.

Here we've written $\log_b a$ for "the logarithm of a to the base b." We can compute logarithms to any base:

> $L = \log_b a$ is the power (exponent) to which the base b must be raised to give the number a.
>
> $$b^L = a$$
>
> $$(\text{base})^{\text{logarithm}} = \text{number}.$$

The only logarithms to base 10 that are very easily computed are those of the powers of 10 (see the table at the beginning of this section), but these are very rapidly increasing. We can get a denser sequence by using base 2, and still denser ones by using bases 1.1, 1.01, 1.001, ... For instance, with base 1.001, it's fairly easy to multiply by 1.001 and find

number	1	1.001	1.002001	1.003003	1.010045	. . .
$\log_{1.001}$	0	1	2	3	10	. . .

1.999013	2.001012	. . .	2.999516	3.002515
693	694	. . .	1099	1100

so that $\log_{1.001} 2$ is roughly $693\frac{1}{2}$ and $\log_{1.001} 3$ is roughly $1099\frac{1}{5}$. Varying the base, we find an interesting pattern:

base b	1.1	1.01	1.001	. . .	1.000001	1.0000001
$\log_b 2$	7.27	69.66	693.49	. . .	693147.53	6931472.15
$\log_b 3$	11.5	110.41	1099.16	. . .	1098612.84	10986123.44

It seems that, for comparison, one should divide these logarithms by 10, 10^2, 10^3, . . . , which is equivalent to using the bases:

base b	1.1^{10}	1.01^{10^2}	1.001^{10^3}	1.0001^{10^4}	. . .	1.0000001^{10^7}
i.e.	2.594	2.7048	2.71692	2.718146	. . .	2.718281693
$\log_b 2$	0.727	0.6966	0.69349	0.693182	. . .	0.693147215
$\log_b 3$	1.15	1.1041	1.09916	1.098667	. . .	1.098612344

The limiting numbers in these rows are

2.71828 18284 59045 23536 . . . , base of *natural logarithms*,
0.69314 71805 59945 30941 . . . , the *natural logarithm* of 2,
1.09861 22886 68109 69139 . . . , the *natural logarithm* of 3.

The base of the natural logarithms is **Napier's number**,

$$e = 2.71828\ 18284\ 59045\ 23536\ 02874\ \ldots .$$

It is the approximate value of $(1+1/N)^N$ for very large N, and the number whose natural logarithm is 1. In mathematics, the natural logarithms are defined as areas under the graph of $1/x$ (Figure 9.5).

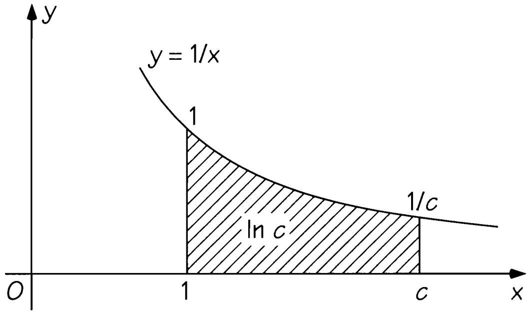


FIGURE 9.5 *The shaded region under the graph* y = 1/x *is ln* c.

This makes the logarithmic property very easy to see. Stretch the shaded region in Figure 9.5 so that it's b times as wide and at the same time squash it vertically in the same ratio so that it becomes the shaded region of Figure 9.6. This process hasn't changed the area, so we can add the cross-hatched area to obtain

$$\boxed{\ln b + \ln c = \ln bc}$$

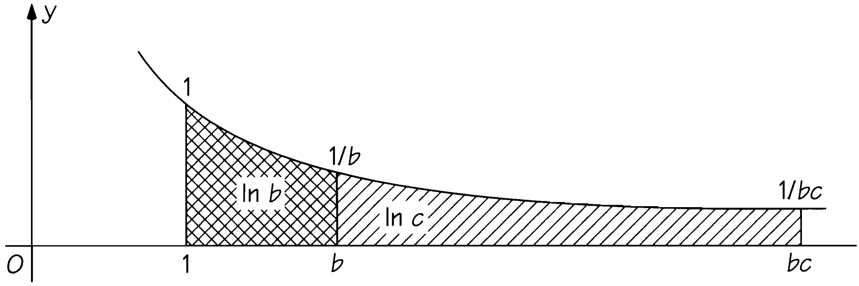


FIGURE 9.6 *Illuminating the logarithmic law.*

POWERS OF *e*

Powers of Napier's number can be found approximately by replacing e by $\left(1 + \frac{1}{N}\right)^N$ for some large N: the larger the value of N, the better the approximation. For instance, e^5 is approximately

$$\left(1 + \frac{1}{2000}\right)^{2000\times5} = \left(1 + \frac{5}{10000}\right)^{10000}$$

and in general, e^x is approximately $\left(1 + \frac{x}{N}\right)^N$ for large N. By the binomial theorem, $\left(1 + \frac{x}{100}\right)^{1000}$

$$= 1+\frac{1000}{1!}\left(\frac{x}{1000}\right)+\frac{1000\times999}{2!}\left(\frac{x}{1000}\right)^2+\frac{1000\times999\times998}{3!}\left(\frac{x}{1000}\right)^3+\cdots$$

$$= 1+\frac{x}{1!}+0.999\,\frac{x^2}{2!}+0.999\times0.998\,\frac{x^3}{3!}+\cdots.$$

Taking larger and larger values of N, we approach the ultimately correct formula

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.$$

In particular, Napier's number has the elegant form

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots$$

from which it's very easy to calculate e with considerable accuracy:

1.00000 00000,	divide by 1
1.00000 00000,	divide by 2
0.50000 00000,	divide by 3
0.16666 66667,	divide by 4
0.04166 66667,	divide by 5
0.00833 33333,	divide by 6
0.00138 88889,	divide by 7
0.00019 84127,	divide by 8
0.00002 48016,	divide by 9
0.00000 27557,	divide by 10
0,00000 02756,	divide by 11
0.00000 00251,	divide by 12
0.00000 00021,	divide by 13
0.00000 00002,	and add up
2.71828 18286	

with an accumulated error of 1 in the tenth decimal place.

Is e TRANSCENDENTAL?

We ask first whether e is *rational*. We see that $7e$ is roughly 19.028, so that 19/7 is quite a good approximation for e. Could it be that $7e$ was *exactly* 19? No! For if so, we'd have

$$\frac{19}{7} = e = 1+\frac{1}{1!}+\frac{1}{2!}+\frac{1}{3!}+\frac{1}{4!}+\frac{1}{5!}+\frac{1}{6!}+\frac{1}{7!}+\frac{1}{8!}+\frac{1}{9!}$$
$$+ \text{ a known error.}$$

All the fractions you can see can be written to the common denominator $9! = 362880$, so the error, which is positive, would be at least $1/362880 = 1/9!$ But in fact the error is exactly

$$\frac{1}{10!} + \frac{1}{11!} + \frac{1}{12!} + \cdots,$$

which is much less than $1/9!$. The point is that by stopping the series at some term we obtain close approximations, n/d, in which the error, while nonzero, has size strictly less than $1/d$, *and* we can arrange that d is divisible by any given number.

In a similar way we can show that e is not equal to any quadratic irrationality. If, for instance, e were exactly equal to $(5 + \sqrt{10})/3$, it would satisfy the equation $3e^2 - 10e + 5 = 0$, and $3e + (5/e)$ would equal the rational number 10. This is disproved using the close approximations obtained from the series

$$3e + \frac{5}{e} = \frac{3+5}{0!} + \frac{3-5}{1!} + \frac{3+5}{2!} + \frac{3-5}{3!} + \frac{3+5}{4!} + \cdots.$$

The same method also shows that $e^{\sqrt{2}}$ and $e^{i\sqrt{2}}$ are irrational. The real part of $e^{i\sqrt{2}}$ is given by the series

$$1 - \frac{2}{2!} + \frac{2^2}{4!} - \frac{2^3}{6!} + \frac{2^4}{8!} - \frac{2^5}{10!} + \cdots,$$

where the powers of 2 cancel exactly, so we do indeed get close approximations in our precise sense by stopping at any term. Note that e^x is defined even for a complex x, and indeed one of the most famous relations in mathematics is Euler's formula $e^{i\pi} = -1$, which we'll prove in a moment. This gives us a hint as to how π was proved to be transcendental. If we could prove, for example, that $e^{i\sqrt{10}}$ were irrational, then this would show that π could not be equal to Brahmagupta's value $\sqrt{10}$, since then $e^{i\sqrt{10}}$ would be -1.

In 1873 Charles Hermite (1822–1901) found clever ways to get close approximations to arbitrary sums of integer powers of e and hence was able to show that e was transcendental. In 1882 Ferdinand Lindemann extended the method to cope with *algebraic* powers of e and proved that π was transcendental. He was able to establish the transcendence of all numbers of the shape αe^{β} where α, β are nonzero algebraic numbers, and also of sums of such numbers with different values of β.

In the 1960s Alan Baker revolutionized transcendence theory by finding *effective* close approximations for sums of natural logarithms of algebraic numbers. In particular he proved a transcendence theorem for sums of numbers of the form $\alpha \ln \beta$ where the α and β are algebraic.

This implies the 1934 theorem of Gelfond and Schneider that $\sqrt{2}^{\sqrt{2}}$ is transcendental. More generally, α^{β} is transcendental if α is an algebraic number other than 0 or 1 and β is an irrational algebraic number.

Lindemann's and Baker's theorems imply that the standard functions (for algebraic numbers b and x)

$$\cos x, \quad \sin x, \quad \tan x, \quad \log_b x, \quad \ln x, \quad e^x$$

take transcendental values except at some very obvious places. In particular, our Gregory numbers $t_{a/b}$ are transcendental, since t_n is the imaginary part of $\ln(a + ib)$.

The definition of e^x as the limit of $\left(1 + \frac{x}{N}\right)^N$ works also for complex numbers x.

EULER'S WONDERFUL RELATION

This equation

$$\boxed{e^{i\pi} + 1 = 0}$$

appears in Euler's *Introductio*, published in Lausanne in 1748. It is justly celebrated as one of the most remarkable identities in all of mathematics.

For the thought will console (as it jolly well ought)
that it's $e^{\pi i} + 1 = 0$

"Diogenes"

However, Euler's equation is no longer mysterious. It means just that

$$\left(1 + \frac{i\pi}{N}\right)^N$$

gets closer and closer to -1 as N gets larger and larger. The idea here is that the triangle on the right side of Figure 9.7, whose top vertex

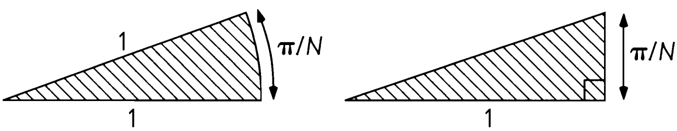


FIGURE 9.7 *The sector and the triangle become more nearly equal.*

is at $1 + \dfrac{i\pi}{N}$, is very close in shape to the circular section on the left of Figure 9.7, in which the length of the circular arc is π/N: N copies of this sector fit together to make an exact semicircle, as in Figure 9.8(a).

On the other hand, our twirling rule for multiplying complex numbers tells us that

$$\left(1 + \frac{i\pi}{N}\right)^N$$

is the point obtained by juxtaposing N triangles of the same shape and slowly increasing size, as in Figures 9.8(b), (c), and (d), where we illustrate $N = 6$, 12, and 24. As N gets larger, the figures get closer to the semicircular arrangement of Figure 9.8(a).

