Simon Gindikin

Tales of MATHEMATICIANS *and* PHYSICISTS

Springer

Tales of Mathematicians and Physicists

Simon Gindikin

Tales of Mathematicians and Physicists

Translated from the Russian by Alan Shuchat

Springer

Simon Gindikin
Department of Mathematics
Rutgers University
Piscataway, NJ 08854
USA

Alan Shuchat (*Translator*)
Department of Mathematics
Wellesley College
Wellesley, MA 02181
USA

Cover design by Alex Gerasev.

Mathematics Subject Classification (2000): 01AXX, 01A70

Library of Congress Control Number: 2006935553

ISBN-10 0-387-36026-3 e-ISBN-10: 0-387-048811-1
ISBN-13 978-387-36026-3 e-ISBN-13: 978-0-387-48811-0

Printed on acid-free paper.

© 2007 Second English edition Springer Science+ Business Media, LLC
Translation by Alan Shuchat based on the third Russian edition: *Rasskazy o Fizikakh i Matematikakh.* Izdatel'stvo Moskovskogo. Zentra Nepreryvnogo Matematicheskogo Obrazovania, Moscow, Russia, 2001.

First edition of *Tales of Physicists and Mathematicians*, Birkhäuser Boston, ©1988.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer Science+Business Media LLC, 233 Spring Street, New York, NY 10013, USA) and the author, except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.
The use in this publication of trade names, trademarks, service marks and similar terms, even if they are not identified as such, is not to be taken as an expression of opinion as to whether or not they are subject to proprietary rights.

9 8 7 6 5 4 3 2 1

springer.com (JLS/SB)

Contents

Preface to the English Edition vii

Preface to the Third Russian Edition xiii

Preface to the First Russian Edition xvii

Ars Magna (The Great Art) 1

Two Tales of Galileo 27

Christiaan Huygens and Pendulum Clocks 79

Secrets of the Cycloid 93

Blaise Pascal 129

The Beginnings of Higher Geometry 151

Leonhard Euler 171

Joseph Louis Lagrange 213

Pierre-Simon Laplace 247

Prince of Mathematicians 263

Felix Klein 311

The Magic World of Henri Poincaré 323

The Enigma of Ramanujan 337

On the Advantages of Coordinates and the Art of Chaining Hyperboloids 349

The Complex World of Roger Penrose 369

Preface to the English Edition

Tales of Physicists and Mathematicians is a translation of a book that was published in Russia in 2001 and is based on articles that appeared from 1960–1980. The first edition of the book, less than half the size of the current one, was published in the Soviet Union in 1981 and in English in 1988. Thus the book has its own history, and I would like to share some of the circumstances under which it appeared to the western reader.

This was a time not only of a surprising flourishing of mathematics in the Soviet Union but also of its surprisingly great prestige in society, perhaps not seen since the time of Plato's Academy in Athens. Mathematics attracted talented youth not only as an area where they could stretch themselves intellectually but also as one that minimized the influence of the official Marxist ideology that deeply penetrated into the lives of the "Soviet people." The profession of scientist, and in particular of mathematician, carried great authority. Here is an interesting observation in this regard. Children of the top Communist elite, including some "members of the Politburo," sometimes chose mathematics or another science as their professions, just as future kings often studied with Plato. Mathematics was lucky: it was never a personal "concern" of Stalin, as were biology, linguistics, and economics, which inevitably led to annihilating, punitive operations against them. In a fantasy of Solzhenitsyn, Stalin looked through a high school mathematics text, choosing the next science to be the subject of his concern. It is hard to imagine what would happen next. The opinion "upstairs" that a high level in the exact sciences was important for the military industry no doubt helped. Gradually, it became the fashion

to have mathematicians in any serious organization. Often they enjoyed some freedom, but this is reminiscent of the freedom of the court jester. The comparative idyll between mathematicians and those in power ended in the late 1960s when many mathematicians signed a letter to the government defending their colleague Alexander Esenin-Volpin, who had been sent to a mental hospital for political reasons.

Mathematical life itself was not without clouds. The most violent anti-Semitism was supported not only by bureaucrats who carried out ideological surveillance and did not take part in real scientific work themselves, but also by some leading mathematicians. A distorted system of entrance exams closed off the way to mathematics for many talented people.

In the 1930s, the work of attracting young people to mathematics began to flourish. Mathematics is the unique area of science where children can begin serious work and obtain outstanding results very early. I recall A. N. Kolmogorov's story of how he became interested in mathematics. He said that one should not seriously study mathematics "too early," "not before the age of 12": at an earlier age there are many competing things to do that are less intellectual. Mathematics competitions (olympiads) and clubs (circles) were organized and many interesting books were written. This mainly took place around the universities in Moscow and Leningrad, and both well-known mathematicians and brilliant young university students played the leading role. Some real changes took place in the 1960s. Olympiads began to be held for students from the whole country and mathematics circles were replaced by mathematics high schools, bringing together many children devoted to mathematics who could be taught with an intensiveness and with results not previously seen. In Moscow and Leningrad, boarding schools opened where children from far away could be taught. A. N. Kolmogorov, I. M. Gel'fand, E. B. Dynkin, and other leading mathematicians gave regular lessons in such classes. Not infrequently, students obtained their first serious results before they finished high school.

The physics-mathematics journal *Kvant* (*Quantum*) began to come out and most of the activities described here were concentrated around it. The articles I wrote that make up this book appeared in *Kvant*. I began with the story of the first two discoveries of the 19-year old Gauss, with complete proofs. It seemed to me that this possibility of following the first steps of a genius was invaluable for young people who were starting along their paths in science. Gradually, I told not only more about mathematics but also about the people who created it. I thought that it was always important to understand the people of science better, but this was especially urgent given the conditions in which we lived.

It was rather unusual for a professional mathematician to write about

its history. There were some highly qualified historians of mathematics in the country, but mathematicians were basically suspicious about historical studies, seeing in them a direction in which the official ideology could influence mathematics. There was no shortage of examples of this. The influential "Communist commissar" at Moscow University was an expert on the mathematical writings of Karl Marx.

I wanted to show the great mathematicians as living human beings. Maybe it sounds strange today that this was in contradiction with the official tradition. It would not be a gross exaggeration to say that a black-and-white picture of the world was created in which scientists were divided into progressive materialists (with no shortcomings) and reactionaries and idealists (with no merit), and whether you belonged to one category or the other was decided at a very high level. Pasternak wrote,[1]

Komu byt' zhivym i hvalimym,	Who is to be honoured and living
Kto dolzhen byt' mërtv i hulim,	And who without honour and dead
Izvestno u nas podhalimam	Nobody knows in our country
Vliyatel'nym tol'ko odnim.	Till Establishment yes-men have said.

Such a world without shades of gray probably made it easier for those at the top to keep an eye on everyone. Russian scientists had a special advantage. Their primacy (real or imagined) was carefully cultivated (disrespect to them could easily be interpreted as slander), and western scientists were rarely "fully" progressive. Today it would be funny to see biographical movies of those years. I remember Euler in a film about Lomonosov,[2] reading with great surprise and delight Lomonosov's text on the conservation of energy and verifying the law by shoving one chair towards another, which began to move on impact. In view of Euler's foreign origins the level of his progressivity was not clear, not withstanding his long work in Russia.

It seemed to me that information about the fact that mathematicians like Euler or Gauss were basically ordinary people who spent a lot of energy solving ordinary problems of life in no way disparaged them. I saw no reason to cover up the history of how the aging great Euler wanted to become a (civil) general on returning to Russia from Prussia but that Catherine the Great explained (through an intermediary!) that he could be given a rank no higher than colonel. A comparison with influential Soviet mathematicians who dreamed of becoming Heroes of Socialist Labor

[1] From "The Wind (Four Fragments about Blok)" in Boris Pasternak, *Selected Poems*, translated by Jon Stallworthy and Peter France, W. W. Norton, New York, 1983, p. 147.

[2] Mikhail Lomonosov (1711–1765) is traditionally thought of as the first Russian scientist and was influential in founding the university that carries his name today, Lomonosov Moscow University.—*Transl.*

(and twice was better!) lay on the surface. "Double heroes" were eligible to have monuments erected to them during their lives (while Euler did not achieve this honor even after his death; see the story on p. 212). At times I succeeded in some counterestablishment action. The story about Chernyshevsky writing complete nonsense about Lobachevsky and his geometry was against the rules, since Chernyshevsky was officially classified as a "revolutionary democrat," which was only one step below a "Marxist revolutionary." More often, I was put in my place: a large part of the article on Pascal, devoted to his *Pensées*, was deleted. A progressive scientist could not be a religious writer (so they bashfully tried to overlook Gogol's religious searching during the last years of his life). An article on von Neumann was rejected, since I refused to say that he was a "servant of American militarism." The last trick for getting the book published was to switch "mathematicians" and "physicists" in the title and declare it to be a book about physics: there was no chance of getting it past the publishing committee on mathematics. Life taught us to fight for survival.

It is always instructive to compare similar events separated in time. Mandel'shtam wrote,[3]

Vsë bylo vstar',	Everything's been told before,
vsë povtoritsya snova,	everything will happen again,
I sladok nam lish' uznavan'ya mig.	and all that's sweet is the instant of recognition.

But in that life such comparisons could be risky. It was hard not to compare the story of the Göttingen professors' letter to the king about violating the constitution (which interrupted the collaboration of Gauss and Weber) with the letter that mentioned Esenin-Volpin. The limits within which Cardinal Bellarmino proposed to place Galileo turned out to be fantastically gentle compared to what the Soviet ideological machine required of scientists. Pascal's tragic thoughts about the sinfulness of science acquired new nuances in the 20th century. The fate of the French scientists who were happy to have a chance to participate in governing France at the time of the Revolution had direct associations with Soviet reality.

While recently rereading what I wrote in preparing this edition, I felt that after such a long time it seems to be the writing of another person. I think that it would have been wrong to change anything. Of course texts exist independently of the context in which they were written, but all the same I decided to use the occasion to recall in this very important stage of my life when this book was written.

[3]From Tristia in Osip Mandelstam, *Complete Poetry of Osip Emilevich Mandelstam*, translated by Burton Raffel and Alla Burago, State University of New York Press, Albany, NY, 1973, p. 103.

I want to thank Alan Shuchat for his enormous selfless work in translating both editions of this book. I am also grateful to my son Daniel, who read the translation and made several suggestions. Ann Kostant's support was very important for the publication of the translation.

Preface to the Third Russian Edition

The first edition of this book appeared in 1981 in the *Kvant* (*Quantum*) collection. It was reprinted several times in large print runs until 1985, more than half a million copies were sold, and it was translated into English, French, and Japanese. The book was based on articles that were published earlier in *Kvant* magazine. In this edition, some material is added that existed in 1981 but was not included then because of strict limitations on size. Some additional chapters were written later. More than twenty years have passed since a significant part of this book was written and today I would have written much of it differently, but I preferred to limit myself only to correcting errors that have been pointed out and to inaccuracies.

Among the additional subjects we note the history of the cycloid, a curve of unusual destiny, which seemed to 17th century mathematicians to be a curve of paramount importance and figured in the research of the strongest mathematicians but turned out in the end to be a curiosity in the history of mathematics. The story of the 17th century, the heroic century of mathematical analysis, is completed by the chapter on Leibniz, one of the most surprising figures in the history of science.

The 18th century that followed is represented by a trio of the most important mathematicians of the century: Euler, Lagrange, and Laplace (the last two worked into the 19th century). By the usual logic of the history of science, this should have been a relatively quiet century of putting in order the unpolished facts accumulated during the preceding revolutionary century of differential and integral calculus. However the great genius Euler, who felt confined by the mathematics of his day, broke all the rules and

made surprising discoveries that were extraordinarily ahead of their time. At the end of the century, scientists turned out to be the objects of a critical historical experiment: the French Revolution tempted some of them with the possibility of taking a direct role in government, and this temptation cost many of them their lives. The fates of Laplace and Lagrange are two examples of the behavior of scientists under these conditions.

The 19th and 20th centuries are represented, apart from Gauss, by stories about Klein, Poincaré, and Ramanujan. Of course, this choice is random enough but their histories are instructive from our view. Finally, we brought to completion two articles about the history of projective geometry and its connections with one of the most modern theories of mathematical physics—Penrose's twistor theory. The mathematical part of this dramatic history assumes a greater degree of preparation than the rest of the book.

I want to remind the reader again that this is not a systematically written book but a collection of articles that were first of all intended for students interested in mathematics, and so wherever possible I tried to include detailed mathematical fragments in the historical tales. Since then, it has turned out that the circle of readers of this book was significantly wider. I discovered, not without surprise, that even some professional mathematicians and physicists found something in it for themselves. On the other hand, there were readers who skipped all the mathematics and found something instructive in the remainder. I would also like to warn against treating this book as a serious work on the history of mathematics: I did not work with original sources, did not thoroughly verify details, and did not furnish the text with citations and references. I only wanted to share with the reader who, like myself, loves mathematics and physics a picture that appeared to me after I became familiar with considerable historico-scientific material in connection with my professional mathematical studies. It would have been ideal for me to present this history not in serious history books (which are doubtless important) but rather in the novels of Dumas.

Although this book does not give a systematic picture of the development of mathematics, it contains significant material for reflecting on some astonishing paths in this development. I have already pointed out certain recurring subjects in the preface to the first edition. The additional chapters touch on some new and important examples (we recall the apocalyptic ideas of Leibniz and Lagrange on the coming end of mathematics). Unknowable laws govern mathematical fashion! How can we understand why Fermat, well-respected by his contemporaries, could not interest any serious 17th century mathematician in his work in arithmetic? Only as the result of a fortunate coincidence was his work continued in the next century by Euler, who passed the baton to Lagrange and Gauss, thus guaranteeing

the continued development of number theory. By contrast, projective geometry, one of the greatest achievements of human thought and discovered in the same 17th century by Desargues and Pascal, was quickly forgotten and rediscovered only in the 19th century.

I do not try to explain in this book the laws of the development of mathematics: I do not know them. I only observe this process with interest, trying to draw the reader into searching for the logic hidden within it. Does there exist a natural time for the creation of a mathematical theory? One can bring many arguments in favor of this proposition. The construction of differential and integral calculus was begun by several 17th century mathematicians at once and in the end was completed independently by Newton and Leibniz; analytic geometry was independently constructed by Descartes and Fermat. Some problems that remained unsolved for many years were suddenly solved in a short interval of time by several mathematicians at once (strangely, often by three). Non-Euclidean geometry was discovered independently by Gauss, Lobachevsky, and Bolyai; the theory of elliptic functions was constructed independently by Gauss, Abel, and Jacobi. On the other hand, there have been great scholars who were very much ahead of their time and made discoveries that did not lead naturally to advancing science. Sometimes these discoveries were welcomed by their contemporaries (in the case of Archimedes or Euler) and sometimes they were forgotten (as in the case of Nicole Oresme in the 14th century, who used coordinates and considered uniformly accelerated motion 250 years before Galileo; see also the above examples about arithmetic and projective geometry). We find the richest information about the laws of mathematical creativity in the history of Ramanujan's surprising life.

What role do personalities play in the history of mathematics? For example, how decisive for the fate of mathematics was Plato's uncompromising position on the question of the subject of mathematics, given his unlimited influence on the science of his day? Was the development of geometry as an axiomatic science predetermined, or under different circumstances could it have evolved as more of an experimental science? Did Plato's almost extreme requirement of using only a straightedge and compass in geometric constructions help or hurt? Without it, what would have been discovered about unsolvable geometric problems, algebraic equations not solvable in radicals, and transcendental numbers?

I belong to the generation of Russian mathematicians who sometimes experience an ambivalent nostalgia for the time when mathematics flourished against the background of all the horrors of Soviet reality (the word "despite" would have been out of place in this context). Mathematics was a prestigious profession that attracted many talented young people who

aspired to an intellectual activity that was relatively free from the influence of the prevailing Marxist ideology. This phenomenon has been talked about a lot during the last ten years, and we will not try to continue this important discussion here.