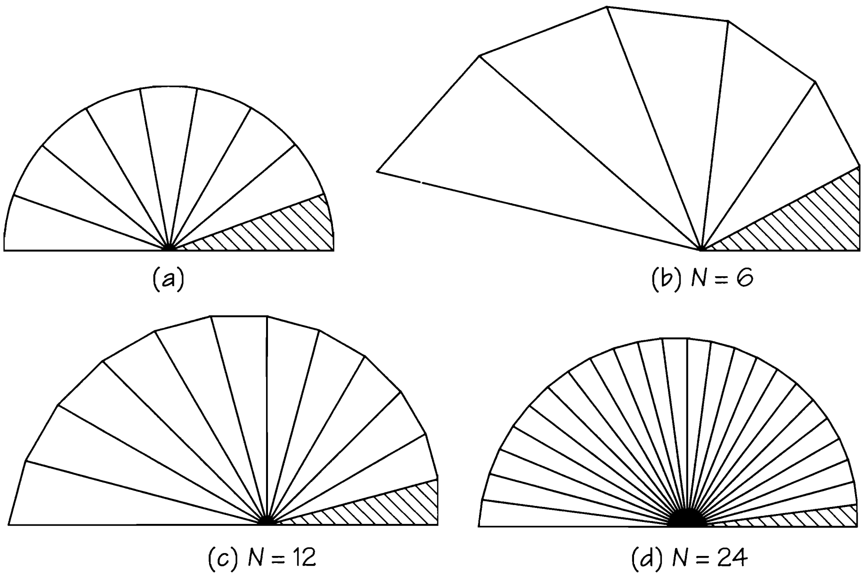


FIGURE 9.8 *Why* $e^{\pi i} = -1$.

HARMONY, FRACTIONS, AND LOGARITHMS

Among the many things attributed to the Pythagorean brotherhood is the discovery that harmonious musical notes are produced when the ratios of the lengths of the column of air, or of the vibrating strings, are simple fractions.

If the ratio is $\frac{2}{1}$, the notes are an **octave** apart; the ratio $\frac{3}{2}$ gives an interval called a **fifth** (the standard musical names come from counting, inclusively, the white notes on a piano; see Figure 9.9).

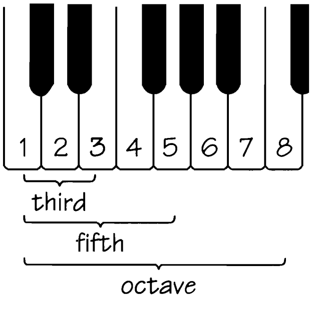


FIGURE 9.9 *Counting notes inclusively.*

The **pitch** of a note, measured in octaves above a certain reference point (say middle C, with frequency 2^9/sec), is just the logarithm to base 2 of its frequency. What does this have to do with logs? Obviously, if two notes are n octaves apart, then the ratio of their frequencies is 2^n, or the difference of their logs to base 2 is exactly n. We can apply this even when n isn't a whole number. Since $\log_2 3$ (log 3 to base 2) is 1.5849625007 . . . , two notes whose frequencies differ by a factor of 3 are 1.5849625007 . . . octaves apart. Two notes at an interval of one fifth are therefore 0.5849625007 . . . octaves apart.

> A **fifth** is
> 0.5849625007 . . . .
> of an octave.

We can use the continued fractions method (Chapter 6) to find that

roughly 7 fifths make 4 octaves,
12 fifths make 7 octaves,
41 fifths make 24 octaves,
53 fifths make 31 octaves.

Is any whole number of fifths precisely equal to a whole number of octaves? This is the same as asking is $\log_2 3$ a rational number? No! If so, we would have $2^a = 3^b$ for some positive whole numbers a and b, which is clearly impossible.

We know that Pythagoras (or his brotherhood) discovered the existence of irrational numbers and also the arithmetic meaning of harmony. We also know that he wondered about the relation between the lengths of fifths and octaves. The "comma of Pythagoras" is the difference between 7 fifths and 4 octaves. It is tempting to speculate that Pythagoras might also have realized that this ratio is irrational.

The mathematical uses of the word "harmonic" (as in our next section) ultimately stem from the Pythagorean tradition, but it would be tedious to explain in detail just how.

HARMONIC NUMBERS

The **harmonic numbers** are $H_1 = 1$, $H_2 = 1 + \frac{1}{2}$, $H_3 = 1 + \frac{1}{2} + \frac{1}{3}$, and, more generally,

$$H_n = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}.$$

No harmonic number is a whole number after the first one: look at the term with the highest power of 2 in it. It has nothing with which to pair. So $H_2, H_3, H_4, \ldots$ have odd numerator and even denominator.

SECOND, THIRD, . . . , HARMONICS

You might expect that expressions such as $H_1+H_2+H_3+\cdots+H_n$ would lead to some new kind of "hyperharmonic number," but in fact they are quite easily worked out in terms of the ordinary harmonic numbers. For example,

$H_1 + H_2 + \cdots + H_7 = \frac{7}{1} + \frac{6}{2} + \frac{5}{3} + \frac{4}{4} + \frac{3}{5} + \frac{2}{6} + \frac{1}{7}$, and if you add 8 to this in the form $\frac{1}{1} + \frac{2}{2} + \frac{3}{3} + \frac{4}{4} + \frac{5}{5} + \frac{6}{6} + \frac{7}{7} + \frac{8}{8}$, you get $8H_8$.

In general, the **second harmonic number**, $H_n^{(2)}$, is

$$H_1 + H_2 + \cdots + H_n = (n+1)(H_{n+1} - 1) = (n+1)(H_{n+1} - H_1)$$

and the **third harmonic number**, $H_n^{(3)}$, is

$$H_1^{(2)} + H_2^{(2)} + \cdots + H_N^{(2)} = \binom{n+2}{2}(H_{n+2} - H_2),$$

and the kth harmonic number is

$$H_n^{(k)} = \binom{n+k-1}{k-1}(H_{n+k-1} - H_{k-1}).$$

HOW BIG IS THE NTH HARMONIC NUMBER?

One answer is "about one nth of the nth prime." For example, the 60th prime is 281 and $281/60 = 4.6833\ldots$, while

$$1+\tfrac{1}{2}+\tfrac{1}{3} + \cdots + \tfrac{1}{60} = 4.6798\ldots.$$

We know that the size of the nth prime is roughly $n \ln n$, so another answer is "roughly $\ln n$." The real answer is rather surprising! H_n is very close to

$$\ln n + 0.5772156 + \frac{1}{2n}.$$

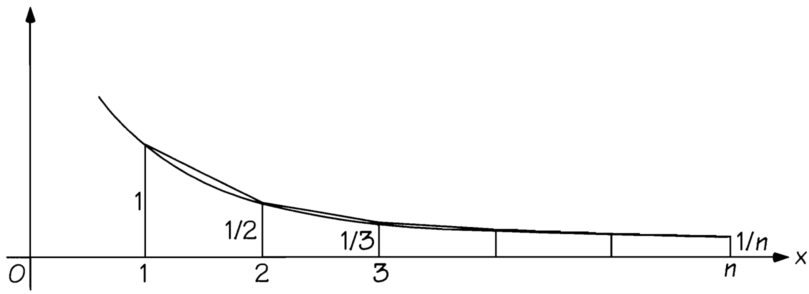


FIGURE 9.10 *An area slightly bigger than ln* n.

To see why this is so, we use the fact that $\ln n$ is the shaded area in Figure 9.5 (with $c = n$). This is just a bit less (roughly 0.077) than the polygonal area in Figure 9.10, which is made of $n - 1$ pieces, whose areas are the averages of

$$1 \text{ and } \frac{1}{2}, \quad \frac{1}{2} \text{ and } \frac{1}{3}, \quad \frac{1}{3} \text{ and } \frac{1}{4}, \quad \ldots, \frac{1}{n-1} \text{ and } \frac{1}{n},$$

So the total area is

$$\frac{1}{2}\left(1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{3} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{n-1} + \frac{1}{n}\right)$$

$$= 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} - \frac{1}{2}\left(1 + \frac{1}{n}\right)$$

$$= H_n - \frac{1}{2} - \frac{1}{2n}$$

and H_n is about $\ln n + \frac{1}{2} + \frac{1}{2n} + 0.077$.

THE EULER–MASCHERONI NUMBER

The number

$$\gamma = 0.577215664901532860606512090082402431042159335939 92\ldots$$

that appears here is one of the most mysterious of all arithmetic constants. Some people credit it to Mascheroni and others to Euler, so we'll call it the **Euler–Mascheroni number**.

The Euler-Mascheroni number appears unexpectedly in several places in number theory. In 1838 Dirichlet proved that the average number of divisors of all the numbers from 1 to n is very close to

$$\ln n + 2\gamma - 1.$$

For example, the total number of divisors of the first 250 numbers is 1421, so the average number is 5.684, while $\ln 250+2\gamma-1$ is about 5.676.

In 1898 de la Vallée Poussin proved that if a large number n is divided by all the primes up to n, then the average fraction by which the quotient falls short of the next whole number is γ, and not ½ as you might expect! For example, if we divide 43 by 2, 3, 5, 7, . . . , 41, then the answers $21\frac{1}{2}$, $14\frac{1}{3}$, $8\frac{3}{5}$, $6\frac{1}{7}$, . . . , $1\frac{2}{21}$ fall short of 22, 15, 9, 7, . . . , 2 by $\frac{1}{2}$, $\frac{2}{3}$, $\frac{2}{5}$, $\frac{6}{7}$, $\frac{1}{11}$, $\frac{9}{13}$, $\frac{8}{17}$, $\frac{14}{19}$, $\frac{2}{23}$, $\frac{15}{29}$, $\frac{19}{31}$, $\frac{31}{37}$, $\frac{39}{41}$, whose average is 0.57416 . . . .

Euler's totient number, the number of proper fractions with denominator exactly n (see Chapter 6), can never get much smaller than

$$\frac{n}{e^{\gamma} \ln \ln n}$$

although it often gets about this small.

STIRLING'S FORMULA

The fact that $1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}$ is roughly equal to $\ln n + \gamma + \frac{1}{2n}$ is one instance of the **Euler–Maclaurin sum formula**, which often gives good estimates for the sums of values of functions. The most famous application is

STIRLING'S APPROXIMATE FORMULA FOR N!

$n!$ is roughly equal to $\sqrt{2\pi n}\,(n/e)^n$

for $\ln n! = \ln 1 + \ln 2 + \cdots + \ln n$ is roughly equal to

$$H_1 + H_2 + \cdots + H_n - n\gamma - \frac{1}{2} Hn$$

which we have seen is $(n + \frac{1}{2})H_n - n - n\gamma$, or, roughly,

$$(n + \tfrac{1}{2})(\ln n + \gamma) - n - n\gamma \doteqdot (n + \tfrac{1}{2}) \ln n - n,$$

so $n!$ is roughly $n^{n+1/2}e^{-n}$.

It's strange that the formula involves both Napier's number, e, and Ludolph's number, π.

As an example, try $n = 8$. $\sqrt{16\pi}\,(8/e)^8 \approx 39902$, which is short of the exact answer, $8! = 40320$, by about one part in a hundred. A more exact form of Stirling's formula augments it proportionally by the fraction $1/12n$: in the present example to 40318.04 . . . .