Today the position of mathematics has changed in an important way. I am able to observe a significant decrease in the standing of mathematics and of science in general in American life. I do not see it as a tragedy that most talented youngsters prefer professions with incomparably better prospects for financial success than a scientific career, but I am frightened by the needlessly utilitarian view of the role of mathematics in education, a view that absolutely misunderstands the unique place of mathematics in the general intellectual development of the individual. Recall that in the past all future rulers, rather than future scholars, studied geometry in Plato's Academy (the Spartans did not share this piety towards mathematics and the Romans did not include it among the values they inherited from Greek civilization). Graduates of mathematics schools in the former Soviet Union were successful far beyond mathematics. Today, many young professional mathematicians have decided to leave mathematics for careers in business. They are often successful, thanks not to some particular bit of mathematical knowledge but rather to the intellectual training they received while preparing for the mathematics profession.

In today's Russia the conditions of life have changed, and mathematics is going through difficult times. Mathematicians run into everyday problems that are unknown to their western colleagues. Glancing at some Russian newspapers one day, I thought that perhaps it was in vain that in the 18th century mathematicians had happily eliminated constructing horoscopes from their professional obligations; today it might have turned out to be a useful occupational addition.

It will soon be 50 years that I have been engaged in mathematics, and I never cease to be enraptured by this amazing science. I am used to expecting that many other people, including the young, share my love for it. This book is above all addressed to them.

I warmly thank the editor of this book, S. M. L'vovskiy, for his invaluable help in preparing this edition.

February 11, 2001
Princeton, NJ

Preface to the First Russian Edition

This book is based on articles published in *Kvant* (*Quantum*) over the course of several years. This explains a certain element of randomness in the choice of the people and events to which the stories collected in the book are devoted. However, it seems to us that the book discusses the principal events in the history of science that deserve the attention of devotees of mathematics and physics.

We cover a time span of four centuries, beginning with the sixteenth. The 16th century was a very important one for European mathematics, when its rebirth began a thousand years after the decline of ancient mathematics. Our story begins at the very moment when, after a 300-year-long apprenticeship, European mathematicians were able to obtain results unknown to the mathematicians of either ancient Greece or the East: they found a formula for the solution of the third-order polynomial equation. The events of the next series of tales begin at the dawn of the 17th century when Galileo, investigating free fall, laid the foundation for the development of both the new mechanics and the analysis of infinitely small quantities. The parallel formulation of these two theories was one of the most notable scientific events of the 17th century (from Galileo to Newton and Leibniz). We also tell of Galileo's remarkable astronomical discoveries, which interrupted his study of mechanics, and of his dramatic struggle on behalf of the claims of Copernicus. Our next hero, Huygens, was Galileo's immediate scientific successor. The subject we take is his work over the course of forty years to create and perfect the pendulum clock. A significant part of Huygens' achievements in both physics and mathematics was directly stimulated by this activity. The 17th century is also represented

here by Pascal, one of the most surprising personalities in human history. Pascal began as a geometer, and his youthful work signified that European mathematics was already capable of competing with the great Greek mathematicians in their own territory—geometry. A hundred years had passed since the first successes of European mathematics in algebra.

Towards the end of the 18th century, mathematics unexpectedly found itself with no fundamental problems on which the leading scholars would otherwise have concentrated their efforts. Some approximation of mathematical analysis had been constructed; neither algebra nor geometry had brought forth suitable problems up to that time. Celestial mechanics "saved the day." The greatest efforts of the best mathematicians, beginning with Newton, were needed to construct the theory of motion of heavenly bodies, based on the law of universal gravitation. For a long time, almost all good mathematicians had considered it a matter of honor to demonstrate their prowess on some problem of celestial mechanics. Even Gauss, to whom the last part of this book is devoted, was no exception. But Gauss came to these problems as a mature scholar, and instead made his debut in an unprecedented way. He solved a problem that had been outstanding for 2000 years: He proved it was possible to construct a regular 17-gon with a straightedge and compass. The ancient Greeks had known how to construct regular n-gons for $n = 2^k, 3 \cdot 2^k, 5 \cdot 2^k$, and $15 \cdot 2^k$, and had spent much energy on unsuccessful attempts to devise a construction for other values of n. From a technical point of view, Gauss' discovery was based on arithmetical considerations. His work summed up a century and a half of converting arithmetic from a collection of surprising facts about specific numbers, accumulated from the deep past, into a science. This process began with the work of Fermat and was continued by Euler, Lagrange, and Legendre. It was startling that the young Gauss, with no access to the mathematical literature, independently reproduced most of the results of his great predecessors.

Observing the history of science from points chosen more or less at random turns out to be instructive in many ways. For example, numerous connections revealing the unity of science in space and time come into view. Connections of a different kind are revealed in the material considered in this book: the immediate succession from Galileo to Huygens, Tartaglia's ideas on the trajectory of a projectile carried by Galileo to a precise result, Galileo's profiting from Cardano's proposal for using the human pulse to measure time, Pascal's problems on cycloids being opportune for Huygens' work on the isochronous pendulum, the theory of motion of Jupiter's moons, which were discovered by Galileo, to which scholars of several generations tried to make some small contribution, and so on.

One can note many situations in the history of science that repeat, often with small variations (in the words of the French historian de Tocqueville, "history is an art gallery with few originals and many copies"). Consider, for example, how the evaluation of a scientist changes over the centuries. Cardano had no doubt that his primary merit lay in medicine and not in mathematics. Similarly, Kepler considered his main achievement to be the "discovery" of a mythical connection between the planetary orbits and the regular polyhedra. Galileo valued none of his discoveries more than the erroneous assertion that the tides prove the true motion of the earth (to a significant extent, he sacrificed his material well-being for the sake of its publication). Huygens considered his most important result to be the application of the cycloid pendulum to clocks, which turned out to be completely useless in practice, and Huygens could have considered himself generally unsuccessful since he could not solve his greatest problem—to construct a naval chronometer (much of what is considered today to be his fundamental contribution was only a means for constructing naval chronometers).

The greatest people are defenseless against errors of prognosis. In fact, a scientist sometimes makes the critical decision to interrupt one line of research in favor of another. Thus, Galileo refused to carry through to publication the results of his twenty-year-long work in mechanics, first being diverted for a year to make astronomical observations and then essentially ceasing scientific research, in the true sense of the word, for twenty years in order to popularize the heliocentric system. A century and a half later, Gauss' work on elliptic functions remained unpublished, again for the sake of astronomy. Probably neither foresaw how long the interruption would be, and neither saw around him anyone who could have threatened his priority. Galileo succeeded in publishing his work in mechanics after 30 years(!), when the verdict of the Inquisition closed off for him the possibility of other endeavors. Only a communication by Cavalieri about the trajectory of a projectile being parabolic forced Galileo to worry a bit, although it did not encroach on his priority. Gauss did not find time to complete his results, also for thirty years, and they were rediscovered by Abel and Jacobi.

The selection of material and the nature of its presentation were dictated by the fact that the book and the articles on which it is based are addressed to lovers of mathematics and physics and, most of all, to students. We have always given priority to a precise account of specific scientific achievements (Galileo's work in mechanics, Huygens' mathematical and mechanical research in connection with pendulum clocks, and Gauss' first two mathematical works). Unfortunately, this is not always possible, even with ancient works. There is no greater satisfaction than following

the flight of fancy of a genius, no matter how long ago he lived. It is not only a matter of this being beyond the reach of the amateur in the case of contemporary works. To be able to feel the revolutionary character of an achievement of the past is an important part of culture.

We wish to stress that the tales collected in this book do not have the nature of texts in the history of science. This is revealed in the extensive adaptation of the historical realities. We freely modernize the reasoning of our scientists: we use algebraic symbols in Cardano's proofs, we introduce free-fall acceleration in Galileo's and Huygens' calculations (in order not to bother the reader with endless ratios), we work with natural logarithms instead of Naperian ones in the story of Napier's discovery, and we use Galileo's latest statements in order to reconstruct the logic of his early studies in mechanics. Throughout, we consciously disregard details that are appropriate for a work in the history of science in order to present vividly a small number of fundamental ideas.

Translator's Note

Wherever possible, citations have been made to English versions of the works discussed in the book. In addition, since many of the quotations that appear were taken from various European languages (including English), I have tried to use existing translations or work directly from the original. It has been difficult to locate the sources of some quotations and these have thus been translated twice, first into Russian and then into English, and inaccuracies may have crept in. There is an apocryphal story about a computer that translated "the spirit is willing but the flesh is weak" into Russian and back again, ending up with "the wine is strong but the meat is rancid." I trust these results are more palatable! *A. S.*

Ars Magna (The Great Art)

In 1545 a book by Gerolamo Cardano appeared whose title began with the Latin words *Ars Magna*. It was essentially devoted to solving third- and fourth-order equations, but its value for the history of mathematics far surpassed the limits of this specific problem. Even in the 20th century, Felix Klein, evaluating this book, wrote, "This most valuable work contains the germ of modern algebra, surpassing the bounds of ancient mathematics."

The 16th century was the century in which European mathematics was reborn after the hibernation of the Middle Ages. For a thousand years the work of the great Greek geometers was forgotten, and in part irrevocably lost. From Arab texts, the Europeans learned not only about the mathematics of the East but also about the ancient mathematics of the West. It is characteristic that in the spread of mathematics across Europe a major role was played by traders, for whom journeys were a means of both obtaining information and spreading it. The figure of Leonardo of Pisa (1180–1240), better known as Fibonacci (son of Bonacci), especially stands out. His name is immortalized by a remarkable numerical sequence (the Fibonacci numbers). Science can lose its royal status very quickly and centuries may be needed to reestablish it. For three centuries European mathematicians remained as apprentices, although Fibonacci undoubtedly did some interesting work. Only in 16th century Europe did significant mathematical results appear that neither the ancient nor the Eastern mathematicians knew. We are talking about the solution of third- and fourth-degree equations.

Typically, the achievements of the new European mathematics were in algebra, a new field of mathematics that arose in the East and was essen-

tially taking only its first steps. For at least a hundred years, it would be beyond the power of the European mathematicians not only to achieve something in geometry comparable to the great geometers Euclid, Archimedes, and Apollonius, but even to master their results fully.

Legend ascribes to Pythagoras the phrase "all is number." But after Pythagoras, geometry gradually came to dominate all of Greek mathematics. Euclid even put the elements of algebra into geometric form. For example, a square was divided, by lines parallel to its sides, into two smaller squares and two equal rectangles. The formula $(a+b)^2 = a^2 + b^2 + 2ab$ was obtained by comparing areas. But to be sure, there was no algebraic notation at the time, and expressing the result in terms of areas was definitive. Mathematical statements were very awkward. In essence, construction problems with straightedge and compass led to solving quadratic equations and to considering expressions that contained square roots (quadratic irrationals). For example, Euclid considered expressions of the form

$$\sqrt{(a+\sqrt{b})}$$

in detail (in different language). To a certain extent, the Greek geometers understood the link between the classical unsolved construction problems (duplicating a cube and trisecting an angle) and cubic equations.

With the Arab mathematicians, algebra gradually became distinct from geometry. However, as we will see below, the solution of the cubic equation was obtained by geometric means (the debut of algebraic formulas for solving even the quadratic equation came only with Bombelli in 1572). The algebraic assertions of the Arab mathematicians are stated as recipes for the solution of one-of-a-kind arithmetic problems, usually of an "everyday" sort (for example, dividing an inheritance). Rules are formulated for specific examples but so that similar problems can be solved. Until recently rules for solving arithmetic problems (the rule of three,[1] and so on) were sometimes stated this way. Stating rules in general form almost inevitably requires a developed symbolism, which was still far off. The Arab mathematicians did not go further than solving quadratic equations and some specially chosen cubics.

The problem of solving cubic equations bothered both the Arab mathematicians and their European apprentices. A surprising result in this direction belongs to Leonardo of Pisa. He showed that the roots of the equation $x^3 + 2x^2 + 10x = 20$ cannot be expressed in terms of Euclidean irrationals of the form

[1] A mechanical way of solving proportion problems.—*Transl.*

$$\sqrt{(a+\sqrt{b})}.$$

This statement is startling for the beginning of the 13th century and foreshadows the problem of solving equations in radicals, which was thought of significantly later. Mathematicians did not see the path that led to solving the general cubic equation.

The state of mathematics at the turn of the 16th century was summed up by Fra Luca Pacioli (1445–1514) in his book, *Summa de Arithmetica* (1494), one of the first printed mathematics books and written in Italian rather than Latin.[2] At the end of the book he states that "the means [for solving cubic equations] by the art of algebra are not yet given, just as the means for squaring the circle are not given." The comparison sounds impressive, and Pacioli's authority was so great that most mathematicians (even including our heroes at first, as we shall see) believed that the cubic equation could not be solved in general.

Scipione dal Ferro

There was a man who was not deterred by Pacioli's opinion. He was a professor of mathematics in Bologna named Scipione dal Ferro (1465–1526), who found a way to solve the equation

$$x^3 + ax = b. \tag{1}$$

Negative numbers were not yet in use and, for example,

$$x^3 = ax + b \tag{2}$$

was thought of as a completely different equation! We have only indirect information about this solution. Dal Ferro told it to his son-in-law and successor on the faculty, Annibale della Nave, and to his student Antonio Maria Fior. The latter decided, after his teacher's death, to use the secret confided to him to become invincible in the problem-solving "duels" that were then quite widespread. On February 12, 1535, Niccolò Tartaglia, one of the major heroes of our story, nearly became his victim.

Niccolò Tartaglia

Tartaglia was born around 1500 in Brescia into the family of a poor mounted postman named Fontana. During his childhood, when his native city was captured by the French, he was wounded in the larynx and thereafter spoke with difficulty. Because of this he was given the nickname "Tartaglia"

[2] Despite its title.—*Transl.*

The only known portrait of Niccolò Tartaglia.

(stutterer). Early on he came under the influence of his mother, who tried to enroll him in school. But the money ran out when the class reached the letter "k," and Tartaglia left school without having learned to write his name. He continued to study on his own and became an "abacus master" (something like an arithmetic teacher) in a private school for commerce. He traveled a lot throughout Italy until landing in Venice in 1534. Here his scientific studies were stimulated by contact with engineers and artillerymen of the famed Venetian arsenal. They asked Tartaglia, for example, at what angle to aim a gun so that it shoots the farthest. His answer, a 45° angle, surprised his questioners. They did not believe that they had to raise the barrel so high, but "several private experiments" proved he was right. Although Tartaglia said he had "mathematical reasons" for this assertion, it was more of an empirical observation (Galileo gave the first proof).

Tartaglia published two books, one a sequel of the other: *La Nuova Scientia* (*The New Science* [*of Artillery*], 1537) and *Quesiti et Inventioni Diverse* (*Problems and Various Inventions*, 1546), where the reader is promised "...new inventions, not stolen from Plato, from Plotinus, or from any other Greek or Roman, but obtained only by art, measurement, and reasoning." The books were written in Italian in the form of a dialogue, which was later adopted by Galileo. In several respects, Tartaglia was Galileo's predecessor. Although in the first of these books he followed Aristotle in saying

that a projectile launched at an angle first flies along an inclined straight line, then along a circular arc, and finally falls vertically, in the second book he wrote that the trajectory "does not have a single part that is perfectly straight." Tartaglia was interested in the equilibrium of bodies on an inclined plane and in free fall (his student Giovanni Benedetti (1530–1590) convincingly showed that the behavior of a falling body does not depend on its weight). Tartaglia's translations of Archimedes' and Euclid's work into Italian and his detailed commentaries played an important role (he called Italian the "national" language, as opposed to Latin). In his personal qualities Tartaglia was far from irreproachable and was very difficult with interpersonal relations. Bombelli (who was admittedly not impartial; more on him later) wrote that "this man was by nature so inclined to speak badly, that he took any sort of abuse as a compliment." According to other information (Pedro Nuñes) "he was at times so excited that he seemed mad."