THE GREAT ENIGMA

Mathematicians are still wrestling with the numbers not covered by Lindemann's and Baker's theorems. For all we knew until quite recently, **Apéry's number**,

$$\begin{aligned}\zeta(3) &= 1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \cdots \\ &= 1.20205\ 69031\ 59594\ 28539\ 97381\ 61511\ 44999\ 07649 \ldots\end{aligned}$$

might have been rational, but we've named it after Apéry because he recently found quite a simple proof that it is not. However, we still don't know if it's transcendental, and we don't know anything about

$$\begin{aligned}\zeta(5) &= 1 + \frac{1}{2^5} + \frac{1}{3^5} + \frac{1}{4^5} + \cdots \\ &= 1.03692\ 77551\ 43369\ 92633\ 13654\ 86457\ 03416\ 80570\ldots, \\ \zeta(7) &= 1 + \frac{1}{2^7} + \frac{1}{3^7} + \frac{1}{4^7} + \cdots \\ &= 1.00834\ 92773\ 81922\ 82683\ 97975\ 49849\ 79675\ 95998\ldots,\end{aligned}$$

etc. However,

$$\begin{aligned}\zeta(2) &= 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6} \\ &= 1.64493\ 40668\ 48226\ 43647\ 24151\ 66646\ldots, \\ \zeta(4) &= 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \cdots = \frac{\pi^4}{90} \\ &= 1.08232\ 32337\ 11138\ 19151\ 60036\ 96541\ldots, \\ \zeta(6) &= 1 + \frac{1}{2^6} + \frac{1}{3^6} + \frac{1}{4^6} + \cdots = \frac{\pi^6}{945} \\ &= 1.01734\ 30619\ 84449\ 13971\ 45179\ 29790\ldots,\end{aligned}$$

are all transcendental.

The greatest enigma is the Euler-Mascheroni number, γ. Nobody has shown that it cannot be rational. We're prepared to bet that it is transcendental, but we don't expect to see a proof during our lifetime.

REFERENCES

Alan Baker. Approximations to the logarithms of certain rational numbers. *Acta Arith.*, **10** (1964): 315–323; *MR* **30** #1976.

Alan Baker. Linear forms in the logarithms of algebraic numbers, I, II, III, IV. *Mathematika*, **13** (1966): 204–216; **14** (1967): 102–107; **14** (1967): 220–228; **15** (1968): 204–216; *MR* **36** #3732; **41** #3402.

W.W. Rouse Ball and H.S.M. Coxeter. *Mathematical Recreations & Essays*, 12th edition. Univ. of Toronto Press, 1974, 347–359.

G.L. Dirichlet. *J. für Math.*, **18** (1838): 273.

Leonhard Euler. *Introductio in Analysin Infinitorum.* Lausanne, 1748.

Ch. Hermite. Sur la fonction exponentielle. *C. R. Acad. Sci. Paris,* **77** (1873): 18–24, 74–79, 226–233.

Ferdinand Lindemann. Ueber die Zahl π. *Math. Ann.,* **20** (1882): 213–225.

J. Liouville. Sur des classes très étendues de quantités dont la valeur n'est ni algébrique, ni même reductible à des irrationelles algébriques. *C. R. Acad. Sci. Paris,* **18** (1844): 883–885, 910–911; *J. Math. pures appl.,* (1), **16** (1851): 133–142.

T.H. O'Beirne. *Puzzles and Paradoxes.* Oxford University Press London, 1965, 195–197.

John R. Pierce. *The Science of Musical Sound.* Sci. Amer. Library, W.H. Freeman, New York, 1983.

Alfred J. van der Poorten. A proof that Euler missed . . . Apéry's proof of the irrationality of $\zeta(3)$. *Math. Intelligencer,* **1** (1978/79): 195–203; *MR* **80i**: 10054.

Ch.-J. de la Vallée Poussin. *Annales de la Soc. Sci. Bruxelles,* **22** (1898): 84–90.

10

Infinite and Infinitesimal Numbers

Sierpiński's Luggage

Wacław Sierpiński, the great Polish mathematician, was very interested in infinite numbers. The story, presumably apocryphal, is that once when he was travelling, he was worried that he'd lost one piece of his luggage. "No, dear!" said his wife, "All six pieces are here." "That can't be true," said Sierpiński, "I've counted them several times: zero, one, two, three, four, five."

Counting from Zero

Many mathematicians prefer to count objects starting at "zero" rather than "one." Although this may be unfamiliar, it is really a much simpler method. In fact, unless we teach them otherwise, machines tend to do it without thinking: One thousand cloakroom tickets will probably be numbered 000 to 999. If you count this way, then the number of objects you've counted is the earliest number that you *didn't* use, rather than the latest one you *did* (see Figure 10.1).

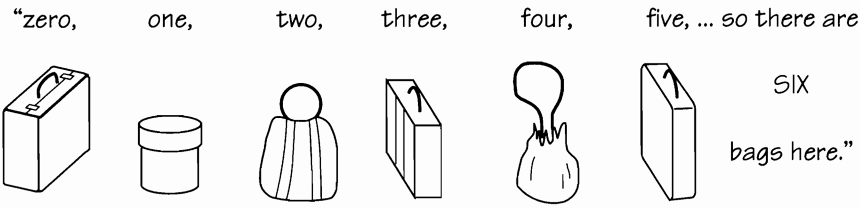


FIGURE 10.1 *What Sierpiński should have said.*

THE EMPTY SET

One of the advantages of the new system is that it works even when you are counting no objects at all. If Sierpiński's luggage all gets lost en route, then, at the other end of his journey he should say:

" ", so there are ZERO bags here!

The usual system of counting doesn't work for counting zero objects, since there isn't a last number that you used.

CANTOR'S ORDINAL NUMBERS

The great German mathematician Georg Cantor was the earliest person to construct a coherent theory of counting collections that may be infinite. For this he extended the ordinary series of numbers used for counting, as follows:

0, 1, 2, . . . as usual,
then ω, $\omega+1$, $\omega+2$, . . . then $\omega+\omega$, $\omega+\omega+1$, . . .

and so on.

The important point about these numbers (and, in essence, their definition) is that, no matter how many of them you've used, there's always a (uniquely determined) earliest one that you haven't. Cantor's opening infinite number,

$$\omega = \{0,\ 1,\ 2,\ \ldots |\ \}$$

is defined to be the earliest number greater than all the finite counting numbers. We'll use

$$\{a,\ b,\ c,\ \ldots |\ \}$$

for the earliest ordinal number after a, b, c, ... The vertical bar signals the place where we've cut off the number sequence a, b, c, . . . , for example,

$$\{0, 1, 2|\ \} = 3, \{0, 1, 2, \ldots|\ \} = \omega, \{0|\ \} = 1,$$
$$\{\ |\ \} = 0, \{0, 1, 2, \ldots \omega|\ \} = \omega + 1.$$

To avoid inventing lots of new words, the symbols $\omega + 1$, $\omega + 2$, . . . are used as proper names for the ordinary numbers following ω, just as "hundred and one" is the proper name of the number you get by adding "a hundred" and "one."

When you count things, you are really ordering them in a special way:

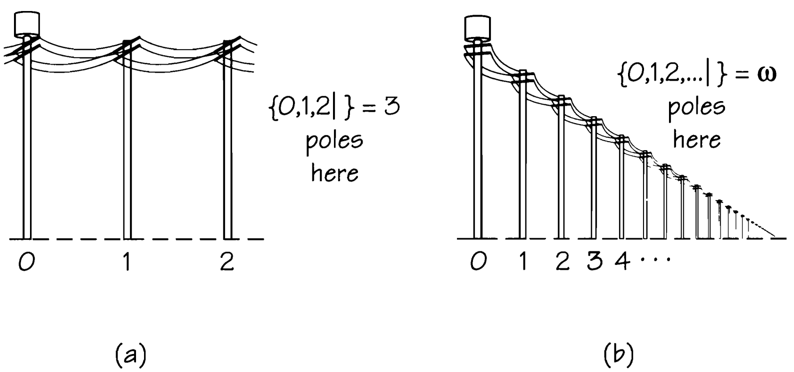
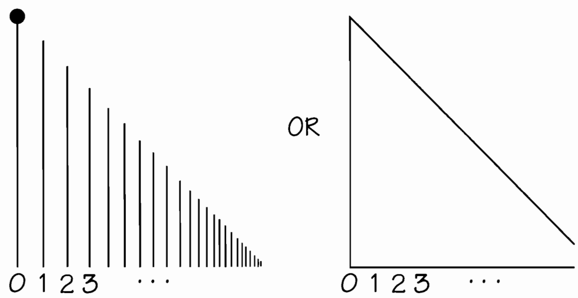


FIGURE 10.2 *Various numbers of poles.*

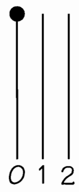
To count the poles in Figure 10.2(a), you'd say, "0, 1, 2, so there are $\{0, 1, 2|\}$ = 3 poles here." But now look at Figure 10.2(b), where we imagine that the road is infinite, with a pole for each of the ordinary integers 0, 1, 2, ... Obviously, we should now say:

"0, 1, 2, . . . , so there are ω poles here."

In the future, we'll represent such an infinite sequence of objects by

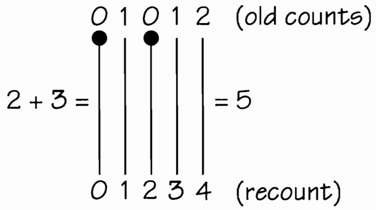


(recalling Figure 10.2(b)), but we'll represent a finite sequence by poles of equal height:

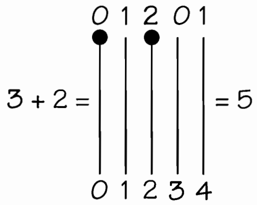


(recalling Figure 10.2(a)).

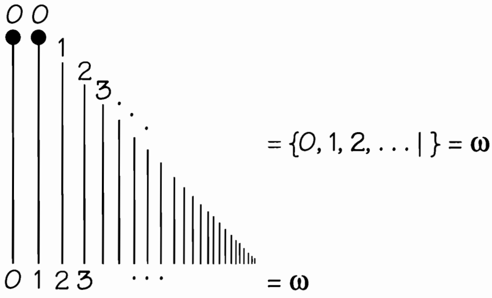
To add two of Cantor's ordinal numbers, you just put their pictures side by side and do a recount: For instance,



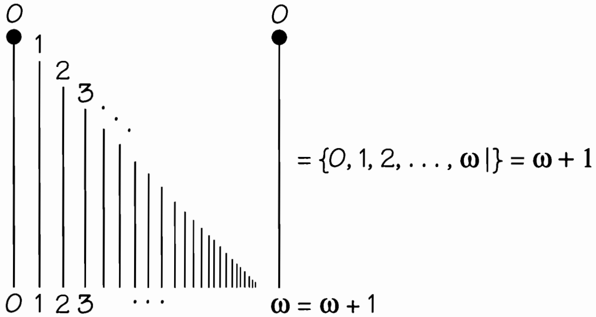
Of course, we get the same answer from 3 + 2, although the recounting's in a different order:



But infinite numbers give some surprises! We find that $1 + \omega$ is the same as ω:

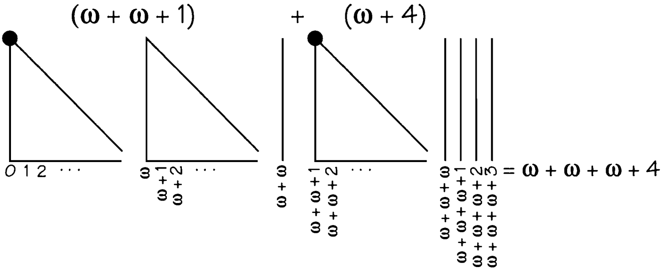


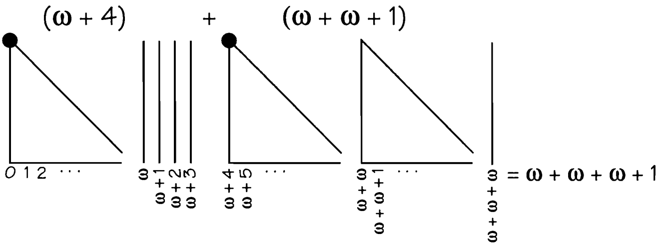
but $\omega + 1$ is bigger:



In other words, this kind of addition usually *fails* to satisfy the commutative law; $\beta + \alpha$ may be larger or smaller than $\alpha + \beta$.