Let us return to the duel before us. Tartaglia was an experienced combatant and hoped to win an easy victory over Fior. He was not frightened even when he discovered that all thirty of Fior's problems contained equation (1), for various values of a and b. Tartaglia thought that Fior himself could not solve these problems, and hoped to unmask him: "I thought that not a single one could be solved, because Fra Luca [Pacioli] assures us of their difficulty that such an equation cannot be solved by a general formula." After fifty days, Tartaglia was supposed to submit the solution to a notary. When the time limit had almost elapsed, he heard a rumor that Fior had a secret method for solving equation (1). He was not pleased by the prospect of hosting a victory meal for Fior's friends, one friend for each problem the victor solved (those were the rules!). Tartaglia put forth a titanic effort, and fortune smiled on him eight days before the deadline of February 12, 1535: He found the method he had hoped for! He solved all the problems in two hours. His opponent did not solve a single one of the problems Tartaglia had given him. Strangely enough, Fior could not handle one problem that could be solved by dal Ferro's formula (Tartaglia had posed it with a certain trick in mind for solving it), but we will see that the formula is not easy to use. Within a day Tartaglia found a method for solving equation (2).

Many people knew about the Tartaglia–Fior duel. In this situation a secret weapon could not help but could rather hurt Tartaglia in further duels. Who would agree to compete with him if the outcome were predetermined? All the same, Tartaglia turned down several requests to reveal his method for solving cubic equations. But one who made the request achieved his goal. This was Gerolamo Cardano, the second hero of our tale.

Gerolamo Cardano

He was born in Pavia on September 24, 1501. His father, Fazio Cardano, an educated lawyer with broad interests, was mentioned by Leonardo da Vinci. Fazio was his son's first teacher. After graduating from the University of Padua, Gerolamo decided to devote himself to medicine. But he was an illegitimate child and so was denied admission to the College of Physicians in Milan. Cardano practiced in the provinces for a long time until August, 1539 when the college admitted him anyway, specially changing the rules to do so. Cardano was one of the most famous doctors of his time, probably only second to his friend, Andreas Vesalius. In his declining years, Cardano wrote his autobiography, *De Vita Propria Liber* (*The Book of My Life*). It contains recollections of his mathematical work, as well as detailed descriptions of his medical research. He claimed that he prescribed cures for up to 5000 difficult diseases and solved some 40,000 problems and questions, as well as up to 200,000 smaller ones. Of course these figures should be taken with a large dose of skepticism, but Cardano was undoubtedly a famous physician. He described cases from his medical practice where he focused on curing noted personalities (Archbishop James Hamilton of Scotland, Cardinal Morone, etc.), claiming that he had only three failures. In modern terms he was evidently an outstanding diagnostician, but he did not pay great attention to anatomical information, unlike Leonardo da Vinci and Vesalius. In his autobiography Cardano places himself alongside Hippocrates, Galen, and Avicenna (the latter's ideas were especially close to his own).

However, medical studies did not fill up Cardano's time. In his free moments he studied everything under the sun. For example, he constructed horoscopes for persons both living and dead (Christ, King Edward VI of England, Petrarch, Dürer, Vesalius, and Luther). These studies harmed his reputation among his successors (according to one unkind legend, Cardano committed suicide in order to confirm his own horoscope). But we must remember that at that time astrology was completely respectable (astronomy was a part of astrology—natural astrology as opposed to the astrology of predictions). The pope himself utilized the work of Cardano the astrologer.

In his scientific activities Cardano was an encyclopedist, but a lone encyclopedist, which was typical for the time of the Renaissance. Only after a century and a half did the first academies appear, in which scholars specialized in more or less narrow fields. Real encyclopedias could only be created with such collaborative efforts. The lone encyclopedist was in no position to verify much of the information he was given. In Cardano's case a large role was played by the peculiarities of his personality and psycho-

Gerolamo Cardano.

logical bent. He believed in magic, premonitions, demons, and in his own supernatural ability. He described in detail the events that convinced him of this (there was no bleeding in any collision he saw, from neither people nor animals, not even in hunting; he learned in advance, from signs, about the events leading up to his son's death, etc.). Cardano believed he possessed a gift of vision (he called it a "harpocratic" feeling) that allowed him to divine both an inflamed organ in an ill patient and the fall of the dice in a game of chance, and to see the mark of death on an interlocutor's face. Dreams, which he remembered in the finest detail and described carefully, played a great role in his life. Contemporary psychiatrists have used these descriptions to try to determine his disease. Cardano writes that constantly recurring dreams, together with the desire to immortalize his name, were his main reasons for writing books. In his encyclopedias *De Subtilitate Rerum* (*On Subtlety*) and *De Rerum Varietate* (*On a Variety of Matters*), he again gave a lot of space to descriptions of the author and his father.

But these books also contain many personal observations and carefully digested communications from others. His readiness to discuss fantastic theories and his peculiar credulity did not only play a negative role. Thanks to them, he discussed things that his more careful colleagues decided to speak of only many years later (see below about complex numbers). It does not always pay to follow authority. It is not clear how familiar Cardano was with the works of Leonardo da Vinci (this also applies to other 16th

century Italian authors; Leonardo became widely known only at the very end of the 18th century). *De Subtilitate Rerum* was brought to France and served as a popular textbook on statics and hydrostatics throughout the 17th century. Galileo employed Cardano's instructions for using the human pulse to measure time (in particular, for observing the oscillations of the cathedral chandelier). Cardano asserted that perpetual motion is impossible, some of his remarks can be interpreted as the principle of virtual displacements (according to Pierre Duhem (1861–1916), the well-known historian of physics), and he studied the expansion of steam. Cardano adhered to the theory, first conceived of in the 3rd century B.C., that explained the tides by the motion of the moon and sun. He was the first to clearly explain the difference between magnetic and electrical attraction (we have in mind the type of phenomenon observed as early as Thales (c.640–c.546 B.C.), such as the attraction of straw to polished amber).

Cardano was no stranger to experimental research either, or to the construction of practical devices. In his declining years he established experimentally that the ratio of the density of air to water is 1/50. In 1541, when King Charles V of Spain conquered Milan and entered the city in triumph, Cardano, as Rector of the College of Physicians, walked alongside him near the baldachin (canopy). In response to the honor shown to him, he offered to supply the royal team with a suspension from two shafts, which would keep the coach horizontal when it rocked (the roads in Charles' empire were long and bad). Such a system is now called a Cardan suspension (Cardan shaft, Cardan joint) and is used in automobiles. The truth requires us to note that the idea of such a system arose in antiquity and that, at the very least, there is a drawing of a ship's compass with a Cardan suspension in Leonardo da Vinci's *Codice Atlantico*. Such compasses became common during the first half of the 16th century, obviously without Cardano's influence.

Cardano wrote a great many books, of which some were published, some remained as manuscripts, and some were destroyed by him in Rome in anticipation of arrest. His voluminous book, *De Libris Propriis* (*On My Own Books*), contained only a description of the books he had written. His books on philosophy and ethics were popular for many years, and *On Consolation* was translated into English and influenced Shakespeare. Some Shakespeare-philes even claim that Hamlet speaks his monologue "To be or not to be. . ." while holding this book in his hands.

Much can be said about Cardano's personality. He was passionate, quick-tempered, and often played games of chance. Cardano gambled at chess for forty years ("I could never express in a few words how much damage this caused my home life, without any compensation") and at dice

for twenty-five ("but dice harmed me even more than chess"). From time to time he threw away his studies for gambling, and fell into unpleasant situations. A collateral product of Cardano's passion was *Liber de Ludo Aleae* (*The Book on Games of Chance*), written in 1526 but published only in 1663. This book contains the beginnings of probability theory, including a preliminary statement of the law of large numbers, some combinatoric questions, and observations on the psychology of gamblers.

Here are a few words about Cardano's nature. He himself writes, "This I recognize as unique and outstanding among my faults—the habit, which I persist in, of preferring to say above all things what I know to be displeasing to the ears of my hearers. I am aware of this, yet I keep it up willfully.... And I have made many, nay, numberless blunders, wherever I wished to mingle with my fellows.... I blundered, almost unavoidably, not solely because of lack of deliberations, and an ignorance of... manners and customs, but because I did not duly regard certain of those conventions which I learned about long afterwards, and with which cultivated men, for the most part, are acquainted."[3] For friends and students he could be yet another person. Bombelli wrote that Cardano had "a more godlike than human appearance."

Cardano and Tartaglia

Towards 1539, Cardano was completing his first mathematical book, *Practica Arithmeticae Generalis*, envisioned to replace Pacioli's book. Cardano burned with desire to adorn his book with Tartaglia's secret. At his request, the bookseller Zuan Antonio da Bassano met with Tartaglia in Venice on January 2, 1539. He asked Tartaglia, in the name of "a worthy man, physician of Milan, named Messer Gerolamo Cardano," to give him the rule for solving equation (1), either to publish in the book or under promise to keep it secret. The response was negative: "Tell his Excellency that he must pardon me, that when I publish my invention it will be in my own work and not in that of others...."[4] Tartaglia also refused to communicate the solutions to Fior's thirty problems and only stated the questions (which could have been obtained from the notary), and refused to solve seven problems sent by Cardano. Tartaglia suspected that Cardano was a straw man for the mathematician Zuanne de Tonini da Coi, who had long been trying

[3]From Jerome Cardan (Gerolamo Cardano), *The Book of My Life*, translated by Jean Stoner, E. P. Dutton, New York, 1930. Reprinted with the permission of E. P. Dutton, a division of NAL Penguin, Inc.

[4]Øystein Ore, *Cardano, the Gambling Scholar*, Princeton University Press, Princeton, NJ, 1953 (© renewed 1981). Reprinted with permission.

unsuccessfully to learn the secret.

On February 12th, Cardano sent Tartaglia comments about his book, *La Nuova Scientia*, and repeated his requests. Tartaglia was implacable, agreeing to solve only two of Cardano's problems. On March 13th Cardano invited Tartaglia to visit him, expressed interest in his artillery instruments, and promised to present him to the Marchese del Vasto, the Spanish governor of Lombardy. Evidently, this perspective enticed Tartaglia, since he accepted the invitation and the critical meeting took place on March 25th at Cardano's home.

Here is an excerpt from the notes of this meeting (one must keep in mind that the record was made by Tartaglia; Ferrari, Cardano's student, claimed that it does not completely correspond to the facts):

"**Niccolò:** I say to you: I refused you not just because of this one chapter and the discoveries made in it, since this is the key that unlocks the way to the study of countless other areas. I would have long ago found a general rule for many other problems, if I had not at present been occupied with translating Euclid into the national language (I have now brought the translation up to Book XIII). But when this task, which I have already begun, is done, I plan to publish the work for practical application together with a new algebra. . . . If I give it to some theorist (such as your Excellency), then he could easily find other chapters with the help of this explanation (for it is easy to apply this explanation to other questions) and publish the fruit of my discovery under his own name. All my plans would be ruined by this.

Messer Gerolamo: I swear to you by the Sacred Gospel, and on my faith as a gentleman, not only never to publish your discoveries, if you tell them to me, but I also promise and pledge my faith as a true Christian to put them down in cipher so that after my death no one shall be able to understand them. If I, in your opinion, am trustworthy then do it, and if not then let us end this conversation.

Niccolò: If I did not believe an oath such as yours then, of course, I myself would deserve to be considered a nonbeliever."

Thus, Tartaglia convinced himself. He communicated his solution in the form of a Latin poem. Is it not true that it is hard to understand from these notes *what* induced Tartaglia to change his decision? Was he really shaken by Cardano's vow? What happened later is not well understood. Having revealed his secret, the uneasy Tartaglia left immediately, refusing to meet the marchese for whom he had undertaken the journey. Could Cardano have hypnotized him? In all likelihood, Tartaglia's account is inaccurate.

Tartaglia was somewhat reassured when on May 12th he received the *Practica Arithmeticae Generalis*, freshly printed, without his recipe. In an ac-

companying letter, Cardano wrote, "I have verified the formula and believe it has broad significance."

Cardano received from Tartaglia a ready-to-use method for solving equation (1), without any hint of proof. He spent a great deal of effort on carefully verifying and substantiating the rule. From our standpoint it is not easy to understand the difficulty: Just substitute into the equation and verify it! But the absence of a well-developed algebraic notation made what any schoolchild today can do automatically, accessible to only a select few. Without knowing the original texts from that time we cannot appreciate how much the algebraic apparatus "economizes" thought. The reader must always keep this in mind, so as not to be deluded by the "triviality" of the problems over which passions seethed in the 16th century.

Cardano put in years of intense work trying to understand the solution of cubic equations thoroughly. He obtained a recipe (after all, they did not know how to write formulas!) for solving equations (1) and (2), as well as

$$x^3 + b = ax \tag{3}$$

and equations containing x^2. He certainly "outstripped" Tartaglia. All this happened against the background of a consolidation of Cardano's position: in 1543 he became professor at Pavia. "My knowledge of astrology," wrote Cardano, "led me to the conclusion that I would not live more than forty years and, in any case, would not reach the age of forty-five.... The year arrived that was supposed to be the last one of my life and that, on the contrary, turned out to be its beginning—namely, the forty-fourth."

Luigi Ferrari

For some time Cardano had been assisted in his mathematical work by Luigi Ferrari (1522–1565). In a list Cardano made of his fourteen students, Ferrari appears as the second chronologically and one of the three most outstanding. Cardano, believing in signs, wrote that on November 14, 1536, when the fourteen-year old Luigi and his brother arrived in Bologna, "a magpie in the courtyard chirred for such an unusually long time that we all expected someone to arrive." Ferrari was a man of phenomenal ability. He had such a stormy temper that even Cardano was sometimes afraid to speak with him. We know that at seventeen, Ferrari returned from a brawl without a single finger on his right hand. He was unreservedly devoted to his teacher and for a long time was his secretary and confidant. Ferrari's contribution to Cardano's mathematical work was quite substantial.

In 1543 Cardano traveled with Ferrari to Bologna, where della Nave allowed him to examine the papers of the late dal Ferro. They became

convinced that dal Ferro had known Tartaglia's rule. It is interesting that they evidently knew almost nothing about dal Ferro's formula. Cardano would hardly have pursued Tartaglia so energetically had he known that the same information could have been obtained from della Nave (he had not consulted him before 1543). Almost everyone now agrees that dal Ferro had the formula, that Fior knew it, and that Tartaglia rediscovered it knowing that Fior had it. However, not one of the steps in this chain has been strictly proven! Cardano spoke of it, but Tartaglia wrote at the end of his life, "...I can testify that the theorem described was not proved before by Euclid or by anyone else but only by one Gerolamo Cardano, to whom we showed it.... In 1534 [elsewhere February 4, 1535–S.G.] in Venice, I found a general formula for the equation...." It is hard to untangle this confused story.

Ars Magna

Familiarity with dal Ferro's papers, strong pressure from Ferrari, or, most likely, an unwillingness to bury the results of many years' work led Cardano to include everything he knew about cubic equations in this book, *Artis Magnae Sive de Regulis Algebraicis* (*The Great Art, or the Rules of Algebra*), which appeared in 1545. It has come to be called simply *Ars Magna* (*The Great Art*).