As a bigger example, we'll add $\alpha = \omega+\omega+1$ to $\beta = \omega+4$, both ways around,





Since two numbers, α, β, in their two orders, can give two distinct sums, you might expect that three ordinal numbers, α, β, γ, could give six different sums,

$$\alpha+\beta+\gamma,\ \alpha+\gamma+\beta,\ \beta+\gamma+\alpha,\ \beta+\alpha+\gamma,\ \gamma+\alpha+\beta,\ \gamma+\beta+\alpha,$$

but it turns out that at least two of these six are equal, so that no three ordinal numbers can have more than five different sums.

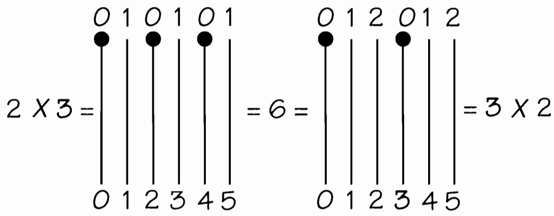
By taking the largest possible number of different sums of n ordinal numbers for $n = 1, 2, 3, \ldots$, we get the sequence

1	2	5	13	33
81	193	449	33^2	33×81
81^2	81×193	193^2	$33^2 \times 81$	33×81^2
81^3	$81^2 \times 193$	81×193^2	193^3	33×81^3
and	from here on	you multiply	the previous	row by 81:
81^4	$81^3 \times 193$	$81^2 \times 193^2$	81×193^3	$33 \times 81^4 \ldots$

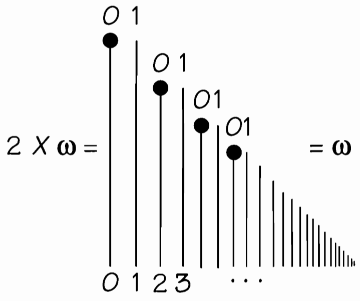
So the largest number of different sums that n ordinals can have behaves rather strangely. For 15 or more numbers, it will be either a power of 193 times a power of 81, or 33 times a power of 81.

MULTIPLYING ORDINAL NUMBERS

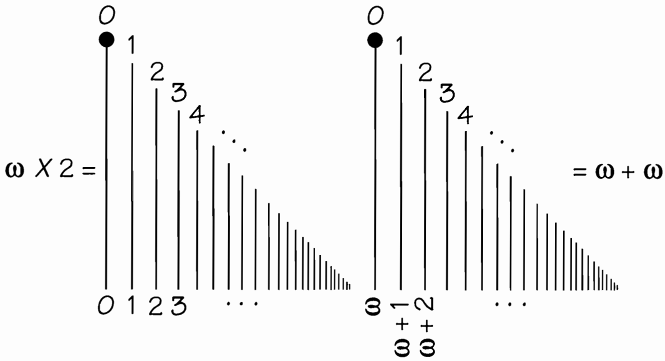
Now let's see how to multiply Cantor's numbers. The product $\alpha\times\beta$ is what you get by placing β copies of α in sequence: for instance,



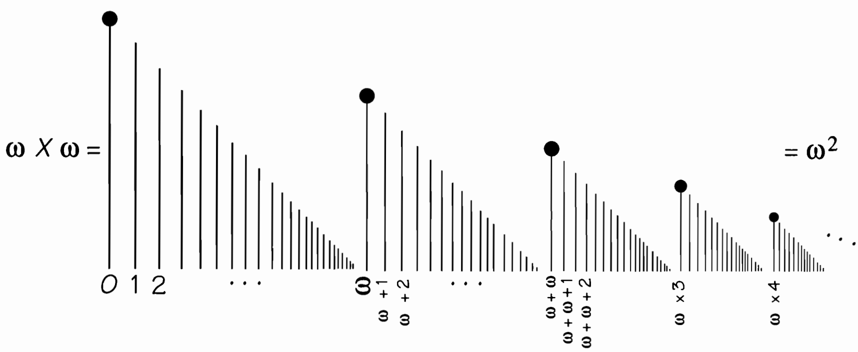
as you might expect, but infinite numbers continue to surprise us. When we take ω copies of 2, we see that $2 \times \omega$ is just ω:



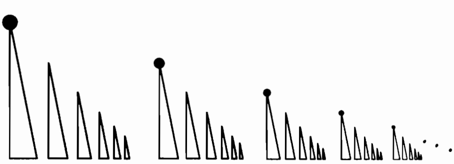
but $\omega \times 2$ (2 copies of ω) is the same as $\omega + \omega$:



What is $\omega \times \omega$ (which we can write as ω^2)? It's a much larger number than the ones we've seen before. It consists of ω copies of ω, placed in sequence:



What about ω^3, ω^4, . . . ? Well, of course, $\omega^3 = \omega^2 \times \omega$. We can get it by having ω copies of a pattern of ω^2:



Then you get ω^4 from ω copies of this; then ω^5 from ω copies of that, and so on—we won't draw the pictures for $\omega^4, \omega^5, \ldots$—and there are lots of other numbers. For instance,

$$\omega^6 \times 49 + \omega^3 \times 8 + \omega^2 \times 3 + \omega \times 57 + 1001$$

lies between ω^6 and ω^7. Figure 10.3 shows a pattern for the number $\omega^2 \times 2 + \omega \times 3 + 7$.

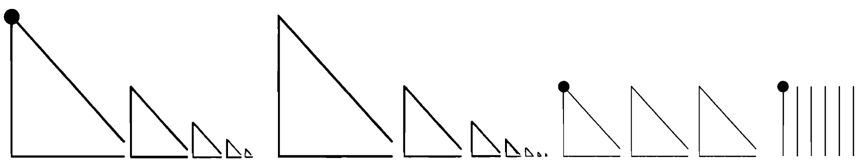


FIGURE 10.3 $(\omega^2 \times 2) + (\omega \times 3) + 7$.

Can we go further? Yes! In Cantor's system you can *always* go further! The number

$$\omega^\omega = 1 + \omega + \omega^2 + \omega^3 + \omega^4 + \cdots$$

is obtained by juxtaposing all the patterns for 1, ω, ω^2, ω^3, ω^4, . . . , in that order. Then you have

$$\omega^\omega + 1,\ \omega^\omega + 2, \ldots\ \omega^\omega + \omega, \ldots\ \omega^\omega + \omega \times 2, \ldots\ \omega^\omega + \omega \times 3, \ldots$$
$$\omega^\omega + \omega^2,\ \omega^\omega + \omega^2 + 1, \ldots\ \omega^\omega + \omega^2 + \omega, \ldots,\ \omega^\omega + \omega^3, \ldots$$
$$\omega^\omega + \omega^\omega = \omega^\omega \times 2,\ \omega^\omega \times 2 + 1, \ldots\ \omega^\omega \times 3, \ldots,\ \omega^\omega \times 4, \ldots$$
$$\omega^\omega \times \omega = \omega^{\omega+1}, \ldots\ \omega^{\omega+1} + \omega, \ldots\ \omega^{\omega+1} + \omega^2, \ldots$$
$$\omega^{\omega+1} + \omega^\omega, \ldots\ \omega^{\omega+2} \ldots,\ \omega^{\omega+3}, \ldots\ \omega^{\omega \times 2}, \ldots\ \omega^{\omega \times 3}, \ldots$$
$$\omega^{\omega^2}, \ldots\ \omega^{\omega^3}, \ldots\ \omega^{\omega^4}, \ldots\ \omega^{\omega^\omega}, \ldots\ \omega^{\omega^{\omega+1}}, \ldots\ \omega^{\omega^{\omega^\omega}}, \ldots\ \omega^{\omega^{\omega^{\omega^\omega}}}, \ldots$$

The "limit" of all these is a number that it is natural to write as

$$\omega^{\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}}$$

where there are ω omegas. This famous number was called ϵ_0 by Cantor. It's the first ordinal number that you can't get from smaller ones by a finite number of additions $\alpha+\beta$, multiplications $\alpha\times\beta$, and exponentiations α^β. Another formula for it is

$$\epsilon_0 = 1 + \omega + \omega^\omega + \omega^{\omega^\omega} + \omega^{\omega^{\omega^\omega}} + \cdots .$$

It is also the first number that satisfies Cantor's famous equation $\omega^\epsilon = \epsilon$. You'd think that this couldn't happen, because

ω^1 is much bigger than 1,
ω^2 even more so than 2,
ω^3 still more so than ω,

but Cantor showed that his equation has lots of solutions. The next is

$$\epsilon_1 = (\epsilon_0 + 1) + \omega^{\epsilon_0+1} + \omega^{\omega^{\epsilon_0+1}} + \omega^{\omega^{\omega^{\epsilon_0+1}}} + \cdots .$$

Then come

$$\epsilon_2,\ \epsilon_3,\ \ldots\quad \epsilon_\omega,\ \epsilon_{\omega+1},\ \ldots\quad \epsilon_{\omega\times 2},\ \ldots\quad \epsilon_{\omega^2},\ \ldots\quad \epsilon_{\omega^\omega},\ \ldots$$

$$\epsilon_{\epsilon_0},\ \epsilon_{\epsilon_0+1},\ \ldots\quad \epsilon_{\epsilon_0+\omega},\ \ldots\quad \epsilon_{\epsilon_0+\omega^\omega},\ \ldots\quad \epsilon_{\epsilon_0\times 2},\ \ldots\quad \epsilon_{\epsilon_1},\ \ldots$$

$$\epsilon_{\epsilon_2},\ \ldots\quad \epsilon_{\epsilon_\omega},\ \ldots\quad \epsilon_{\epsilon_{\epsilon_0}},\ \ldots\quad \epsilon_{\epsilon_{\epsilon_1}},\ \ldots\quad \epsilon_{\epsilon_{\epsilon_\omega}},\ \ldots\quad \epsilon_{\epsilon_{\epsilon_{\epsilon_0}}},\ \ldots$$

and eventually

$$\epsilon_{\epsilon_{\epsilon_{\epsilon_\ddots}}},$$

which is the first solution of $\epsilon_\alpha = \alpha$.

HOW FAR CAN WE GO?

The ordinal numbers go on for an awfully long time! No matter how big the set of them you've already got, there's always another one, and another, and another, and . . . . The precise situation was guessed by Cantor and proved a quarter of a century later by his student Zermelo in 1904: there are enough ordinals to count the members of any set of objects, no matter how big it is. Zermelo's proof showed that this depends on a hitherto unrecognized principle in mathematics: the so-called **axiom of choice**.

On the surface this axiom sounds quite innocuous. It says that if you have any collection of nonempty sets of things, you can make a new set by choosing just one from each set of the given collections. On the other hand, Zermelo's result was so astonishing that many mathematicians, from his day to ours, have had grave doubts about it.

COUNTING THE SAME SET IN DIFFERENT WAYS

There are lots of ways to count the full set of integers, positive, negative, and zero, using Cantor's ordinal numbers. You might, for instance, count them in just that way, positive, negative, and then zero:

1	2	3	4	. . .	−1	−2	−3	−4	. . .	0
0	1	2	3		ω	$\omega+1$	$\omega+2$	$\omega+3$		$\omega\times 2$

The answer this way (the first number missing) is $\omega \times 2 + 1$. It's a bit more economical to include zero with the positive numbers:

0	1	2	3	. . .	−1	−2	−3	−4
0	1	2	3		ω	$\omega+1$	$\omega+2$	$\omega+3$

That way you get the answer $\omega \times 2$. More economically still:

0	1	−1	2	−2	3	−3	
0	1	2	3	4	5	6	. . .

and we get just ω.

You can see that the answer you get depends not only on the *objects* you count, but also the *order* you count them in. The positive integers can be counted in lots and lots of different ways. The simplest is just to put them in order of size:

1 2 3 4 5 . . . ans.: ω

Or we might prefer odd numbers first:

1 3 5 7 . . . 2 4 6 8 . . . ans.: $\omega \times 2$

We might even discriminate further, classifying numbers according to exactly which power of 2 divides them. This gives

	1	3	5	7	9	
then	2	6	10	14	18	
then	4	12	20	28	36	
then	8	24	40	56	72	ans.: ω^2
		......				

Equally we could classify them by the odd factor, i.e., reading this by columns, getting the order

	1	2	4	8	16	
then	3	6	12	24	48	
then	5	10	20	40	80	
then	7	14	28	56	112	ans.: again ω^2
		......				

We can be even more profligate. Let's first have the powers of 2:

1 2 4 8 16 .... (ω, so far)

Then 3 times these, 3^2 times them, 3^3 times, etc.

3 6 12 24 48 ... 9 18 36 72 ... 27 54 108 ... 81 ... 243 ... (ω^2 more)

Then 5 times all the numbers so far, 25× them, 125× ... and so on:

5	10	20	40 ...	15	30	60 ...	45	90	180 ...
25	50	100 ...		75	150		225	450	...
125	250	500 ...		375 ...			1125	...	(ω^3 here)
625	1250 ...			...			...		
......									

Then all these times successive powers of 7, (ω^4 more)
and all those times successive powers of 11, (ω^5 more)
and so on using the primes in order.

This way we get in all

$$\omega + \omega^2 + \omega^3 + \omega^4 + \omega^5 + \cdots = \omega^\omega$$

for the final answer.

CARDINAL NUMBERS

If you find Cantor's ordinal numbers a bit wasteful, you might prefer his **cardinals**. This time, any two ways of counting the same set give the same answer. The precise definition is that two collections *A* and *B* give the same cardinal number just when there's a **one-to-one correspondence** between them. Every object in *A* must be correlated with a unique object in *B*, and vice versa.

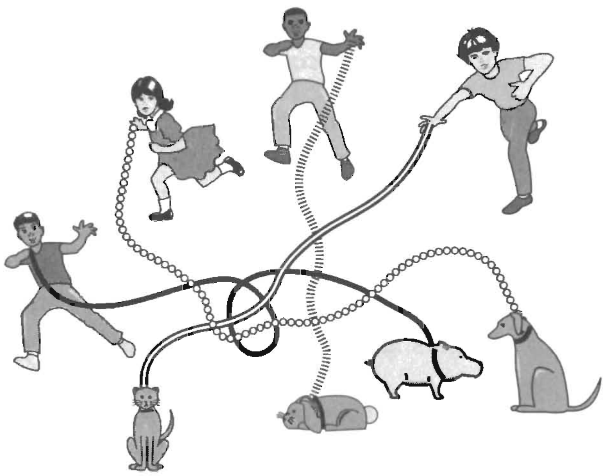


FIGURE **10.4** *Which child owns which pet?*

In Figure 10.4 each child has a unique pet and each pet has a unique owner: The one-to-one correspondence is shown by the leashes.

So we group the ordinal numbers into **classes** as follows: Because ordinal numbers can be used to count *any* set, the cardinal numbers are obtained just by collecting all the ordinal numbers you can get by counting a given set.

$$\begin{array}{llllll}\{0\} & \{1\} & \{2\} & \ldots & \{\omega \ldots \omega^2 \ldots \omega^\omega \ldots \epsilon_0 \ldots\} & \{\omega_1 \ldots\} & \{\omega_2 \ldots\} \\ 0 & 1 & 2 & \ldots & \aleph_0 & \aleph_1 & \aleph_2\end{array}$$

The names we've written under the classes are Cantor's names for the corresponding cardinal numbers.

We know that when you count a finite set you always get the same answer, so each finite ordinal number is in a class by itself, and we can afford to use the same names for the ordinal and cardinal numbers. In ordinary language you don't use *quite* the same names:

the ordinal numbers are *first*, *second*, *third*, . . . ;
the cardinal numbers are *one*, *two*, *three*, ...

Unfortunately, ordinary language doesn't really suit our purpose since we are using the labels

0, 1, 2, . . .
for the *first*, *second*, *third*, ...

The first infinite cardinal number is $\aleph_0$ (pronounced aleph zero: aleph is the first letter of the Hebrew alphabet).

A set has $\aleph_0$ members if they can be put in one-to-one correspondence with the finite ordinal numbers 0, 1, 2, ... So there are $\aleph_0$ integers. We've also seen that there are $\aleph_0$ *positive* integers. We've seen one-to-one correspondences showing that the *positive* integers, the *nonnegative* integers, and *all* the integers each have a cardinal number $\aleph_0$.

More surprisingly, there are only $\aleph_0$ positive *rational* numbers, a/b, as we see by enumerating them in the order of $a + b$ (omitting those not in their lowest terms):

$$\frac{1}{1}\ \frac{1}{2}\ \frac{2}{1}\ \frac{1}{3}\ \frac{3}{1}\ \frac{1}{4}\ \frac{2}{3}\ \frac{3}{2}\ \frac{4}{1}\ \frac{1}{5}\ \frac{5}{1}\ \frac{1}{6}\ \frac{2}{5}\ \frac{3}{4}\ \frac{4}{3}\ \frac{5}{2}\ \frac{6}{1}\ \frac{1}{7}\ \frac{3}{5}\ \frac{5}{3}\ \frac{7}{1}\ \frac{1}{8}\ \frac{2}{7}\ \frac{4}{5} \cdots$$
$$0\ 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9 \ldots$$

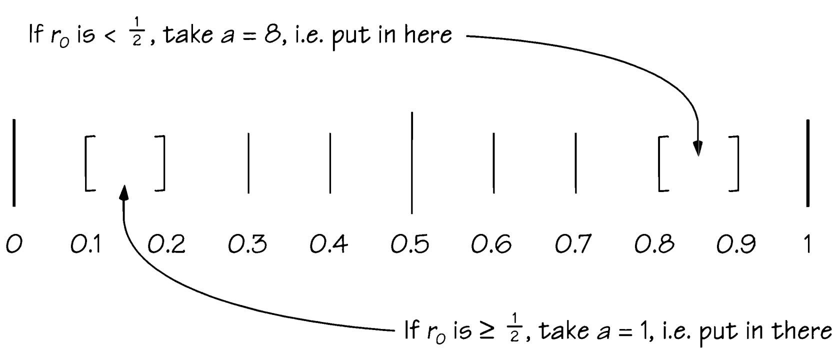
The number $\aleph_0$ wouldn't increase if we included 0 and the negative rational numbers as well. Indeed, Cantor showed that there are still precisely $\aleph_0$ *algebraic* numbers!

A set is often called **countable** if it is either finite or has exactly $\aleph_0$ members. Are there any *uncountable* sets? Collections that contain *more* than $\aleph_0$ objects? Yes! Cantor made the surprising discovery that there are *strictly* more than $\aleph_0$ real numbers! Although people had speculated about the infinite for thousands of years, this was the first time that anyone realized that there was more than one infinite number!

However, no countable sequence $r_0, r_1, r_2, \ldots$ can contain every real number! We prove this by producing a real number r that is not equal to any of the $r_0, r_1, r_2, \ldots$ in the sequence. The number r will lie between 0 and 1, and we find it by specifying the successive digits in its decimal expansion

$$r = 0 \cdot abcd \ldots.$$

Whatever number r_0 is, we can choose a so as to guarantee that r will differ from r_0 by at least 3/10.



In a similar way, we can choose b so that r differs from r_1 by at least 3/100, and c so that r differs from r_2 by at least 3/1000, and so on indefinitely, and the completed number $r \cdot abcd \ldots$ cannot be any of the numbers $r_0, r_1, r_2, \ldots.$

So the number of real numbers must be strictly more than $\aleph_0$! Exactly how many are there? We'll only be able to tell you when we've shown how to do arithmetic with cardinal numbers.

COUNTING CARDS

If your hand contains 5 red cards and 8 black ones, then it contains $5 + 8 = 13$ cards in all. Seizing on this, Cantor defined the sum of any two cardinal numbers by saying that if a set can be split into two sets having cardinals M and N, then its cardinal is $M+N$. For instance, {0, 1, 2, . . .} can be split in this way, one set containing only 0 and the other containing the positive numbers. So

$$\aleph_0 = 1 + \aleph_0.$$

Equally, by splitting it into odds and evens, we find

$$\aleph_0 = \aleph_0 + \aleph_0.$$

Each card in a normal deck can be specified by giving its rank and its suit. Since there are 13 ranks A, 2, . . . , J, Q, K and 4 suits ♠, ♥, ♦, ♣, there are $13 \times 4 = 52$ cards in all. Cantor generalized this by saying that *any* set whose objects can be uniquely specified in terms of "ranks" and "suits" has cardinal number $R \times S$, where R is the number of ranks and S is the number of suits. For instance, any number from {0, 1, 2, . . .} is uniquely specified by giving its quotient and remainder on division by 5. As there are $\aleph_0$ quotients and 5 possible remainders, we have

$$\aleph_0 = \aleph_0 \times 5.$$

Indeed, we can see that

$$\aleph_0 = \aleph_0 \times \aleph_0$$

by observing that any positive number is a unique odd number ($\aleph_0$ possibilities) multiplied by a unique power of two (another $\aleph_0$ possibilities).

The cloakroom tickets in a book are specified by 3 digits running from 000 to 999. Since each digit has 10 possibilities, there are exactly $10^3 = 1000$ tickets in the book. Cantor generalized this by saying that any set of "tickets" that can be uniquely specified by giving X digits from a population of size Y will have cardinal number

$$Y^X$$

How many real numbers are there between 0 and 1? This looks easy! Let us suppose there are precisely C real numbers. They are specified by their decimal expansions

$$.abcd \ldots$$

in which there are $\aleph_0$ digits each chosen from a set of 10 possibilities, so the answer should be

$$10^{\aleph_0}.$$

However, closer examination reveals that *some* real numbers have *two* decimal expansions, for example, 0.500000 . . . = 0.499999 . . . . In fact, there are exactly $\aleph_0$ such exceptional numbers, so what we've really shown is that

$$10^{\aleph_0} = \aleph_0 + C.$$

However, we can prove that C is unchanged by adding $\aleph_0$. This is because each of our real numbers is either rational ($\aleph_0$ possibilities) or irrational, so that $C = \aleph_0 + X$, where X is the number of irrational numbers. Now since $\aleph_0 + \aleph_0 = \aleph_0$, we deduce that

$$\aleph_0 + C = \aleph_0 + (\aleph_0 + X) = (\aleph_0 + \aleph_0) + X = \aleph_0 + X = C.$$

By using base 2 instead of 10 we could show that C is also equal to

$$2^{\aleph_0},$$

which is its standard name.