At the beginning, Cardano lays out the history of the problem: "...In our own days Scipione del Ferro of Bologna has solved the case of the cube and first power equal to a constant, a very elegant and admirable accomplishment. Since this art surpasses all human subtlety and the perspicuity of mortal talent and is a truly celestial gift and a very clear test of the capacity of men's minds, whoever applies himself to it will believe that there is nothing that he cannot understand. In emulation of him, my friend Niccolò Tartaglia of Brescia, wanting not to be outdone, solved the same case when he got into a contest with his [Scipione's] pupil. Antonio Maria Fior, moved by my many entreaties, gave it to me. For I had been deceived by the words of Luca Paccioli, who denied that any general rule could be rediscovered other than his own. Notwithstanding the many things which I had already discovered, as is well known, I had despaired and had not attempted to look any further. Then, however, having received Tartaglia's solution and seeking its proof, I came to understand that there were a great many other things that could also be had. Pursuing this thought and with increased confidence, I discovered these others, partly by myself and partly through Lodovico Ferrari, formerly my pupil."[5]

[5]Girolamo Cardano, *The Great Art or the Rules of Algebra*, translated by T. Richard Witmer,

In modern form, the method by which Cardano solved equation (1) can be presented in the following way. We will seek a solution to (1) in the form $x = \beta - \alpha$. Then $x + \alpha = \beta$ and

$$x^3 + 3x^2\alpha + 3x\alpha^2 + \alpha^3 = \beta^3. \tag{4}$$

Since $3x^2\alpha + 3x\alpha^2 = 3x\alpha(x + \alpha) = 3x\alpha\beta$, we can rewrite (4) as

$$x^3 + 3\alpha\beta x = \beta^3 - \alpha^3. \tag{5}$$

Let us try to choose the pair (α, β) in terms of (a, b) so that (5) coincides with (1). In order for this to happen, (α, β) must be a solution of the system

$$3\alpha\beta = a, \qquad \beta^3 - \alpha^3 = b,$$

or, equivalently,

$$\beta^3(-\alpha^3) = -\frac{a^3}{27}, \qquad \beta^3 + (-\alpha^3) = b.$$

By Vieta's theorem,[6] β^3 and $-\alpha^3$ will be the roots of the auxiliary quadratic equation

$$y^2 - by - \frac{a^3}{27} = 0.$$

Since we are seeking *positive* roots of (1), $\beta > \alpha$. This means that

$$\beta^3 = \frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}, \qquad -\alpha^3 = \frac{b}{2} - \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}.$$

Thus,

$$x = \sqrt[3]{\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}} - \sqrt[3]{-\frac{b}{2} + \sqrt{\frac{b^2}{4} + \frac{a^3}{27}}}.$$

When a and b are positive, the root x is thus also positive.

The calculation presented here follows only the idea of Cardano's argument. He himself argued geometrically: If we divide a cube of side $\beta = \alpha + x$ by planes, parallel to its faces, into one cube of side α and one of side x, then in addition to those two cubes we obtain three rectangular parallelepipeds with sides α, α, x and three with sides α, x, x. Their volumes

MIT Press, Cambridge, MA, 1968., pp. 8–9. Here and in other places there are slight variations in names.—*Transl.*

[6] François Viète, or Vieta (1540–1604), lived after Cardano but Cardano essentially knew this result, now known as Vieta's theorem. It is a special case of a result Vieta later proved.

are related according to (4), and if we combine the parallelepipeds of different types pairwise, then we obtain (5). "When, moreover, I understood that the rule of Niccolò Tartaglia handed to me had been discovered by him through a geometrical demonstration, I thought that this would be the road to pursue in all cases."[7] Cardano may have known of an analogous argument for the quadratic equation, due to al-Khowârizmî.[8]

Equation (2) can be solved using the substitution $x = \beta + \alpha$, but it can happen that the original equation has three real roots, while the auxiliary quadratic equation has none. This is called the *irreducible* case, and it gave Cardano (and probably Tartaglia) much trouble.

Cardano solved equation (3) by an argument that was daring at the time because it played on the negativity of the root. No one before had used negative numbers so decisively, and even Cardano himself far from used them freely. He considered equations (1) and (2) separately!

Cardano also thoroughly investigated the general cubic equation $x^3 + ax^2 + bx + c = 0$, noting that, in contemporary terms, the substitution $x = y - a/3$ eliminates the x^2 term.

Cardano decided to consider not only negative numbers (he called them "purely false") but also complex numbers (these he called "truly sophisticated"). He remarked that if we operate on them according to certain natural rules, then we can ascribe complex roots to a quadratic equation having no real roots. Cardano may have arrived at complex numbers in connection with the irreducible case. (N. Bourbaki, for example, suggests this). If, in this case, we are "undaunted" in carrying out all the operations on the complex numbers that arise during the calculation, then at the end we obtain the correct values of the real roots. But there is no indication whatever that Cardano considered more than quadratic equations here. However, the argument presented for the cubic equation soon appeared—in the hands of Rafael Bombelli (1526–1573), a successor of Cardano, a hydraulics engineer from Bologna, and the author of the famous *Algebra* (1572).

Cardano understood that the cubic equation $x^3 + ax^2 + bx + c = 0$ can have three real roots and that their sum then equals $-a$. Cardano was unprecedented in making such general assertions. In algebra, as opposed to geometry, practically no proofs were given (traces of this remain today in high school mathematics!). Here is yet another of Cardano's observations: If all the terms on the left side of an equation (with positive coefficients) have greater degree than all the terms on the right side, then there is a unique

[7] Cardano, *The Great Art*, p. 52.

[8] Mohammed ibn Musa al-Khowârizmî (c.780–c.850), Persian mathematician and astronomer whose name is preserved in the word "algorithm."—*Transl.*

positive root. A whole series of important algebraic concepts comes from *Ars Magna*, e.g., the multiplicity of a root. In general, Cardano's significance in the history of mathematics is determined most of all not by specific achievements (he did not have many) but by the fact that in *Ars Magna* he saw the path along which algebra would develop.

Remarks on Cardano's Formula

Let us analyze the formula as it applies to solving $x^3 + px + q = 0$ over the real numbers. Unlike Cardano, we can allow ourselves to ignore the signs of p and q. Thus,

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

In calculating x, we must first find the square roots and then the cube roots. We obtain real square roots if $\Delta = 27q^2 + 4p^3 > 0$. The two square root terms, differing by a sign, appear in different summands. Real cube roots are unique, so when $\Delta > 0$ we obtain a unique real value for x.

By studying its graph, it is not hard to see that the cubic trinomial $x^3 + px + q$ in fact has a unique real root when $\Delta > 0$. For $\Delta < 0$ there are three real roots. For $\Delta = 0$ we have one double real root and one single real root, and when $p = q = 0$ we have the triple real root $x = 0$.

We continue with the case $\Delta > 0$ (one real root). It turns out that even if an equation with integer coefficients has an integer root, then calculating the root by the formula can lead to intervening irrational numbers. For example, $x^3 + 3x - 4 = 0$ has the unique real root $x = 1$. For this unique real root, Cardano's formula gives the expression

$$x = \sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}}.$$

This means that

$$\sqrt[3]{2 + \sqrt{5}} + \sqrt[3]{2 - \sqrt{5}} = 1.$$

But try to prove this directly! Perhaps you will find some trick, but straightforward transformations lead to cubic radicals that cannot be removed.

It may be that this explains why Fior could not solve Tartaglia's cubic equation. It probably could have been solved by guessing the answer (this is what Tartaglia had in mind), while dal Ferro's recipe led to intervening irrationals.

The situation is even more confusing in the case of three real roots. This is called the irreducible case. Here $\Delta = 27q^2 + 4p^3 < 0$ and the numbers under the cube root signs are complex. If we find the complex cube roots, then after addition the imaginary parts vanish and we obtain real numbers. But how can we reduce everything to operations on real numbers? For example, finding the square root $\sqrt{a+ib}$ can be reduced to purely real operations on a and b. If that were the case with $\sqrt[3]{a+ib} = u+iv$, then all would be in order. But when we express u and v in terms of a and b, we obtain new cubic equations that again give rise to the irreducible case. We have a vicious circle! In the end, in the irreducible case we cannot express the roots in terms of the coefficients without going beyond the real number system. In this sense the cubic equation with three real roots is unsolvable in radicals over the reals (as opposed to the quadratic equation). This situation does not often receive the attention it deserves.

The Fourth-Degree Equation

Ferrari's personal contribution, the solution of the fourth-degree equation, was also reflected in *Ars Magna*.

In modern terms, Ferrari's method for solving

$$x^4 + ax^2 + bx + c = 0 \tag{6}$$

is as follows (it is easy to reduce the full fourth-degree equation to (6)).

Introducing an auxiliary parameter t, we rewrite (6) in the equivalent form

$$\left(x^2 + \frac{a}{2} + t\right)^2 = 2tx^2 - bx + \left(t^2 + at - c + \frac{a^2}{4}\right). \tag{7}$$

We now choose a value for t so that the quadratic trinomial (in x) on the right side of (7) has two equal roots. In order for this to happen, its discriminant must be zero:

$$b^2 - 4 \cdot 2t\left(t^2 + at - c + \frac{a^2}{4}\right) = 0.$$

We have obtained an auxiliary cubic equation in t. Find some root t_0 by Cardano's formula. We can now rewrite (7):

$$\left(x^2 + \frac{a}{2} + t_0\right)^2 = 2t_0\left(x - \frac{b}{4t_0}\right)^2. \tag{8}$$

Equation (8) can be decomposed into a pair of quadratic equations giving the four desired roots.

Thus, by Ferrari's method, the fourth-degree equation reduces to solving an auxiliary cubic and two quadratic equations.

Ferrari and Tartaglia

After meeting in 1539, Cardano and Tartaglia rarely corresponded. Once a student told Tartaglia he had heard that Cardano was writing a new book. Tartaglia immediately wrote Cardano a cautioning letter but received a calming answer. Another time Cardano wanted a clarification dealing with the irreducible case, but received nothing substantive in response. It is not hard to imagine the effect on Tartaglia when *Ars Magna* appeared in 1545. In the last part of *Quesiti et Inventioni Diverse* (1546), Tartaglia published the correspondence and notes of discussions dealing with his relations with Cardano, and heaped abuse and rebuke on him. Cardano did not react to this attack, but on February 10, 1547, Ferrari answered Tartaglia. He took exception to Tartaglia's rebukes, pointed out defects in his book, rebuked him in one place for appropriating someone else's result, and found a repetition betraying a bad memory (apparently a serious accusation for the time). Finally, he challenged Tartaglia to a public debate on "Geometry, Arithmetic and the disciplines which depend on them, such as Astrology, Music, Cosmography, Perspective, Architecture, and others."[9] Ferrari was ready to discuss not only what was written in these areas by the Greek, Latin, and Italian authors but even the works of Tartaglia himself, if the latter in turn agreed to discuss the works of Ferrari.

By tradition, such a "cartel" (challenge) required "questions" in response. They appeared on February 19th. Tartaglia wanted to draw Cardano himself into the skirmish: "...I have expressed this in such calumnious and sharp words to incite his Excellency, and not you, to write me in his own hand. I have many accounts to settle with him...." The discussion of the conditions for the duel dragged on. Tartaglia began to understand that Cardano was remaining on the sidelines. Then he started to emphasize Ferrari's lack of independence, calling him Cardano's "creation" or "creature," since Ferrari had called himself that in the first cartel. All questions were addressed to both: "You, Messer Gerolamo, and you, Messer Lodovico...." Much of the correspondence is interesting. For example, the second cartel reproduces a conversation between Cardano and Tartaglia, supposedly overheard by Ferrari: "What more do you want? 'I don't want

[9] Ore, p. 88.

it divulged,' you say. And why? 'So that no one else shall enjoy my invention.' . . .Really, since we are born not for ourselves only but for the benefit of our native land and the whole human race, and when you possess within yourself something good, why don't you want to let others share it?"[10]

The correspondence continued for a year and a half, and suddenly Tartaglia resolutely agreed to a duel in Milan. Why? At the same time, in March 1548, he received a flattering invitation to his native Brescia, to give public lectures (which he had not had occasion to do previously) and to conduct private lessons "in which only certain doctors and people of definite authority will participate." Things had not been going well for him, and there is the opinion that his patrons forced Tartaglia to accept the challenge in hopes that a victory would strengthen his position. The dispute took place on August 10, 1548, in Milan in the presence of many well-known personalities, including the governor of Milan, but in Cardano's absence. Only Tartaglia's brief notes have been preserved, from which it is almost impossible to recreate the true picture. It seems that Tartaglia sustained a shattering defeat. But do not be mistaken—the debate had no relation to the problem over which the argument had arisen, just as debates as well as physical duels generally have little relation to clarifying the truth. It was hard for the tongue-tied Tartaglia to stand up in public to the sparkling young Ferrari.

The Fate of Our Heroes

Tartaglia was not retained in Brescia, and within a year and a half he returned to Venice without having received even an honorarium for his lectures. His defeat in the debate had hurt him very much. At the end of Tartaglia's life (he died in 1557), *Trattato Generale di Numeri et Misure* (*A General Treatise on Number and Measurement*) began to appear, and its publication was completed only after his death. Very little is said about cubic equations, and no trace of a great treatise on the new algebra, which Tartaglia had spoken about all his life, was discovered in his carefully preserved legacy.

By contrast, Ferrari became very famous after the duel. He gave public lectures in Rome, headed the taxation department in Milan, received an invitation to serve Cardinal Mantui, and took a hand in bringing up the emperor's son. But he left no further trace in science! Ferrari died in 1565 at the age of forty-three; according to legend, he was poisoned by his sister. Speaking of his death, Cardano recalls the lines of Martial, the Roman

[10]Ibid., p. 94.

epigrammatist:

> For those without measure life is short, rarely do they attain old age. Whatever you may love or desire, let not your indulgence go beyond bounds.[11]

Cardano outlived them both, but the end of his life was not easy. One of his sons (the doctor Giambattista, on whom Cardano had placed his greatest hopes) poisoned his wife out of jealousy and was executed in 1560. Cardano did not recover from this blow for a long time. Another son, Aldo, became a criminal and robbed his own father. In 1570, Cardano himself was sentenced to prison and his property was confiscated. The reason for his arrest is unknown, but the initiative may have come from the Inquisition. While waiting for arrest, Cardano destroyed 120 of his own books. He ended his days in Rome, a "private person" (his expression) receiving a modest pension from the pope. Cardano devoted the last year of his life to his autobiography, *De Vita Propria Liber* (*The Book of My Life*). The last item it mentions is dated April 28, 1576, and on September 21st Cardano died.

In his autobiography, Cardano mentions Tartaglia four times. In one place he approvingly cites his thought that "no one knows everything, and moreover knows nothing that he suspects many people do not know." Elsewhere he says that Tartaglia preferred him as "a rival and victor, and not a friend and a man obliged to do him well." Still, Tartaglia turns out to be among Cardano's critics who "did not go beyond grammar." Finally, on the very last pages, we read, "I confess that in mathematics I received a few suggestions, but ever few, from brother Niccolò." It seems that Cardano's soul was uneasy!

Epilogue

The Cardano–Tartaglia problem lay forgotten for a long time. The formula for solving the cubic equation was associated with *Ars Magna* and gradually became known as *Cardano's formula*, although at some time dal Ferro's name was involved (his authorship was stressed by Cardano himself). Such errors in appropriating names are not rare (e.g., recall the axiom of Archimedes, who did not claim this discovery).

The question of the origin of the formula for the cubic equation arose again at the beginning of the 19th century. The insulted Tartaglia, who had been practically forgotten, was rediscovered. The almost forgotten story received publicity, and not only professionals but even amateurs were ready

[11] Cardano, *My Life*, p. 144; Ore, p. 82.

to fight for Tartaglia's honor. The detective aspect of history was very attractive. For how many years should Cardano's promise have been in effect? Was six years really long enough? Why did Tartaglia not publish his formula for ten years? However, as the story spread through the popular literature it became distorted, and Cardano in time turned into an adventurer and villain who stole Tartaglia's discovery and gave it his own name. As we have seen, the situation was more complicated, and such an interpretation, at the very least, oversimplifies the picture.

It was not only a matter of wanting to restore the true picture of the events in a situation where the participants undoubtedly did not tell the whole truth. For many it was important to establish the degree of Cardano's guilt. This question touches on the perennially topical question of proprietary rights in scientific discovery. Concerning today's practice, what strikes us is the difference between the rights of the scientist and of the inventor. The scientist cannot control the future use of his published results but only lay claim to the remembrance of his name. Controlling future use is one of the reasons for keeping inventions secret. At the juncture of the Middle Ages and the Renaissance, mathematical results were kept secret in order to use them in duels.

Towards the end of the 19th century, part of the discussion began to take on the character of serious historico-mathematical research. Some original material was published for the first time ("cartels" and "questions"). Mathematicians understood how great a role Cardano's work had played in 16th century science. What Leibniz had remarked earlier became clear: "Cardano was a great man for all his faults; without them he would have been perfect."