In fact, there are one-to-one correspondences which show that

$$\left.\begin{array}{l} \aleph_0 = n + \aleph_0 = \aleph_0 + \aleph_0 \\ C = n + C = \aleph_0 + C = C + C \end{array}\right\} n = 0, 1, 2, \ldots$$

$$\left.\begin{array}{l} \aleph_0 = n \times \aleph_0 = \aleph_0 \times \aleph_0 \\ C = n \times C = \aleph_0 \times C = C \times C \end{array}\right\} n = 1, 2, 3, \ldots$$

$$\left.\begin{array}{l} \aleph_0 = \aleph_0^n \\ C = C^n = n^{\aleph_0} = \aleph_0^{\aleph_0} = C^{\aleph_0} \end{array}\right\} n = 2, 3, 4, \ldots$$

So far we've only counted the real numbers x, $0 \leq x < 1$, but since an arbitrary real number is uniquely an integer plus one of these numbers x, the total number of real numbers is

$$\aleph_0 \times C = C.$$

These are precisely $C = 2^{\aleph_0}$ real numbers
and also just C complex numbers, since $C \times C = C$.

Since this is vastly more than the number, $\aleph_0$, of algebraic numbers.

Almost all real and complex numbers are transcendental!

Cantor now faced the problem: Every cardinal number was supposed to appear in his sequence

$$0, 1, 2, \ldots, \aleph_0, \aleph_1, \aleph_2, \ldots, \aleph_\omega, \ldots$$

(there being one infinite cardinal number $\aleph_\alpha$ for each ordinal number α). Which of these is C?

We cannot attempt to answer the question until we've told you what the numbers $\aleph_\alpha$ are! Here are the definitions:

$\aleph_0$ is the number of finite ordinal numbers.

$\aleph_1$ is the number of ordinal numbers that are either finite or in the $\aleph_0$ class.

$\aleph_2$ is the number of ordinal numbers that are either finite or in the $\aleph_0$ or $\aleph_1$ classes.

$\aleph_\omega$ is the number of ordinal numbers that are either finite or in one of the classes $\aleph_0, \aleph_1, \aleph_2, \ldots$, and so on. . . . .

In fact, we cannot tell you the answer now that we've told you what the numbers are!

Is $C = \aleph_1$?

Cantor guessed that it was, and this is called the **continuum hypothesis**. In 1940 the Austrian mathematician Kurt Gödel showed that Cantor's guess can never be *disproved* from the other axioms of mathematics! Unfortunately, in 1963, the American mathematician Paul Cohen showed that it couldn't be *proved* either!

So we need a new axiom! The prevailing opinion today is that the continuum hypothesis should be considered false.

Similar remarks hold for the **generalized continuum hypothesis**

Is $2^{\aleph_\alpha} = \aleph_{\alpha+1}$ for every α?

SURREAL NUMBERS

Just as the *real* numbers fill in the gaps between the integers, the ***surreal*** numbers fill in the gaps between Cantor's ordinal numbers. We get them by generalizing our use of the $\{\,|\,\}$ notation for the ordinal numbers.

The symbol

$$\{a, b, c, \ldots | d, e, f, \ldots\}$$

means "the ***simplest*** number strictly greater than all the numbers a, b, c, . . . and strictly less than all the numbers d, e, f, ..." Just what we mean by "simplest" will emerge after we've done some examples.

The ordinal numbers are those where there aren't any numbers to the right of the bar:

$\{\,|\,\} = 0$, the simplest number of all
$\{0|\,\} = 1$, the simplest number greater than 0
$\{0,1|\,\} = 2$, the simplest number greater than 1 (and 0)

and so on.

But now we can put numbers on either side of the bar. Thus

$\{0|1\}$ is the simplest number between 0 and 1, namely $\frac{1}{2}$,

and we may have no numbers on the *left* of the bar:

$\{\,|0\}$ is the simplest number less than 0, namely -1.

The same surreal number may have many definitions: $\{1|\,\}$ is another name for $2 = \{0,1|\,\}$, because the simplest number greater than 1 is also greater than 0. Some other names for 2 are $\{1|3\}$, $\{1\frac{1}{2}|4\}$ and $\{1|\omega\}$.

How do we tell when two different names give the same number? The answer is rather surprising: we play a game!

SURREAL NUMBERS ARE GAMES!

Any number

$$\{a, b, c, \ldots | d, e, f, \ldots\} = g$$

may also be regarded as a game played between two players, Left and Right: the **moves**

from g to a, b, c, . . . are legal only for Left, and those
from g to d, e, f, . . . are legal only for Right.

Suppose that Left moves from g to b, say. Then b will have a similar definition:

$$b = \{A, B, C, \ldots \mid D, E, F, \ldots\},$$

and so Right may now move from b to any of D, E, F, ... Suppose he moves to

$$D = \{\alpha, \beta, \gamma, \ldots \mid \delta, \epsilon, \zeta, \ldots\}.$$

Then Left may move to any of α, β, γ, . . . , and so on. The first person unable to move loses, and the other is then the winner.

THE GAME OF HACKENBUSH

The game of Hackenbush is played with a picture made of nodes joined by blue and red edges. In our pictures the blue edges are printed black. The picture must have the property that from any node you can reach the ground (the dotted line) by travelling along a chain of edges (see Figure 10.5). The players, Left and Right, move alternately; at his move Left may delete any bLack edge and Right may delete any Red edge. At each move, any edges no longer connected to the ground must be deleted. If you can't move when it's your turn, you lose.

We use the same notation

$$\{a, b, c, \ldots \mid d, e, f, \ldots\} = g$$

for Hackenbush positions as we do for numbers since it will turn out that they're much the same thing. Left can move from g to a or b or c or . . . , while Right can move from g to d or e or f or . . . .

Let's analyze some easy positions. At the end of the game you

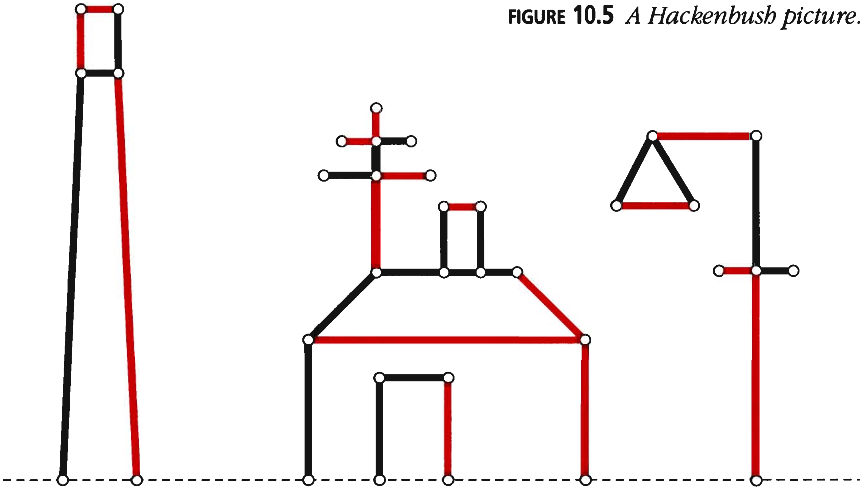


FIGURE 10.5 *A Hackenbush picture.*

may be faced with the empty position in which there are no red or black edges at all (Figure 10.6) neither player has any legal moves.

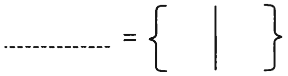
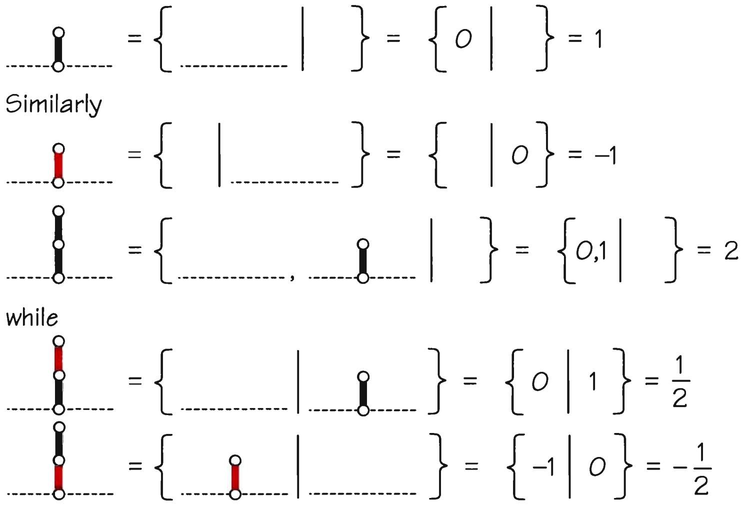


FIGURE 10.6 *The empty position.*

It's natural to call this 0.

From the Hackenbush position containing just 1 black edge (and no red ones), Left can move to 0, and Right can't move at all:



In general, you get the **negative** of a Hackenbush game just by interchanging red and black edges throughout. This reverses the roles of the two players. So if

$$g = \{a, b, c, \ldots | d, e, f, \ldots\},$$

then

$$-g = \{-d, -e, -f, \ldots | -a, -b, -c, \ldots\}.$$

Thus

$$1\tfrac{1}{2} = \{1|2\} \text{ gives } -1\tfrac{1}{2} = \{-2|-1\}.$$

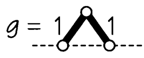
What do these equalities mean?

It turns out that every Hackenbush position has a **value** which is a surreal number, and that every surreal number is its value of a Hackenbush position, indeed of a Hackenbush *chain*.

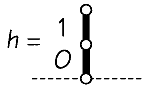
So from now on we'll think of numbers in terms of the game of Hackenbush. This gives us a way of telling which of two surreal numbers is the larger, and in particular whether two surreal numbers are equal. To do this you find Hackenbush positions, g and h, with corresponding values and then play the game found by combining g and $-h$ into a single picture and asking: who wins?

If Left can win, no matter who starts, we say that $g > h$.
If Right can win, no matter who starts, we say that $g < h$.
If the *second* player to move can always win, we say that $g = h$.

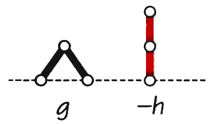
Here's a game



whose value is $\{1|\ \}$, and here's one

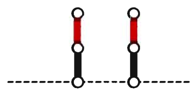


whose value is $\{0,1|\ \}$. We'll use our method to verify that $\{1|\ \}$ and $\{0,1|\ \}$ are indeed two different names for the same number. The two moves for Left in g both leave the same value, 1. On the other hand, as we've already seen, h has value $\{0, 1|\ \}$. The compound position, $g - h$, is

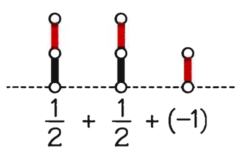


You can see that each player can guarantee himself two moves, and so *whoever is second to move can win*: $g = h$.

To **add** two numbers, we juxtapose two Hackenbush games. For instance, $\frac{1}{2} + \frac{1}{2}$ is represented by



You can now verify that $\frac{1}{2} + \frac{1}{2} = 1$ by checking that the second player to move can win from the position



The standard Hackenbush game for any number is just a chain of edges proceeding upwards from the ground. We can now be a little more clear about what we mean by "simplest": the simpler of two numbers is the one whose corresponding Hackenbush chain is shorter. Figure 10.7 shows some of the simpler Hackenbush chains.