Moritz Cantor (1829–1920 not to be confused with Georg Cantor, the creator of set theory), the greatest historian of mathematics and author of a multivolume history of mathematics, held Cardano in very great esteem but not without regret, stating that his human qualities left something to be desired ("genius but no character"). Cantor proposed, as had Ferrari, that Tartaglia did not rediscover dal Ferro's rule but learned it ready-made and secondhand. He remarked that Tartaglia did not have any significant mathematical works to his credit and, aside from the rule itself and facts that could have been borrowed from *Ars Magna* (which had appeared earlier), his publications and the manuscripts he left contain only elementary remarks about cubic equations. Of course this is no proof, and moreover Tartaglia had many virtues beyond mathematics. Cantor was also suspicious of the fact that Tartaglia's and dal Ferro's solutions were as similar as two drops of water. Gustav Eneström (1852–1923) took exception to Cantor and even conducted some sort of research experiment that showed such

a coincidence is possible. Ettore Bortolotti (1866–1947) did much to explain unclear points, and presented arguments that could confirm several seemingly irresponsible statements by Tartaglia.

For a century and a half, passions have calmed down and then heated up again. The desire to obtain a single answer to the question has not died away, but such an answer may simply not exist. As for the formula for the solution to the cubic equation, the name *Cardano's formula* has firmly taken hold.

Appendix[12]

Four months before his death Cardano completed his autobiography, which he had anxiously written during the entire previous year and which was supposed to sum up his complex life. He felt death approaching. According to some reports, his personal horoscope associated his end with his seventy-fifth birthday. He died on September 21, 1575,[13] a few days before his birthday. There is a version that he committed suicide in anticipation of his inevitable death or even to confirm the horoscope. In any case, Cardano the astrologist took his horoscope seriously. In his book he described waiting for death at age forty-four, as his earlier horoscope had foretold.

Cardano worried about whether his life had been successful. On the one hand, he lived on a meager papal pension in Rome, in enforced exile from the cities where he had spent the best part of his life, he had recently been in prison, and he was unhappy with his children. On the other hand, Cardano was sure of his own significance. He criticized much from his past, although it is not hard to discover the places where he succeeded in convincing himself that he was right. Cardano's leading idea is the predestination of his life. This is the source of his detailed analysis of the influence of the stars, his association with a "guardian angel," the scrupulous account of signs and omens, and the little events that allowed him to build a logically constructed picture of life. In a certain sense, Cardano's aim was, using the scholar's and astrologer's art, to analyze himself in detail as an object of the action of higher powers. A new style was established in science, where conclusions are drawn from the facts as they appear. Therefore, Cardano supplies the reader with detailed information about his physical features, drinking patterns, habits, etc., in order for the author and reader to have the same opportunities to draw conclusions. Cardano's book is a remarkable literary monument of the 16th century, and allows us to understand much

[12] From *De Vita Propria Liber* (*The Book of My Life*).

[13] Some sources say September 20th, e.g., Ore, p. 23, and Cardano, *My Life*, p. xiii.—*Transl.*

about how one of the wisest men of his time perceived life.

Cardano's book was translated into Russian in 1938 and published by the State Literary Publishing House. Let us see how Cardano talks about himself.[14] A few examples have been given in the main part of this chapter.

He begins, "This Book of My Life I am undertaking to write after the example of Antoninus the Philosopher, acclaimed the wisest and best of men, knowing well that no accomplishment of mortal man is perfect, much less safe from calumny; yet aware that none, of all ends which man may attain, seems more pleasing, none more worthy than recognition of the truth. No word, I am ready to affirm, has been added to give savor of vainglory, or for the sake of mere embellishment." Cardano describes in detail his native Milan and his ancestors. He says of his birth, "...I was... born on the 24th day of September in the year 1500 [this is evidently a slip of the pen, since Cardano was born in 1501—S.G.], when the first hour of the night was more than half run, but less than two-thirds.... I was almost dead. My hair was black and curly. I was revived in a bath of warm wine which might have been fatal to any other child." He describes in detail the positions of Mars, Mercury, and the moon, which presaged that "...consequently I ought to have been a monster, and indeed was so near it that I came forth literally torn from my mother's womb." The "sinister planets" Venus and Mercury foretold that he would be "gifted... with a certain cunning only, and a mind by no means at liberty; my every judgment is, in truth, either too harsh or too forbidding."

In another chapter, he describes his parents: "My father went dressed in a purple cloak, a garment which was unusual in our community; he was never without a small black skullcap. When he talked he was wont to stammer. He was a man devoted to various pursuits. His complexion was ruddy, and he had whitish eyes.... From his fifty-fifth year on he lacked all his teeth.[15] He was well acquainted with the works of Euclid; indeed, his shoulders were rounded from much study." What surprising details! "My mother was easily provoked; she was quick of memory and wit, and a fat, devout little woman."

Later, Cardano gives a short description of his life, after which it is the turn of his discoveries. Here are some details: "I am a man of medium height; my feet are short, wide near the toes, and rather too high at the heels, so that I can scarcely find well-fitting shoes.... My chest is somewhat narrow and my arms slender. The thickly fashioned right hand.... A neck a little long and inclined to be thin, cleft chin, full pendulous lower lip, and

[14]The following quotations are generally taken from Stoner's English translation of *My Life*, cited earlier.—*Transl.*

[15]It says elsewhere that this was after an attempt to poison him.

eyes that are very small and apparently half-closed; unless I am gazing at something. . . . My hair and beard were blond. . . . Old age has wrought changes in this beard of mine, but not much in my hair, . . . ," etc. Cardano describes the illnesses he suffered, and says, "Now I have fourteen good teeth and one which is rather weak; but it will last a long time, I think, for it still does its share." He had ten ailments in all; the tenth was insomnia, which he cured by abstaining from certain kinds of foods.

Cardano tells us that he is timid by nature but gained courage through physical exercise, that he stays in bed for ten hours but sleeps for eight, that he prefers fish to meat and counts off twenty-one kinds of fish that he eats, and that with big fish he eats "the head and belly, but the spine and tail of small ones."

"Vowing to perpetuate my name, I made a plan for this purpose. . . in some hope of the future. I have scorned the present," we read. Chance, the intrigues of his opponents, and his own astrological findings that he would not live past forty-five interfered with Cardano's aspirations to perpetuate his name. Everything changed when it turned out that the prediction had not come true. Cardano decisively changed his way of life. He delivered lectures early in the morning. "That over, I went walking in the shade beyond the city walls, and later lunched, and enjoyed some music. In the afternoon, I went fishing. . . . While there I also studied and wrote, and in the evening returned home." Cardano explains why he preferred a career in medicine to law, as his father had wished: "I deemed medicine a profession of sincerer character than law, and a pursuit relying rather upon reason and nature's everlasting law, than upon the opinion of men." He talked about his teaching and debating: "While at the University of Bologna, I usually lectured extemporaneously. . . . Excellent as I may have appeared in these respects, I possessed neither grace in my manner of speech nor talent for making a clever conclusion."

He characteristically lists his virtues: "Notwithstanding, whatsoever my good fortune, or however many the happy issues that attended me, I have never modified my carriage. . . [nor adopted] a more luxurious mode of dress. . . . To the duties of life I am exceptionally faithful, and particularly in the writing of my books. . . . I have never broken a friendship; neither, if the relationship happened to be discontinued, have I divulged the secrets of my erstwhile friends. . . ." He describes his friends and patrons in detail, but ostentatiously does not enumerate his enemies and rivals. But they appear repeatedly on the pages of the book, even as early as the next chapter, entitled "Calumny, Defamations, and Treachery of My Unjust Accusers."

Cardano begins with intrigues and experiences some difficulty in choosing examples. He wanted to talk about great and secret intrigues, but in-

trigues that have already been uncovered cannot be considered secret, and great intrigues are hard to conceal. Philosophizing, he chooses the case of obtaining a professorship in Bologna, when it was rumored that he "lectures to the empty benches. . . that he is a man of bad manners, disagreeable to all, and, in the main, a fool. He is given to disgraceful practices, and he does not show even tolerable skill in medicine. . . and accordingly has no practice." All this would have been believed if the papal legate in Bologna had not remembered that Cardano had cured his mother. This undermined confidence in the remaining information. Nevertheless the intrigues continued even in Bologna, and Cardano in the end was denied the post, although he reassured himself: ". . .all this wound up pleasing those who tried so hard for it, but it is not at all to their benefit." Concerning "calumny and lying deformation," Cardano does not dwell on concrete cases, stating that his "calumniators [were] tormented by their own guilty conscience. . . they have left me more time for collecting my literary works. . . they have provided me with an opportunity to. . . devote myself to the investigation of many things not fully revealed to man," and that "I do not hate them."

His pleasures are briefly listed: pen knives (he spent more than twenty gold crowns on them), different types of pens (more than two hundred crowns), precious stones, china, globes of painted glass, rare books, navigation, fishing, the philosophy of Aristotle and Plotinus, the occult, Petrarch's poetry, etc. He preferred solitude to company, not only out of devotion to science but so as not to lose time. We have already mentioned his partiality to gambling with chess and dice.

A separate chapter is devoted to clothing. Cardano finds a description in Horace that resembles him very much. A rather long discussion with references ends in saying that one must have "four changes of clothes: one warm, one very warm, one light, and one very light. Thus, one gets fourteen different combinations. . . ." He describes his gait, stating that the reason for its unevenness is his constant reflection. He discusses his attitude towards religion and philosophy, stressing the influence of Plato, Aristotle, Plotinus, and especially Avicenna. He lists the "general rules" that he mastered during his life: to thank God and ask Him to help, not to limit oneself to redeeming a loss but always to obtain something in addition, to make the most of time, to respect one's elders, "always to set certainties before uncertainties," "never. . . to persist willingly in any course which is turning out for the worse," and so on. Cardano lists the houses in which he lived, and colorfully describes his poverty and losing his inheritance from his father.

Cardano writes in detail about his wife and children. He writes that he saw his future wife in a dream before meeting her, and that the dream

presaged an unhappy marriage. We have already talked about the fate of his children. He describes his travels, usually in connection with his medical activities, and explains the value of the trips.

The longest chapter is devoted to perils and accidents. Cardano describes them in detail, visibly trying to impress the reader to prepare the way for more profound phenomena: "...it is not the significant event which ought to be wondered at, but rather the frequent recurrence of similar instances." Three times, he miraculously escaped danger at nearly the same place: from a tile falling from a wall, from a giant piece of masonry, and from an overturning carriage. Twice he nearly drowned under very romantic circumstances. Cardano was subjected to the attack of a rabid dog, fell into a pit, fell from a carriage onto the road, and was exposed to plague. These accounts read like detective stories. After this comes a sequence of terrible intrigues devised by his competitors, the doctors in Pavia: a scandal involving his daughter's husband, a beam that could have fallen while he entered the Academy, and a poisoning attempt that was averted by the boys who tasted his food. However, everything was unexpectedly ended by the illness or even death of his enemies. In Rome, dangers pursued Cardano because he did not know the streets and "the behavior here [is] so uncouth." But he finally decides that Providence is protecting him, and he stops fearing danger: "Who does not now perceive that all these things have been, as it were, precursors to bliss about to be overtaken...?"

Cardano includes in his book a study of happiness, with examples from his life. He lists the honors shown him, mainly flattering invitations. On the other hand, he recounts unpleasant episodes from his medical practice and discusses their association with dreams. There is an unexpected discussion of the Cardano family coat of arms, to which he decided to add a swallow on the day of his arrest: "I chose the swallow as in harmony... with my own nature: It is harmless to mankind, it does not shun association with the lowly, and is ever in contact with humankind without becoming familiar...." A questionable comparison! Cardano also lists his teachers and students.

Cardano again discussed his characteristics and surprising events in his life: While a child he had visions of ringlets in the sky, he could not warm up his legs below the knees, and blood did not flow in his presence (he even began to intervene purposely in fights and was never injured); the events leading up to his death of his oldest son; and finally the many dreams which preceded events that later occurred. The descriptions of his dreams are very colorful and detailed.

Later, Cardano lists the ten sciences he was acquainted with in his life and describes forty cases from his medical practice. Then comes a chapter,

"Concerning Natural Though Rare Circumstances of My Own Life." Here is the first of these occurrences: "...I was born in this century in which the whole world became known, whereas the ancients were familiar with but little more than a third part of it." Also, his house collapsed but his bedroom was spared, his bed caught fire twice. He analyzes in detail the gift of prognostication that constantly appeared in his life, from medicine to card games.

The concluding part of the book again deals with supernatural events, discusses scientific achievements, and list his books. Cardano again talks about himself, about his guardian angel, lists testimonials to himself, discusses "worldly things," and devotes several pages to maxims for life. Here are some examples: "Friends are your support in adversity, flatterers bring you advice.... An illustrious man ought to live under the aegis of a prince.... When you wish to wash yourself first prepare a linen towel for wiping.... Evil is but a lack of good, and good is of itself a virtue which is within our power to possess, or rather which is indispensable." After these sayings comes "A Lament on the Death of My Son." At the end, we again hear of Cardano's inadequacies, of the changes that come with age, and of the "Quality of Conversation."

Two Tales of Galileo

I The Discovery of the Laws of Motion

. . .it was Galileo who laid the first foundation of [dynamics]. Before him, only forces acting on bodies in equilibrium were considered; and although the acceleration of falling bodies and the curved motion of projectiles could only be attributed to the constant force of gravity, no one had been able to determine the laws by which these daily phenomena follow from such a simple cause. Galileo was the first to take this important step, and in so doing began a new and vast arena for the development of mechanics. . . today, it is the most significant and secure part of this great man's glory. The discoveries of the moons of Jupiter, the phases of Venus, sunspots, etc., only required a telescope and perseverance; however, it needed an extraordinary genius to unravel the natural laws of these phenomena that had always been before everyone's eyes but had nonetheless always escaped the philosophers' reach. Lagrange[1]

Prologue

Vincenzio Galilei, a well-known Florentine musician, had reflected for a long time on what field to choose for his oldest son Galileo. The son was undoubtedly talented in music, but the father preferred something more reliable. In 1581, when Galileo turned seventeen, the scales were leaning in the direction of medicine. Vincenzio understood the expenses of instruction would be great, but that his son's future would be assured. The place of instruction was chosen to be the University of Pisa, perhaps a bit provincial but familiar to Vincenzio. He had lived for a long time in Pisa, and Galileo was born there.

[1] Joseph-Louis Lagrange, *Mécanique Analytique*, 5th ed., Vol. 1, Blanchard, Paris, 1965, p. 207.

Galileo Galilei.

The road to becoming a doctor was not easy. Before beginning to study medicine, it was necessary to learn—by heart—Aristotelian philosophy. In Galileo's opinion, "it seems that there is not a single phenomenon worth attention that he [Aristotle] would have encountered without considering." At the time, Aristotelian philosophy was taught in an appalling way, namely as a selection of statements considered to be the ultimate truth, devoid of motivation or proof. One could not even talk about disagreeing with Aristotle.

What interested Galileo most of all was what Aristotle wrote about the physics of the world around him, but he did not want to believe every word of the great philosopher blindly. He mastered it by using logic: "Aristotle himself taught me to be satisfied in my mind only when the arguments convinced me, and not just the teacher's authority." He also read other authors, and Archimedes and Euclid were among those who impressed him the most.

Mysteries of Motion

Of everything that takes place in the world around us, Galileo was most interested in motion in its various forms. Bit by bit, he gathered everything the ancients wrote about motion, but regrets, "there is nothing older than motion in nature, but rather little that is significant has been written about

it." Questions came to the inquisitive youth at each step. . . .

"In 1583, at the age of about twenty, Galileo found himself in Pisa, where, on his father's advice he applied himself to the study of philosophy and medicine. And one day, being in the cathedral of that city and curious and clever as he was, he decided to observe the motion of a [hanging] chandelier that swerved from the perpendicular—whether the time it took to swing back and forth along long arcs was the same as along medium and short arcs. It seemed to him that the time to travel a long arc might be less because of the greater speed with which, as he saw, the lamp moved along the higher and more sloped sections. And since the lamp moved gradually, he had a gross estimate, as he was wont to say, of how it moved back and forth, using the beating of his own pulse, as well as the tempo of music he had with great profit practiced. And on the basis of these calculations he saw that he was not mistaken in believing that the times were the same. But not satisfied with this, on returning home he thought of doing the following, in order to make certain.

"He attached two lead balls to strings of exactly the same length so that they could swing freely. . . , and displacing them from the vertical by different numbers of degrees, for example, one by 30 and the other by 10, he released them at the same instant. With the help of a friend he observed that while one made a certain number of oscillations along long arcs, the other made exactly the same number along small ones.

"In addition he had two similar pendula, but of rather different lengths. He observed that while the short one made a certain number of oscillations, for example, 300 along its longest arcs, in the same time that the long one always made the same number, say 40, both along its longest arcs and its shortest ones; repeating this several times. . . , he concluded from this that the time to go back and forth is the same for the same pendulum, the longest or the shortest, and that there are almost no notable differences in this which must be attributed to interference by the air, which resists a faster moving heavy object more than a slowly moving one.

"He also saw that neither different absolute weights, nor different specific gravities of the balls made any manifest change in this—all, provided they hang from strings of equal length from their centers to their points of suspension, keep much the same time to travel along every arc; as long as one does not take a very light material, such as cork, whose motion in air. . . is more easily resisted and which more quickly comes to rest."

This story is due to Vincenzio Viviani[2] (1622–1703), who in 1639, at the age of seventeen, was at the villa of Arcetri near Florence, where Galileo

[2]Letter to Prince Leopoldo de' Medici, 1659, in Galileo Galilei, *Opere*, A. Favaro, ed., Vol. 19, Barbèra, Florence, 1938, pp. 648–649.

found himself after the verdict of the Inquisition. For two years, Evangelista Torricelli (1608–1647) was also there, and the two helped the blind scientist complete his projects. They obtained a series of results under Galileo's influence (famous barometric experiments and research on cycloids). Viviani was apparently especially close to Galileo, who willingly discussed various subjects with him, often recalling the distant past. Afterwards, Viviani had many occasions to retell what he had heard in those days. These tales are not considered sufficiently reliable, and it is not always clear who was the source of the inaccuracies—the narrator or the listener. Viviani's main goal in life was to immortalize his teacher's memory.

Let us return to Viviani's story. He described Galileo's discovery that a pendulum is isochronous: For a fixed length, the period of its oscillation does not depend on its amplitude. It is instructive to see how Galileo kept time, with music and his pulse (it seems that Cardano was the first to suggest this method). We in the 20th century, who are accustomed to wristwatches, should not overlook such difficulties. Rather precise clocks were constructed immediately afterwards, based on Galileo's discovery of the pendulum's property (we will have a chance to talk about pendulum clocks later). Incidentally, in laboratory experiments which we will discuss below, Galileo used slowly dripping streams of water to measure time (a variation on water clocks).

Galileo discovered a connection between the length of a pendulum and the frequency of its oscillations: The square of the period is proportional to the length. Viviani wrote that Galileo obtained this result "guided by geometry and by his new science of motion," but no one knows how he could have reached such a theoretical conclusion. Perhaps Galileo observed this relationship experimentally. He apparently did not know that the oscillations of a pendulum are only isochronous for small angles of deflection. For large angles, the period begins to depend on the angle and for 60°, for example, it is noticeably different from the period for small angles. Galileo could have noticed this in the series of experiments Viviani described. The error in Galileo's claim that a mathematical pendulum is isochronous was discovered by Huygens.

Galileo's medical studies did not go very well, although he tried to justify his father's hopes and expenditures. In 1585, he returned to Florence without having received a doctor's diploma. There, he continued to study mathematics and physics, first in secret from his father and then with his consent. Galileo was in contact with scientists, including the marchese Guidobaldo del Monte. Thanks to the latter's support, in 1589 Ferdinando de' Medici, the Grand Duke of Tuscany, appointed Galileo as professor of mathematics at the University of Pisa. Galileo remained in Pisa until

moving to Padua in 1592. He considered his eighteen years in Padua the happiest period of his life. From 1610 until the end of his life, he was "Philosopher and First Mathematician of his Highness the Grand Duke of Tuscany." In both Pisa and Padua, the study of motion was Galileo's main work.

Free Fall

Galileo was above all interested in free fall, one of the most common forms of motion in nature. At the time, one had to begin with what Aristotle said on the matter. "Bodies having a greater degree of heaviness or lightness but in all other respects having the same shape, traverse an equal space more rapidly in the same proportion as the quantities mentioned." Thus, according to Aristotle, the velocity of a falling body is proportional to its weight. A second assertion is that velocity is inversely proportional to "the density of the medium." This assertion led to complications, since in a vacuum, whose "density" is zero, the velocity should have been infinite. As to this, Aristotle declared that a vacuum cannot exist in nature ("nature abhors a vacuum").

Aristotle's first assertion was sometimes disputed, even during the Middle Ages. But Giovanni Benedetti's criticism was especially convincing. Benedetti was Tartaglia's student and Galileo's contemporary, and Galileo became familiar with his treatise in 1585. The main idea of Benedetti's refutation looks like this. Suppose we have two bodies, one heavy and one light: The heavy one should fall faster. Now combine them. It is natural to assume that the light body slows down the heavy one, and that the velocity at which the combined body falls should be intermediate between the individual velocities. But according to Aristotle, the velocity should be greater than that of each body! Benedetti declared that velocity depends on specific gravity, and even estimates that for lead it is eleven times greater than for wood. Even Galileo believed this for a long time.

Galileo began studying free fall in Pisa. Here is what Viviani writes: "...Galileo completely gave himself up to reflection, and to the great embarrassment of all philosophers, he was persuaded, by means of experiments, solid proofs, and arguments, of the falsity of very many of Aristotle's conclusions about motion that up to that time had been considered perfectly obvious and unquestionable. These include, among others, that two bodies of the same material but different weight, moving in the same medium, do not move with speeds proportional to their weights as Aristotle proposed, but with the same speed. He proved this by repeated experiments from the top of the Tower of Pisa, in the presence of other lecturers and philoso-

phers and the entire scholarly fraternity."[3] To this time, Galileo is often drawn throwing balls from the Tower of Pisa. This legend has acquired many spicy details (for instance, the bartender who started the rumor that Professor Galilei would jump from the tower). Note that so far only bodies of the same substance are being discussed.

Galileo studied Benedetti's observation that the velocity of free fall increases according to the body's motion and decided to find a mathematically precise description of this change. Here we should say that from the start Galileo saw his problem as how to quantify Aristotle's physics: "Philosophy is written in that great book, which ever lies before our eyes (I mean the universe); but we cannot understand it if we do not first learn the language and grasp the symbols in which it is written. The book is written in the mathematical language and the symbols are triangles, circles, and other geometrical figures. . . ."[4] However, it soon became clear that quantification requires a systematic review of all the facts.

How, then, to find the law by which the velocity of free fall changes? An experiment was only the beginning of scientific research. Aristotle and his followers considered experimentation unnecessary and worthless, for both establishing and verifying the truth. Galileo could have tried to conduct a series of experiments on bodies in free fall, carry out careful measurements, and search for a law that would explain them. This is the way Galileo's contemporary, Kepler, working with Tycho Brahe's many observations, discovered that the planets move along ellipses. But Galileo chose a different route. He decided to guess the law first from general considerations, and then to verify it experimentally. No one had done this before, but gradually this way of doing research has become one of the leading methods of establishing truth in science.

Now let us see how Galileo tried to guess the law. He decided that nature "tries to take the simplest and easiest way in all its adaptations," which means that the law by which the velocity grows must be "in the simplest and most universally clear form." But since the velocity grows according to the distance traveled, what could be simpler than assuming it is proportional to the distance: $v = cs$, where c is a constant. This was wrong from the start; after all, it would imply that free fall begins with zero velocity, while the velocity is apparently large from the very beginning. But here is an argument that convinced him there is no contradiction: "...is it not true that if a block be allowed to fall upon a stake from a height of four cubits and drives it into the earth, say, four finger-breadths, that coming from a height of two cubits it will drive the stake a much less distance,

[3] Ibid., letter of 1654, p. 606.
[4] Galileo, *Opere*, Vol. 4, p. 171.

and from the height of one cubit a still less distance; and finally if the block be lifted only one finger-breadth how much more will it accomplish than if merely laid on top of the stake without percussion? Certainly very little. If it be lifted only the thickness of a leaf, the effect will be altogether imperceptible. And since the effect of the blow depends upon the velocity of this striking body, can anyone doubt the motion is very slow and the speed more than small whenever the effect [of the blow] is imperceptible?"[5]

For a long time, Galileo studied the various consequences of this assumption and unexpectedly discovered that . . .*according to such a law motion cannot take place at all*! Let us also try to see why this is so. The proportionality coefficient depends on the choice of the unit of time. For simplicity, we will assume that $c = 1$, that distance is measured in meters (m) and time in seconds (sec). Then at all moments of time, $v = s$.

Consider a point A one meter from the origin O. Let us estimate at what time the body reaches this point after it begins to move. At A the velocity equals 1 meter per second (m/sec). Take the point A_1 halfway between O and A. At each point in the interval A_1A the instantaneous velocity will be less than 1 m/sec, and since this interval has length $\frac{1}{2}$ m, more than $\frac{1}{2}$ sec is required to traverse it. Now take the point A_2 halfway between O and A_1. On A_2A_1 the instantaneous velocity will be less than $\frac{1}{2}$ m/sec (all its points are less than $\frac{1}{2}$ m from O), and since its length is $\frac{1}{4}$ m it too requires more than $\frac{1}{2}$ sec. Of course, you have already guessed the rest of the argument: A_3 is the midpoint of OA_2, the length of A_3A_2 is $\frac{1}{8}$ m, the velocity is less than $\frac{1}{4}$ m/sec, and again more than $\frac{1}{2}$ sec is required, etc. The division process can be continued endlessly, and we can choose any number of intervals, each requiring more than $\frac{1}{2}$ sec, without reaching O. This means that a body leaving O cannot arrive at A!

We assumed that A is at a distance of 1 m from O. But it can be shown analogously that a body leaving O can reach no point whatsoever. This remarkable argument was the beginning of classical mechanics!

However, Galileo himself published an unconvincing argument. He tried to reach a contradiction by saying that since velocity is proportional to distance, all intervals beginning at the origin must be traversed in the same time, which is impossible. Either Galileo was not yet used to working with instantaneous velocity, or he originally had some other argument that he could not reconstruct when he wrote down these results, in his old age, after a long interruption (we will see why this happened). He left more

[5]Galileo, *Discorsi e Dimostrazioni Matematiche, Intorno à Due Nuove Scienze* (*Dialogues Concerning Two New Sciences*), translated by Henry Crew and Alfonso de Salvio, Dover, New York, 1954, pp. 163–164.

than a few claims that were either unmotivated or supported by doubtful arguments.

Well, Galileo had every reason to be offended by the perfidy of nature, which did not choose the simplest path, but he did not lose his belief in nature's reasonableness. He considered a no-less-simple assumption, that the velocity grows proportionally with time: $v = at$. He called this naturally accelerated motion, but the term "uniformly accelerated motion" has survived. Galileo considered the graph of the velocity on the time interval from O to t and remarked that if we take moments of time t_1, t_2 equidistant from $\frac{t}{2}$, then the velocity at t_2 is greater than $\frac{at}{2}$ by the same amount that it is less than $\frac{at}{2}$ at t_1. From this he concluded that at the midpoint the velocity equals $\frac{at}{2}$, and the distance traveled equals $\frac{at}{2}t = \frac{at^2}{2}$ (not too rigorous an argument!). This means that *if we consider equally spaced moments of time* $t = 1, 2, 3, 4, \ldots$ *then the corresponding distances traveled from the origin will be proportional to the squares* $t^2 = 1, 4, 9, 16, \ldots$, *and the distances traveled between adjacent times proportional to the odd integers* $1, 3, 5, 7, \ldots$.

Again, let us follow Galileo's logic. First he separated the questions "how" and "why." For Aristotle's followers, the answer to the first question had to be an immediate consequence of the answer to the second. But Galileo, soberly evaluating his chances, did not investigate the origin of accelerated free-fall motion in nature, but instead only tried to describe the law by which it occurs. The main thing was to search for a simple, general principle from which this law can be deduced. He sought "a completely unquestioned principle, that can be taken as an axiom." Galileo's statements in a letter to Paolo Sarpi[6] in the autumn of 1604 can be interpreted to say that he already knew the law by which the distance traveled changes during free fall but was dissatisfied because he could not deduce it from an apparently unquestioned principle: "A body experiencing natural motion increases its velocity in the same proportion as its distance from the initial point."

Here it was important to choose a fundamental independent variable, relative to which the change of all quantities characterizing motion could be considered. It is very natural to begin by choosing the distance traveled as this variable; after all, an observer sees how the velocity grows as the distance grows. We have already said that the measurement of time did not yet play a significant role in people's lives and that precise clocks were not available. We do not always take into account how gradually the sensation of constantly passing time took root in human psychology. Galileo showed

[6](1522–1643), an influential Venetian priest and theologian who was a benefactor of Galileo.—*Transl.*

great flexibility in reorienting himself comparatively quickly from distance to time. During 1609–1610 he discovered the true principle that free fall is uniformly accelerated (with respect to time!).

We must not overemphasize the final form of Galileo's ideas of velocity and acceleration. The idea of an instantaneous, continuously changing velocity is not easy to sense, and it gained acceptance slowly. It was hard to convince oneself that rejecting abrupt changes in velocity did not lead to reasoning about continuous procedures that was overflowing with contradictions. It is difficult for us today to judge Galileo's courage in working with a varying velocity so decisively. Such masters of analytical arguments as Cavalieri, Mersenne, and Descartes did not believe him. Descartes categorically rejected motion with zero initial velocity, in which a body "passes through all degrees of slowness." The process of calculating distances under a varying velocity was even more complicated, and required integration. Galileo possessed only a form of integration similar to Archimedes' method or to Cavalieri's "indivisibles." In this case he adopted an artifice, passing to the average velocity in a way that was not well grounded and then using the usual formula for uniform motion. Not only the new mechanics but also the new mathematical analysis found its origins in the discovery of the law of free fall. Since Galileo restricted himself to the case of constant acceleration, the concept of acceleration in general was not needed. The acceleration of free fall as a universal constant does not appear in Galileo's work.

When it comes to the role of force in nonuniform motion, Galileo's statements lack complete clarity. He rejected Aristotle's principle that the velocity is proportional to the acting force, because when there is no force, uniform rectilinear motion is maintained. The law of inertia (Newton's first law) carries Galileo's name. Galileo constantly turned to the example of a projectile that would fly along a line if not for the earth's attraction. He wrote that "the degree of velocity displayed by the body inexorably lies in its own nature, at the same time as the reasons for its acceleration or deceleration are external. . . horizontal motion is eternal, for if it is uniform then it does not weaken for any reason, does not slow down, and is not extinguished." Galileo, in his *Letter to Francesco Ingoli*,[7] poetically describes various phenomena on board a ship, moving uniformly in a straight line, that do not reveal this motion: Drops of water fall exactly into the mouth of a jar placed below, a stone falls straight down from the mast, steam rises straight up, butterflies fly with the same speed in all directions, and so on. We have the sense that Galileo confidently supported the principle

[7] (1578–1649), a priest and professor of civil and canon law who wrote opposing the Copernican system.—*Transl.*

of inertia in "terrestrial" mechanics, but not in celestial mechanics (more about this later).

Newton ascribed to Galileo not only the first law of mechanics, but also the second, although this was an overstatement: Galileo did not make a clear connection between force and acceleration (when they are different from zero). As far as free fall is concerned, Galileo thoroughly answered "how," but not "why."