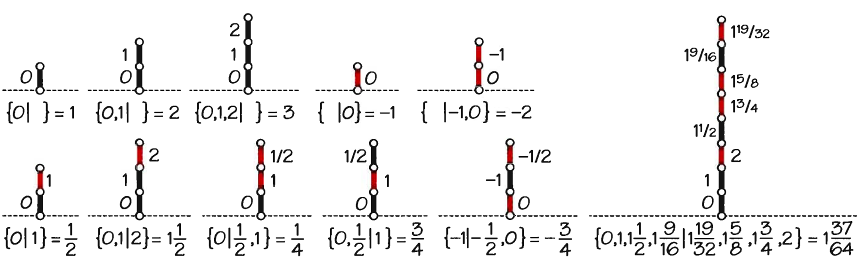
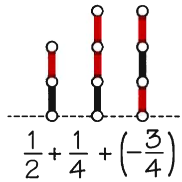


FIGURE 10.7 *Values of some Hackenbush chains. A small number written to the left or right of an edge is the value that results when the appropriate player deletes that edge.*

You might like to check that $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$ by playing



Elwyn Berlekamp has a simple rule for the correspondence between positive real numbers and Hackenbush chains (Figure 10.8). The first pair of edges of opposite color is treated as a binary point, and the blue and red edges above this pair are read as digits 1 and 0 after the binary point and an extra 1 is appended if the chain is finite. The integer part is just the number of blue edges below this pair.

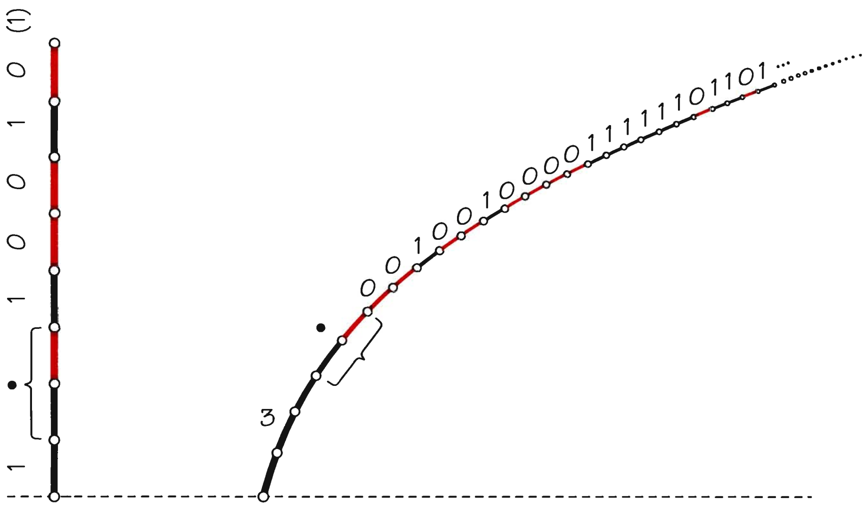


FIGURE 10.8 *Berlekamp's Rule. The last picture in Figure 10.7 has (binary) value* 1.1001101 = $1 + \frac{1}{2} + \frac{1}{16} + \frac{6}{14} = 1\frac{37}{64}$. *The value of* π *in binary is* (3).001001000011111101101 . . . .

Hackenbush chains can be infinite! Indeed, we allow the height of our Hackenbush chains to be *any* of Cantor's ordinal numbers,

$$0, 1, 2, \ldots, \omega, \omega + 1, \omega + \omega, \ldots, \omega^2, \ldots.$$

Figure 10.9 shows some examples.
Even though the players sometimes have an infinite number of options, you'll find that these games can't go on forever.

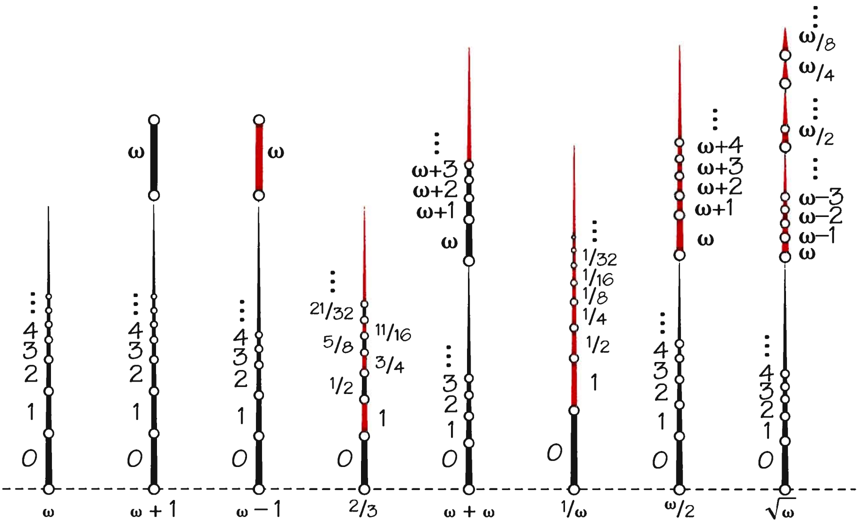


FIGURE 10.9 *Surreal values of some infinite Hackenbush chains.*

Is the game we've labelled $\omega + 1$ really the sum of ω and 1? Yes! If the second player replies to the opening moves with the moves indicated by the arrows in Figure 10.10, he or she can win.

Likewise, the game in Figure 10.11 shows that $\frac{1}{2}\omega + \frac{1}{2}\omega = \omega$: The (bad) moves that Right might make in $-\omega$ are countered by any Left move.

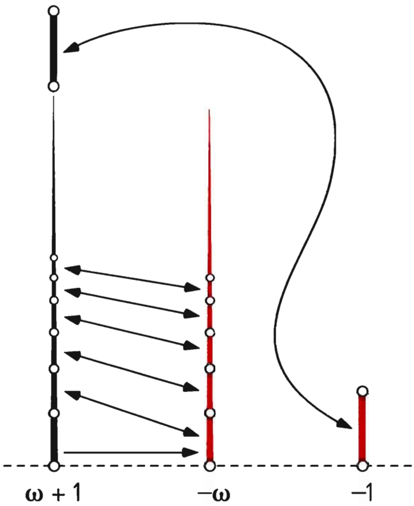


FIGURE 10.10 *The sum of* ω *and* 1 *is* $\omega + 1$.

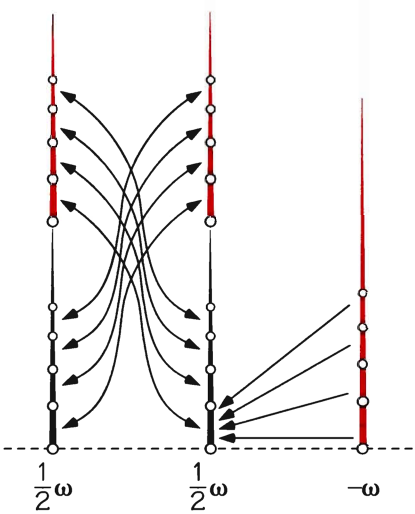


FIGURE 10.11 *The sum of* $\frac{1}{2}\omega$ *and* $\frac{1}{2}\omega$ *is* ω.

The formal definitions of addition, negation, and multiplication are given in Figure 10.12. Using them you could check that $1/\omega$ deserves its name by multiplying it by ω to get 1. You could also check, by squaring

$$\sqrt{\omega} = \{0,\ 1,\ 2,\ \ldots\ |\omega,\ \frac{1}{2}\ \omega\ \frac{1}{4}\ \omega,\ \frac{1}{8}\ \omega,\ \ldots\}$$

that it, too, deserves its name. This number is the value of a Hackenbush chain consisting of one ω-sized blue chain surmounted by ω ω-sized red chains.

$$\text{If } \alpha = \{\ldots, a, \ldots \mid \ldots, A, \ldots\} \text{ and } \beta = \{\ldots, b, \ldots \mid \ldots, B, \ldots\}, \text{ then}$$
$$\alpha + \beta = \{\ldots, a + \beta, \ldots, \ldots, \alpha + b, \ldots \mid \ldots, A + \beta, \ldots, \ldots, \alpha + B, \ldots\}$$
$$\text{and } -\alpha = \{\ldots, -A, \ldots \mid \ldots, -a, \ldots\},$$
$$\text{while } \alpha\beta = \left\{ \begin{matrix} \ldots, a\beta + ab - ab, \ldots \\ \ldots, A\beta + \alpha B - AB, \ldots \end{matrix} \middle| \begin{matrix} \ldots, a\beta + \alpha B - aB, \ldots \\ \ldots, A\beta + \alpha b - Ab, \ldots \end{matrix} \right\}$$

FIGURE 10.12 *We suppose that α has been named in terms of various numbers $a < \alpha$ and $A > \alpha$, and similarly for β. For example, one of the numbers on the right in the definition of $\alpha\beta$ might be $\alpha_{13}\beta + \alpha\beta_7 - \alpha_{13}\beta_7$. For example,*

$$8 \times 25 = \{0, 1, \ldots, 7 \mid\} \times \{0, 1, \ldots, 24\}$$
$$= \{7 \mid\} \times \{24 \mid\} = \{7 \times 25 + 8 \times 24 - 7 \times 24 \mid\} = \{199 \mid\} = 200.$$

The definitions are inductive, it being supposed that the simpler products have already been computed.

NIMBERS AND THE GAME OF NIM

We digress to consider yet another type of number, which has both finite and infinite manifestations. You can play **Neutral Hackenbush** with pictures that have thin red edges, which may be deleted by either Left or Right. If your picture consists only of chains (Figure 10.13), then you're really playing C.L. Bouton's game of Nim, usually played with heaps of beans, the move being to take any number of beans from a heap. The player who takes the last heap is the winner.

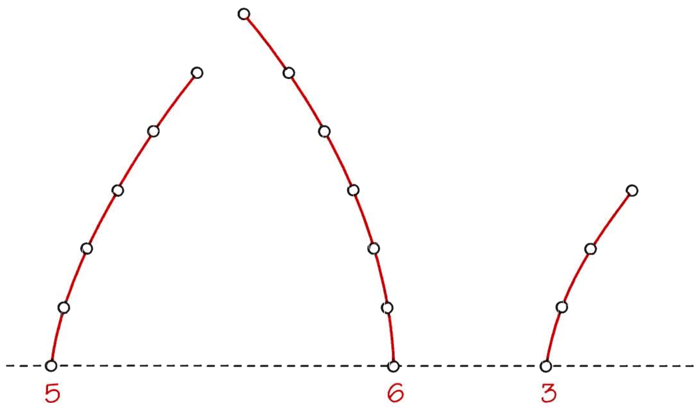


FIGURE 10.13 *Nim played as a Neutral Hackenbush game.*

We'll stick with the Hackenbush version and will write n for the value of a Neutral Hackenbush stalk with n edges. These "values" are really numbers of a new kind: we'll call them **nimbers**. They have a new kind of arithmetic for which we'll use $+$, $\times$ and $=$. The Nim arithmetic tables for nimbers below 16 appear in Tables 10.1, 10.2, and 10.3.

When we define powers $a^1 = a$, $a^2 = a \times a$, $a^3 = a \times a \times a$, . . . , there are many surprises. We find

$$2^2 = 3,\ 4^4 = 5,\ 16^{16} = 17,\ 256^{256} = 257,\ \ldots .$$

And there are more nimbers than you might think: one for each of Cantor's ordinal numbers. Infinite nimber arithmetic yields more surprises:

$$\omega^3 = 2!$$

How do you tell which are the good positions to move in Nim? If you're playing Nim with just one heap, the answer would be easy. It's good to move to 0 (since this ends the game) and bad to move to any other nimber,

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ \ldots$$

(since your opponent could then move to 0).

$a+b$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14
2	2	3	0	1	6	7	4	5	10	11	8	9	14	15	12	13
3	3	2	1	0	7	6	5	4	11	10	9	8	15	14	13	12
4	4	5	6	7	0	1	2	3	12	13	14	15	8	9	10	11
5	5	4	7	6	1	0	3	2	13	12	15	14	9	8	11	10
6	6	7	4	5	2	3	0	1	14	15	12	13	10	11	8	9
7	7	6	5	4	3	2	1	0	15	14	13	12	11	10	9	8
8	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7
9	9	8	11	10	13	12	15	14	1	0	3	2	5	4	7	6
10	10	11	8	9	14	15	12	13	2	3	0	1	6	7	4	5
11	11	10	9	8	15	14	13	12	3	2	1	0	7	6	5	4
12	12	13	14	15	8	9	10	11	4	5	6	7	0	1	2	3
13	13	12	15	14	9	8	11	10	5	4	7	6	1	0	3	2
14	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1
15	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0

TABLE 10.1 *A Nim addition table. All the entries are nimbers.*

Here's all you need to know about addition of nimbers:

> Two equal nimbers always add to **0**.
> If the 'larger' of two different nimbers
> is **1** or **2** or **4** or **8** or **16** or . . . ,
> you add them just as you add
> the corresponding ordinary numbers.

a + b	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	0	2	3	1	8	10	11	9	12	14	15	13	4	6	7	5
3	0	3	1	2	12	15	13	14	4	7	5	6	8	11	9	10
4	0	4	8	12	6	2	14	10	11	15	3	7	13	9	5	1
5	0	5	10	15	2	7	8	13	3	6	9	12	1	4	11	14
6	0	6	11	13	14	8	5	3	7	1	12	10	9	15	2	4
7	0	7	9	14	10	13	3	4	15	8	6	1	5	2	12	11
8	0	8	12	4	11	3	7	15	13	5	1	9	6	14	10	2
9	0	9	14	7	15	6	1	8	5	12	11	2	10	3	4	13
10	0	10	15	5	3	9	12	6	1	11	14	4	2	8	13	7
11	0	11	13	6	7	12	10	1	9	2	4	15	14	5	3	8
12	0	12	4	8	13	1	9	5	6	10	2	14	11	7	15	3
13	0	13	6	11	9	4	15	2	14	3	8	5	7	10	1	12
14	0	14	7	9	5	11	2	12	10	4	13	3	15	1	8	6
15	0	15	5	10	1	14	4	11	2	13	7	8	3	12	6	9

TABLE 10.2 *A Nim multiplication table. All the entries are nimbers.*

And here's all you need to know about multiplication of nimbers:

> If the 'larger' of two different nimbers is
> 1 or 2 or 4 or 16 or 256 or 65536 or 429497296 or . . . ,
> you multiply them just as you multiply
> the corresponding ordinary numbers.
> The product of one of these *special* nimbers
> with itself is obtained by taking $1\frac{1}{2}$ times
> its ordinary value.

a^b	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3	1
3	1	3	2	1	3	2	1	3	2	1	3	2	1	3	2	1
4	1	4	6	14	5	2	8	11	7	10	3	12	13	9	15	1
5	1	5	7	13	4	2	10	9	6	8	3	15	14	11	12	1
6	1	6	5	8	7	3	13	15	4	14	2	11	10	12	9	1
7	1	7	4	10	6	3	14	12	5	13	2	9	8	15	11	1
8	1	8	13	14	10	1	8	13	14	10	1	8	13	14	10	1
9	1	9	12	10	11	2	14	4	15	13	3	7	8	5	6	1
10	1	10	14	13	8	1	10	14	13	8	1	10	14	13	8	1
11	1	11	15	8	9	2	13	5	12	14	3	6	10	4	7	1
12	1	12	11	14	15	3	8	6	9	10	2	4	13	7	5	1
13	1	13	10	8	14	1	13	10	8	14	1	13	10	8	14	1
14	1	14	8	10	13	1	14	8	10	13	1	14	8	10	13	1
15	1	15	9	13	12	3	10	7	11	8	2	5	14	6	4	1

TABLE 10.3 *Nim powers* a^b. *Here* a *is a nimber, but* b *is an ordinary number.*

Using these rules and the laws of algebra, you can do nim arithmetic:

$$5 + 6 = (4 + 1) + (4 + 2) = (4 + 4) + (1 + 2) = 0 + 3 = 3$$

$$\begin{aligned} 5 \times 6 &= (4 + 1) \times (4 + 2) \\ &= (4 \times 4) + (4 \times 2) + (1 \times 4) + (1 \times 2) \\ &= 6 + 8 + 4 + 2 = 8 \end{aligned}$$

It turns out that any collection of Nim heaps, a, b, c, . . . , can be replaced by a single heap of size $a + b + c + \cdots$, so the answer to our question is: You should try to move to a position a, b, c . . . for which $a + b + c + \cdots = 0$

So, if your opponent moved to a position

$$3, 5, 7$$

he should lose, since

$$(3 + 5) + 7 = 6 + 7 = 1$$

from Table 10.1. If you respond by moving to *any* of

$$2, 5, 7 \quad 3, 4, 7 \quad 3, 5, 6$$

and continue playing this well, you'll eventually win, since

$$2 + 5 + 7 = 3 + 4 + 7 = 3 + 5 + 6 = 0$$

Nim addition has applications to many other games, and Nim multiplication helps to analyze a few more, and these notions have some other applications, outside game theory. Here's a simple one.

Each sequence in Table 10.4 has been selected as the first sequence

$$. . \ 000 \ . . . \ v \ w \ x \ y \ z$$

that differs from each of its predecessors in at least three places. The individual digits may be arbitrary nimbers

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, \ldots$$

(where 11, for example, is to be thought of as a single digit), but the order is as if they were ordinary integers, written in a sufficiently large base. This is called the **integral lexicographic code** of minimal **distance** 3. Codes similar to this have important practical applications in the digital transmission of information (see Table 10.4).

It's a strange fact that if you add two codewords, digit by digit, without carrying, you'll always get a new codeword. Even stranger, if you multiply each separate digit of a codeword by a nimber, you get a codeword.

...000000
...000111
...000222
...000333
...0 ... ...
...000nnn
...0 ... ...
...001012
...001103
...001230
...001321
...001456
...0 ... ...
...002023
...002132
...0 ... ...
...003031
...0 ... ...
...004048
...0 ... ...
...0 ... ...
...010123
...0 ... ...
...100132
...0 ... ...

TABLE 10.4 *The codewords of the integral lexicographic code of distance 3.*

The values of other Neutral Hackenbush pictures are also nimbers. For instance, the little tree on the left of Figure 10.14 is equivalent to the stalk on the right, since the branches **2** and **3** at *B* can be replaced by **2 + 3 = 1**

Infinite Green Hackenbush involves the arithmetic of infinite nimbers, which we won't teach you in full. The good move in Figure 10.15 is to replace $\omega \times 2$ by $\omega + 6$ since

$$(\omega + 6) + (\omega + 3) + 5 = 0$$

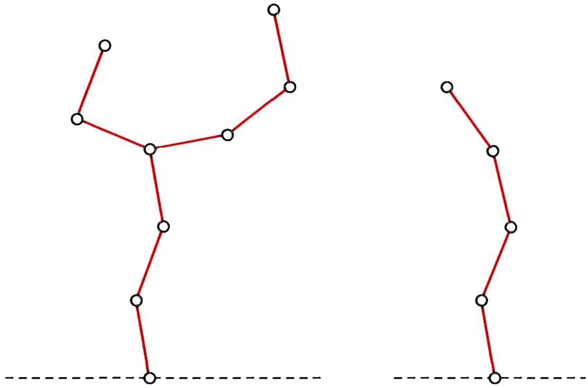


FIGURE 10.14 *Value of a Neutral Hackenbush tree.*

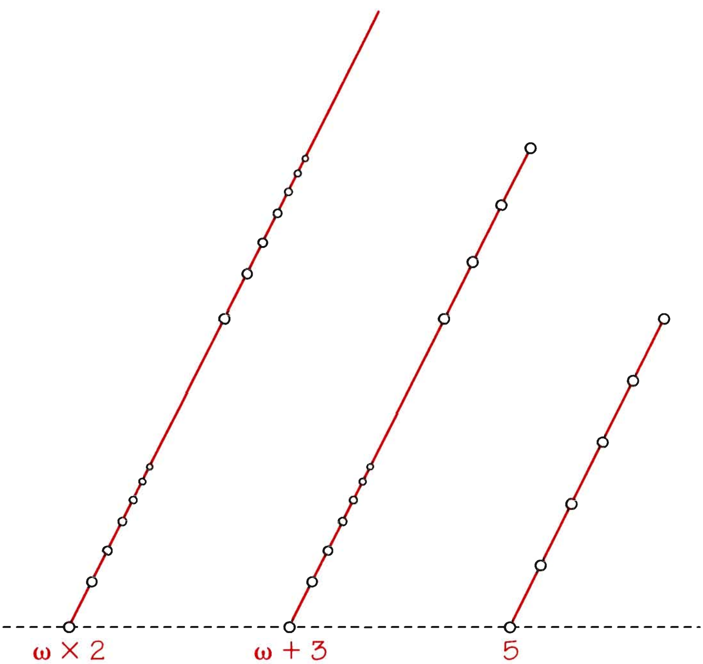


FIGURE 10.15 *Winning an infinite Neutral Hackenbush game.*

ORDERS OF INFINITY

Have the surreal numbers any other uses? Well, one is to measure the rate of growth of a function $f(X)$, as X tends to infinity. We say that $X, X^2, X^3, \ldots$ grow at the rates $1, 2, 3, \ldots$, but the function e^x grows more rapidly than any of these. Let's define it to have growth rate ω. The table shows the growth rates of some functions.

X	X^2	X^3	$\sqrt{X}$	e^X	$e^X \times X$	e^X/X	e^{2X}	$e^{½X}$	$e^{\sqrt{X}}$	e^{X^2}	e^{e^X}	$\ln X$
1	2	3	½	ω	$\omega + 1$	$\omega - 1$	2ω	ω^2	$\sqrt{\omega}$	ω^2	ω^ω	$1/\omega$

The idea of rates of growth as numbers first appeared in the infinitary calculus of Paul Dubois-Raymond.

In calculus problems one often deals with a power series involving a small number x, for instance,

$$\ln(1+x) = x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \frac{1}{4}x^4 + \cdots .$$

Here we regard this as the sum of the first-order infinitesimal x, the second-order one $-\frac{1}{2}x^2$, the third-order one $\frac{1}{3}x^3$, and so on. We say that $x - \frac{1}{2}x^2 + \frac{1}{3}x^3$ is a third-order approximation to $\ln(1 + x)$. In our notation, taking $x = 1/X$, these are *decreasing* functions of X and so have *negative* growth rates:

x	x^2	x^3
-1	-2	-3

We expect that the future will bring other applications and other species of infinite numbers, but we shall stop here, since we cannot write infinitely many Chapters!

REFERENCES

E.R. Berlekamp, J. H. Conway & R. K. Guy, *Winning Ways for your Mathematical Plays,* Academic Press, New York, 1982.

G. Cantor. *Über unendliche, lineare Punktmannigfaltigkeiten, Arbeiten zur Mengenlehre aus dem Jahren 1872–1884,* Teubner-Archiv zur Mathematik, Leipzig, 1984.

J. H. Conway, *On Numbers and Games,* Academic Press, New York, 1976.

Donald Knuth. *Surreal Numbers,* Addison-Wesley, Reading MA, 1974.

Index