Motion Along an Inclined Plane

Galileo considered his most fundamental conclusion to be that during consecutive equal time intervals, a falling body travels distances that are proportional to consecutive odd numbers. He wanted to verify this, but how? He could not continue to throw balls from the Tower of Pisa, since he was already living in Padua. In the laboratory, free fall takes place very quickly. But Galileo found a clever way out: He replaced free fall with the slower movement of a body along an inclined plane. He noticed that assuming free fall is uniformly accelerated implies that the movement of a point mass along an inclined plane is also uniformly accelerated. This is essentially today's usual argument of resolving forces, showing that a point mass slides along an inclined plane with constant acceleration $g \sin \alpha$, where α is the angle of inclination to the horizontal and g is the acceleration of free fall. Galileo's reasoning was more awkward: He did not introduce the acceleration of free fall but instead manipulated a large number of proportions, as was then common. He drew a whole series of consequences from the uniform acceleration of a point on an inclined plane that could be verified conveniently in the laboratory (because when the angle is small, it takes a long time for the point to slide down the plane). A key assertion is that if inclined planes have the same height, then the sliding times are related according to the distances traveled (why?).

Motion along an inclined plane was a question of independent interest for Galileo, and he made many observations. For example, if points move along chords AE_i and BF_j of a circle (see Figure 1), where AB is vertical and a diameter, then all sliding times equal the time for free fall along AB (prove it!). A rather complicated argument leads Galileo to the proof that if A, B, and C are consecutive points along a circle, then a point slides faster along the polygonal line ABC than along the chord AC. This is associated with a well-known error of Galileo: He assumed that a point moves most quickly along a quarter of a circle, but this is true instead for an arc of a cycloid.

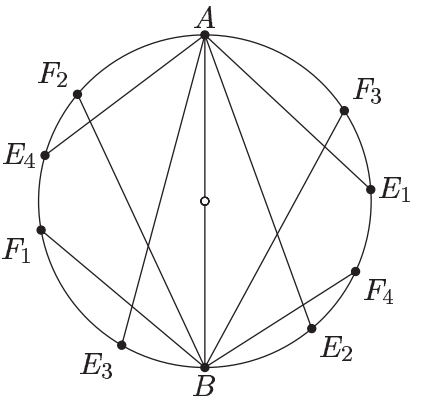


Figure 1.

Projectile Motion

Galileo called the motion of a projectile *forced* (as opposed to free fall). Aristotle assumed that a body thrown at an angle to the horizontal first moves along an inclined line, then along a circular arc, and finally along a vertical line. Tartaglia may have been the first to claim that the trajectory of a projectile "does not have a single part that is perfectly straight."

Galileo constructed a theory of forced motion immediately after his theory of free fall. His route was the same: The theory, a model of the phenomenon, preceded experiments. Galileo's guess was brilliantly simple: The motion of a projectile launched at an angle to the horizontal is made up of the uniform rectilinear motion that would occur if it were not for the force of gravity, together with free fall. In the end, the body moves along a parabola. Let us note that this argument essentially uses the law of inertia—Galileo's law.

In considering complex motion, Galileo had a brilliant predecessor who served as his model: "I assume nothing except the definition of the motion I wish to treat and of whose properties I wish to demonstrate, imitating in this Archimedes in his *Spiral Lines*, where he, having explained what he means by motion made in the spiral that is compounded from two uniform motions, one straight and the other circular, goes on immediately to demonstrate its properties."[8] This refers to the spiral of Archimedes, described by a point moving along the radius of a rotating circle (e.g., a fly walking towards the center of a phonograph record).

[8]From a 1639 letter to Giovanni Battista Baliani, *Galileo at Work*, translated by Stillman Drake, University of Chicago Press, Chicago, 1978, pp. 395–396.—*Transl.*

Using the properties of a parabola, Galileo constructed a "table for gunners, having important practical significance." It was not without reason that Padua belonged to the Venetian republic, and Galileo was in constant contact with the Venetian arsenal. Many of his claims, reached theoretically, could be verified experimentally. He proved Tartaglia's assertion that a shot travels farthest for a 45° angle, and showed that for angles adding up to 90°, shots fired with the same velocity travel the same distance.

Galileo and Kepler

Galileo's discoveries must have startled his contemporaries. The conic sections—ellipses, parabolas, and hyperbolas, the acme of Greek geometry—seemed to be the fruit of mathematical fantasy, unrelated to reality. And here Galileo showed that parabolas inevitably arise in a perfectly "worldly" situation. (Even in the 19th century, Laplace presented an application of the conic sections as a most unexpected use of pure mathematics.) It is remarkable that literally at the same time, conic sections arose in a completely different problem and in a no less surprising way. In 1604–1605 Johannes Kepler (1571–1630) discovered that Mars moves along an ellipse, with the sun at one focus (within ten years Kepler extended this to all the planets). This is an important coincidence and for us the two discoveries go hand in hand, but it is likely that before Newton no one seriously put these results together. Moreover, Galileo did not accept Kepler's law and did not communicate his own discovery to Kepler, in spite of their regular correspondence (only published after Kepler's death).

Galileo and Kepler corresponded for many years. Kepler was one of the scientists closest in spirit to Galileo. Above all it was essential that Kepler unreservedly accepted the Copernican system. As early as 1597, Galileo, in connection with receiving Kepler's book *Mysterium Cosmographicum* (*Cosmographic Mystery*), shared with Kepler his secret desire to publish his arguments supporting the Copernican system. He wrote: "...I do not publish them, because I am deterred by the fate of our teacher Copernicus, who, though he had won immortal fame with a few, was ridiculed and condemned by countless people (for there are so many fools). I would dare to publish my speculations if there were more people like you." In response, Kepler sent a passionate appeal: "Be of good cheer, Galileo, and come out publicly." He proposed joining forces: "If I judge correctly, there are only a few among the distinguished mathematicians of Europe who would part company with us...." And a book can also be printed in Germany, not just in Italy. The problem was seen differently in far off Prague than in Italy,

where for the sixth year Giordano Bruno[9] awaited his fate in prison.

The path Kepler took towards his discovery is very instructive. Kepler had two faces as a scientist. On the one hand, he was a great visionary, attempting to understand the greatest mysteries of the universe. He was certain that the great mystery he discovered was that there are six planets because there are five regular polyhedra! "I have never succeeded in finding words to express my delight at this discovery." Kepler proposed six spheres alternating with the regular polyhedra, so that for each sphere we have one inscribed and one circumscribed polyhedron. He placed the spheres in correspondence with the successive planets. There is a special scheme hidden in the order of the polyhedra (the cube corresponds to Saturn, the tetrahedron to Jupiter, etc.). Kepler compares the ratios of the radii of the spheres with the known relative dimensions of the orbits and, strangely enough, finds only a small discrepancy (except for Mercury). These arguments, published in *Mysterium Cosmographicum*, met with favor from many people and with no objection from Galileo, and Tycho Brahe, the "king of astronomers," invited Kepler to collaborate with him.

Another side of Kepler's scientific life, so different from the other, was associated with this invitation. He scrupulously worked through Tycho's many observations, which were unprecedentedly accurate for observations made without a telescope (their accuracy is estimated at $\pm 25''$). He needed to reexamine the planetary orbits, using Tycho's data. Tycho (whom Kepler called the "phoenix of astronomy") evidently expected a corroboration of his own compromise theory, in which the sun moves around the earth and the remaining planets move around the sun. But Kepler carried out his computations in the Copernican system.

Since Copernicus, like Ptolemy, formed the planetary orbits from circles, his system retained the use of epicycles. Kepler wanted to simplify the system (his summary work, appearing in 1618–1621, was called *Epitome Astronomiae Copernicanae* (*Epitome of the Copernican System*)). Surprisingly, Earth's orbit differs only slightly from a circle, but the sun is somewhat off-center. Copernicus knew all this, but Kepler made the size of this displacement precise. He carefully studied the nonuniformity of the earth's motion in its orbit and for a long time searched for a law to describe it. He tried an inverse proportional dependence on the distance from the sun as well as various other possibilities, but did not yet discover the law of areas (Kepler's second law). Then he computed Mars' orbit and compared it to different curves. Kepler's attitude towards observational data was startlingly sober and confident. Once he rejected a hypothesis because he

[9](1548–1600), a priest, philosopher and astronomer who dreamt of other worlds and who was burned at the stake as a heretic.—*Transl.*

discovered a discrepancy of 8′ from Tycho's data (such a discrepancy is almost invisible to the naked eye). Einstein said, "he first had to recognize that even the most lucidly logical mathematical theory was of itself no guarantee of truth, becoming meaningless unless it was checked against the most exacting observations in natural science."[10] Kepler looked over different sorts of ovals, finally discovering that an ellipse with the sun at one focus fit the data. "Groping incessantly everywhere in the surrounding gloom, I emerged, finally, into the clear light of truth." Is not Kepler's way very different from Galileo's? To a great extent, Galileo proceeded from general principles and qualitative results. In his declining years, Galileo recalled: "I have always esteemed Kepler a free mind (perhaps even too much) and acute, but my way of philosophizing is utterly different from his; it may be that writing on the same matters, and strictly as concerns celestial motions, we have both hit on the same idea, but only in a very few instances, whereby we assigned of some true effect the same true reason; but this will be found to be the case in less than one in a hundred of my thoughts."[11]

Galileo believed that uniform circular motion rules the universe. He believed neither in elliptical orbits nor in the nonuniform motion of the planets in their orbits, and did not take data from observational and computational astronomy into account.

Kepler was the first to consider the mutual attraction of bodies and associate it with motion; he even conjectured on how the interaction decreases with distance (as $\frac{1}{r}$, which is false). He explained the tides by lunar attraction. All this was totally unacceptable to Galileo, who rejected forces acting from afar and, in particular, attempted to explain terrestrial phenomena by the influence of heavenly bodies. This related particularly to the tides, which Galileo erroneously assumed to be an important proof of the earth's motion. Galileo identified such an explanation with astrology, in which events in human life are explained by the influence of the planets. "Among the great people discussing this startling phenomenon of nature, the one who surprises me most is Kepler, who, possessing a free and sharp mind and being quite familiar with the motions ascribed to the Earth, admits a fundamental power of the Moon on water, hidden properties, and such similar childishness." Kepler turned out to be right, but the real arguments appeared later.

[10] Albert Einstein, *Out of My Later Years*, Philosophical Library, New York, 1950, p. 226.

[11] From a 1634 letter to Fulgenzio Micanzio. This translation, together with other comments on Galileo's relation to Kepler, appears in Giorgio de Santillana's translation of *Dialogo... Sopra i Due Massimi Sistemi del Mondo* (*Dialogue on the Two Great World Systems*), University of Chicago Press, Chicago, 1953, pp. 349–350.—*Transl.*

We must keep in mind that Kepler's arguments on mutual attraction are often confused. In one respect, he seriously lagged behind Galileo: He assumed, following Aristotle, that velocity is proportional to force.

Terrestrial and Celestial Mechanics

Around 1610, Galileo obtained results in mechanics towards which he had been working for twenty years. He began work on a comprehensive treatise but unexpected events diverted him from this for more than twenty years! Galileo built a telescope, and at the beginning of 1610 he discovered the moons of Jupiter. For that entire year, astronomical discoveries followed one after the other. Galileo thought he had decisive proof supporting the Copernican system. He completely devoted the following twenty-three years of his life to confirming it, until in 1633 the verdict of the Inquisition interrupted this activity. All those years Galileo thought about mechanics, since this was required by his work *Dialogo... Sopra i Due Massimi Sistemi del Mondo* (*Dialogue on the Two Great World Systems*). At times his new philosophy even contradicted results about "terrestrial" motion. Thus he did not find a place in the universe, "where all parts are in the most excellent order," for rectilinear motion, which under these conditions seemed to him "unnecessary and unnatural." The reason for this was that motion along a line cannot be periodic and the state of the universe must always be changing. He left room for rectilinear motion only in unstable situations, while in nature circular motion must be the rule. Galileo considered the law of inertia he discovered for "local motion" to be valid only near the earth.

Galileo also considered the parabolic law of motion for projectiles to be approximate. He assumed that the trajectory would in fact end at the center of the earth. Because of this he made strange pronouncements about the motion of a projectile having to follow a circular arc or spiral, even after he discovered the trajectory was parabolic. Fermat objected to this, communicated through Carcavi (1637). In reply, Galileo called his statement "poetic fiction" and promised to publish his claim that the trajectory was parabolic, but wrote in conclusion: "But in the motion I defined it is found that all the properties that I demonstrate are verified... in that when we make experiments upon the earth and at heights and distances practicable to us, no sensible difference is discovered, though a sensible, great, and immense difference would be made by approaching to and closely nearing the center."[12] The approximate nature of the parabolic trajectory was

[12]For an English translation of this letter, see Drake, *Galileo at Work*, pp. 376–381.—*Transl.*

clarified by Newton, but Galileo's expectations were incorrect.[13]

The main question about motion that interested Galileo all these years was linked to the standard objection of the opponents of the earth's motion: Why don't objects fly off a moving earth? Galileo had no doubt that the force of gravity was responsible for this, but how could he give a motivated explanation? Let a body move on a sphere of radius R with velocity v. This is how Galileo began his arguments. Let us first fix a starting point. If it were not for the force of gravity the body would continue in rectilinear motion, with velocity v along a tangent. To guarantee motion on the sphere (to hold onto the body), we must add motion in the direction of the center. Galileo was used to combining motions! What was left to do? Just to note that (by the Pythagorean theorem) for the second motion the distance traveled is $s(t) = \sqrt{R^2 + v^2t^2} - R$, and if the time t is small, then this is almost the same as $\tilde{s}(t) = \frac{v^2t^2}{2R}$, since $\frac{s-\tilde{s}}{t^2} \to 0$ as $t \to 0$. Now we cannot fail to recognize Galileo's formula for distance under uniformly accelerated motion, with acceleration $a = \frac{v^2}{R}$. So if $g > a$, then the body will stay on the surface of the sphere. But Galileo did not carry out the second half of this reasoning, giving instead a very confused motivating argument. The formula for centripetal acceleration, along the path Galileo outlined, was obtained by Huygens in 1659.

Discorsi (Two New Sciences)

Finding himself in exile in Siena in 1633, several weeks after the verdict of the Inquisition and his renunciation, Galileo thought about the results he had obtained in mechanics long ago and decided to publish them immediately. He continued work in Arcetri and Florence, without regard to his enforced isolation, worsening health, and progressing blindness. He wrote, "Even though I am silent, I do not live my life completely idly." His book *Discorsi e Dimostrazioni Matematiche, Intorno à Due Nuove Scienze Attenenti alla Mecanica e i Movimenti Locali* (*Discourses and Mathematical Proofs, Concerning Two New Sciences Pertaining to Mechanics and Local Motions*) was completed in 1636, was sent across the border with great precautions (it was not clear how the Inquisition regarded the book), and came out in Holland in July of 1638. As with his previous book, *Massimi Sistemi*, which was

[13]Since Galileo delayed publication for a long time, the first mention of the parabolic trajectory appeared in 1632 in Cavalieri's *Lo Specchio Ustorio* (*The Burning Glass*), which very clearly took from Galileo the idea of adding rectilinear motions and the principle of inertia. Galileo was offended by the absence of the obligatory references, and spoke of his discovery that the trajectory is parabolic as the main goal of forty years' work. Cavalieri's apologies quickly satisfied Galileo.

the reason for his persecution, Galileo wrote *Discorsi* in dialogue form.[14] The dialogues take place over the course of six days and involve the same heroes: Salviati (taking the author's point of view), Sagredo, and Simplicio (who is on Aristotle's side; his name means "simpleton"). On the third and fourth days they read an academic treatise on motion (by Galileo), *De Motu Locali* (*On Local Motion*), and discuss it in detail. Incidentally, in the title of the book "mechanics" and "motion" are separate, since at that time mechanics meant only statics and strength of materials. The form chosen by Galileo for the discussion allowed many to realize how he approached his own discoveries.

The aged Galileo strove to realize the ideas he had laid aside for so long. But much was already beyond his power, and he needed assistants. He entrusted his son Vincenzio with constructing a clock based on the discoveries in his youth about the pendulum, but he did not see his ideas brought to fruition. The Inquisition had limited Galileo's contacts with the outside world. After the completion of *Discorsi* at the villa at Arcetri, which Galileo called his prison, much-wanted guests began to appear. These were his old friend and trusted student Benedetto Castelli, and Cavalieri. Also, Viviani and Torricelli had not abandoned their teacher, but for some time had been helping him complete his work and continue his research.

Torricelli computed the velocity vector for a projectile launched at an angle, using the addition rule for velocities, and since the velocity is directed along a tangent he obtained an elegant way to extend the tangent to a parabola. The era of differential and integral calculus had arrived and problems about extending tangents to curves came into the foreground of mathematics. Various methods were worked out to solve them. One became the kinematic method, in which the curve was represented as the trajectory of a complex motion and the tangent was found by adding velocities, as Torricelli first did for parabolas. The French mathematician Gilles Persone de Roberval (1602–1675) worked miracles with this method. "Mechanical" curves, obtained as the trajectories of various motions, firmly came into use in mathematical analysis. It is worth recalling that Galileo consciously limited himself to motions that arose in nature: "For anyone may invent an arbitrary type of motion and discuss its properties; thus, for instance, some have imagined helices and conchoids as described by certain motions which are not met with in nature, and have very commendably established the properties which these curves possess... but we have decided to consider the phenomena of bodies ... such as actually occurs

[14]There is some confusion over the English titles of these books since both have been translated as *Dialogues...*, as noted in earlier references. To avoid confusion, we will use the abbreviated Italian titles.—*Transl.*

in nature."[15] The value of a general view of motion was demonstrated by Newton.

Discorsi determined the development of mechanics for a long time. It was on the desks of both Huygens and Newton, Galileo's great successors. It is difficult to imagine how much the development of mechanics would have been delayed if these unfortunate events had not occurred and if Galileo had not written down his great discoveries.

Mathematical Appendix

The history of the discovery of the law of free fall has another side: This is the history not only of a discovery that was carried through but also of one that was neglected. Once Galileo understood that motion could not take place according to $v(t) = cs(t)$, he lost interest in this law. He was only interested in motions that occur in nature! Soon the Scottish Lord John Napier (1550–1617) became interested in motion that follows an analogous law.

Napier considered rectilinear motion taking place according to the law $v(t) = l(t)$, where $v(t)$ is the instantaneous velocity at time t and $l(t)$ is not the distance traveled, but rather the distance between the moving point at time t and a fixed point O on the line. The case Galileo considered corresponds to the case when the moving point is at O at the initial time $t = 0$, i.e., $l(0) = 0$, $l(t) = s(t)$. For Napier, $l(0) > 0$, and $l(t) = l(0) + s(t)$.

It turns out that for $l(0) > 0$, such motion can occur in principle and possesses remarkable mathematical properties (although it does not "occur in nature"). Let us examine this. First of all, if the initial distance $l(0)$ is multiplied by c, then at each moment of time $l(t)$ and $v(t)$ are also multiplied by c. Strictly speaking, this requires proof! But it is clear that after we multiply l and v by a constant, the law $v(t) = l(t)$ still holds. Next, we restrict ourselves to the case $l(0) = 1$. Then, as we will show,

$$l(t_1 + t_2) = l(t_1)l(t_2).$$

It is convenient to take t_1 as a new time origin. By what we said above, at the new time t_2 (the old $t_1 + t_2$) the distance to O should be $l(t_1)$ times greater than at the old t_2. But this means that $l(t_1 + t_2) = l(t_1)l(t_2)$. This is how the exponential function first appeared in science.

We have $l(t) = e^t$, where $e = l(1)$, i.e., the distance from 0 at $t = 1$. Using this and $v = l$, it is not difficult to show that $e > 2$ (prove it!). In fact, $e = 2.71828\ldots$ and became known as Napier's number. By considering motion

[15] *Discorsi*, p. 160.

according to the law $v(t) = kl(t)$, we can obtain exponential functions with other bases.

For any positive number a, we will call the time t for which $l(t) = a$ the (*natural*) *logarithm* of a (denoted by $\ln a$).[16] By the above, $\ln ab = \ln a + \ln b$. For twenty years Napier composed tables of logarithms, and in 1614 published *Mirifici Logarithmorum Canonis Descriptio* (*A Description of the Wonderful Law of Logarithms*), whose forward contained an apology for the inevitable errors and finished with the words, "Nothing is perfect at first."

Napier's discovery is remarkable not only because he constructed logarithm tables but also because he showed that new functions can appear through the study of motion. Beginning with the work of Galileo and Napier, mechanics became a constant source of new functions and curves for mathematics.

II The Medicean Stars

In November 1979, the Vatican decided to rehabilitate Galileo, who had been sentenced by the Inquisition in 1633. Then he had been acknowledged to be "vehemently suspect of heresy" because "he held and defended as probable" the opinion that "the Sun is the center of the world and does not move from east to west, and that the Earth moves and is not the center of the world."[17] At a session of the Vatican Academy of Sciences in November 1979, devoted to Einstein's centenary, Pope John Paul II noted that Galileo "suffered greatly—we cannot conceal this now—from oppression on the part of the Church," but qualifying Galileo's repentance as "divine illumination in the mind of the scientist," he asserted that Galileo's tragedy emphasizes the "harmony of faith and knowledge, religion and science." In October 1980, it was announced that the pope had ordered a supplementary investigation into the circumstances of the proceedings against Galileo. Conversations about Galileo's acquittal had already taken place at the second Vatican Council (1962–1965). They had wanted to time his acquittal to the 400th anniversary, in 1964, of his birth, but evidently did not succeed, since the question turned out to be not uncontroversial. Incidentally, Galileo's works (together with those of Copernicus and Kepler) had been taken off the *Index Librorum Prohibitorum* (*Index of Forbidden Books*)

[16] Napier's reasoning was not quite the same as this and Napier's logarithms are different from natural logarithms.

[17] Complete English translations of Galileo's sentence and renunciation appears in de Santillana, *The Crime of Galileo*, University of Chicago Press, Chicago, 1955, p. 306 ff.—*Transl.*

as early as 1835. The judgment against Galileo and his renunciation have continued to disturb people, often far from science, for three and a half centuries. The attention given by literature to this question is typical (it suffices to recall Berthold Brecht's play *Galileo*). The problem of Galileo is alive even today, notwithstanding his recent "rehabilitation."

At the dawn of the 17th century, the question of a world system was not a simple one. In the 4th century B.C., Aristotle claimed that the seven visible heavenly bodies revolve uniformly around the earth, on crystal spheres to which they are attached, and that the fixed stars occupy an eighth sphere. Astrologers classified the planets as follows: two luminaries, the moon and sun; two harmful planets, Mars and Saturn; two favorable ones, Jupiter and Venus; and one neutral one, Mercury.

Aristotle's rules, and especially those of his followers, did not explain any deviations from his scheme, e.g., the surprising "retrograde" motion, in which a planet appears to reverse direction. Contradictions to precise observational data gradually accumulated. In the 2nd century A.D., Ptolemy constructed a system that took observations into account as much as possible. He assumed that the planets move along auxiliary curves (epicycles) whose centers (deferents) in turn revolve around the earth. The desire to account for new data led to an increasingly complicated system. We must acknowledge the persistence and cleverness by which scientists succeeded in saving this system.

Nicolaus Copernicus (1473–1543) proposed a completely unexpected route. His carefully worked-out scheme, consistent with observations, contained all the fundamental aspects of today's view of the solar system: The planets, including the earth, revolve around the sun; the earth executes a daily motion; and the moon revolves around the earth. Such an approach simplified everything incredibly, although obscure aspects remained as a result of making the system agree with observational data. In the opinion of Copernicus, the planets move with nearly uniform circular motion (as for Aristotle) but undoubtedly deviate from it. Epicycles were still required, but they played a less essential role than for Ptolemy. The epicycles disappeared only with Kepler, who discovered that the orbits are elliptical. The Copernican system was not a purely descriptive theory based on qualitative phenomena. It contained many computations, e.g., the distance to the sun, the periods of revolution, etc. Only such a theory could compete with Ptolemy's, which completely accounted for the data.

The Pythagoreans had already indicated the possibility of the earth's motion. Therefore, the Church called the teaching that the earth moves *Pythagorean*. They preferred not to use Copernicus' name for this teaching, for the following reason. His book *De Revolutionibus Orbium Coelestium*

(*On the Revolutions of the Heavenly Spheres*), which appeared the year he died, was preceded by a foreword that Copernicus himself may not have written. The foreword calls his system a convenient mathematical scheme for astronomical computations and no more, and the motions it considers are called imaginary. This means that "real" motions are not discussed in the book. That is not a mathematician's job! This question must be solved by philosophers and theologians, in accordance with holy writ. The book was dedicated to Pope Paul III. The Church had arranged this compromise, and the book was not declared heretical. Mathematicians were allowed to use imaginary schemes in their computations. This included even the Jesuit astronomers, who in particular used Copernicus' tables for the calculations that were necessary to reform the calendar.

The assertion that the earth is stationary and the sun moves had to remain immutable. The Church was not as uncompromising regarding the remaining planets (they are not mentioned in the Scriptures). Tycho's system, in which the sun moves around the earth but the remaining planets move around the sun, was tolerated. Tycho essentially gave up the crystal spheres, asserting that comets do not belong to the "world beneath the moon" but fly in from the outside (Galileo, by the way, took a different view).

Thus the Copernican system is a convenient mathematical fiction but a Pythagorean teaching, and is heresy. Here lay the boundary. Galileo was not ready to agree to this compromise: "Copernicus, in my view, should not soften, for the motion of the Earth and the immobility of the Sun form the essence and general foundation of his teaching. Therefore they must either condemn it completely or else leave it as it is!" Galileo insisted that the earth's motion is not imaginary, but real.

Galileo's path leading to the decisive struggle for a heliocentric world system was not a simple one. He believed in the Copernican system early on, but for a long time did not decide to publish his arguments in its support (his 1597 letter to Kepler tells us this). The 17th century began with the burning of Giordano Bruno. By 1610, Galileo had approached the peak of his scientific activity: He completed his twenty-year-long study of natural motion (free fall and projectile motion) brilliantly. He was beginning work on his great discoveries, and unexpectedly left them for an indefinite period. What happened? Events took place in Galileo's scientific life that forced this completely practical man to put off publishing the discoveries to which he had devoted his youth. Galileo decided that he now had decisive arguments to support the Copernican system and that from then on his life would be wholly given over to propagandizing these ideas. Let us recall these important arguments.

New Glasses

In talking about the lives of great scientists, we rarely pay attention to everyday matters. One of the reasons for Galileo's move from Pisa to Padua was to increase his salary. His material situation became more secure. His initial salary of 180 florins increased, although slowly; he received additional income from the young aristocrats with whom he worked privately and who often lived in his home. But paying his sisters' dowries was a heavy burden that lay on Galileo's shoulders, even as his own family was growing and required greater resources. In 1609 Galileo was anxious about the drawn-out negotiations over increasing his salary. Some sort of invention, with an unquestioned practical application, would open the purse strings of the stingy and practical Venetian *signori*. Galileo was no stranger to technical problems. His home had an excellent workshop, he had recently designed convenient proportional dividers (a "geometric and military compass"), and had himself seen to their manufacture and distribution. He could have thought of artillery tables, based on the parabolic trajectory of a projectile in flight. But he unexpectedly discovered a completely different idea.

In 1608, optical tubes, sometimes called "new glasses," appeared in Holland that allowed people to look at distant objects. Leonardo da Vinci had spoken of glasses through which one could see the moon enlarged, and Roger Bacon of glasses that made a man as big as a mountain. Eyeglass makers Hans Lippershey and Zacharias Janssen contended for the honor of having invented them. Towards the beginning of 1609, such a tube could be bought in Holland for a few *soldi*. By midyear, the tubes had appeared in Paris. Henri IV spoke pessimistically about the innovation, explaining that at the moment he was more in need of glasses that magnified nearby objects rather than distant ones. Then some foreigner tried to sell an optical tube to the Republic of Venice, without going into details about its origins. The first tubes were very imperfect, and Paolo Sarpi, Galileo's friend, expressed a negative opinion about being able to use them "in war, on land, and at sea." Galileo heard about the tubes when he was in Venice.

"Upon hearing this news I returned to Padua, where I then resided, and set myself to thinking about the problem. The first night after my return I solved it, and on the following day I constructed the instrument and sent word of this to those same friends at Venice with whom I had discussed the matter the previous day. Immediately afterwards I applied myself to the construction of another and better one, which six days later I took to Venice. . . ."[18] Elsewhere he describes the situation more triumphantly:

[18] A translation of Galileo's accounts of his discovery appears in Stillman Drake, *Discoveries*

"...sparing neither labor nor expense, I succeeded in constructing for myself so excellent an instrument that objects seen by means of it appeared nearly one thousand times larger and over thirty times closer than when regarded with our natural vision. It would be superfluous to enumerate the number and importance of the advantages of such an instrument at sea as well as on land."[19]

In fact, the properties of the tubes were more modest. Galileo's first tube magnified objects three times (3×), while the one taken to Venice was 8×. Galileo decided, with his perfected tube, to push his request to the members of the *Signoria* (this may have been Sarpi's idea). On August 21st, the most respected people in Venice observed the far quarters of the city from the *campanile* of St. Mark's Cathedral, and on August 24th Galileo triumphantly gave his tube to Leonardo Donato, Doge of Venice. Galileo did not skimp in publicizing his gift. He said that he extracted his idea "from the most secret considerations of perspective."

Many later said that Galileo overestimated his contribution or even appropriated a foreign invention as his own (Brecht's play talks about this). At least in his publications, Galileo always acknowledged that he constructed his tube having heard of the Dutch invention (but without detailed information and not having seen "the Flemish glass"). Later he stressed the originality of his approach: "Indeed, we know that the Fleming who was first to invent the telescope was a simple maker of ordinary spectacles who, casually handling lenses of various sorts, happened to look through two at once, one convex and the other concave, and placed at different distances from the eye. In this way he observed the resulting effect and thus discovered the instrument. But I, incited by the news mentioned above, discovered the same thing by means of reasoning." The name "telescope" was proposed by Cesi (see below) in 1611, when Galileo demonstrated the tube in Rome; earlier Galileo had used the term *occhiale*.[20] One can assume that Galileo demonstrated the superiority of theory over practice: For many years no one was able to build tubes of comparable power. (Because of this, in particular, there was no confirmation of Galileo's astronomical observations.)

Galileo's tube fulfilled its purpose: He was given an annual salary for life of a thousand florins, unheard of for a mathematician. Galileo was supposed to make twelve tubes for the *Signoria*, and to give none to anyone else.

and Opinions of Galileo, Doubleday, New York, 1957 (copyright 1957 by Stillman Drake). This quotation (p. 244) is taken from *Il Saggiatore* (*The Assayer*).—*Transl.*

[19]Ibid., pp. 21–52 passim; taken from *Sidereus Nuncius* (*The Starry Messenger*). Reprinted with permission.—*Transl.*

[20]Spyglass, or eyeglass.—*Transl.*

The Starry Messenger

Soon Galileo had a 20× tube and wrote, "forsaking terrestrial observations, I turned to celestial ones." At the end of 1609 Galileo looked through his tube at the moon and discovered "that the surface of the moon is not smooth, uniform, and precisely spherical as a great number of philosophers believe it (and the other heavenly bodies) to be, but is uneven, rough, and full of cavities and prominences, being not unlike the face of the earth. . . ." Moreover, Galileo turned his attention to the ashen light on the part of the moon not lit by the sun. He assumed this light to be "the reflection of the Earth." It turned out later that at the same time an Englishman, Thomas Harriot (1560–1621), and his student William Lower (c.1570–1615) had begun to observe heavenly bodies by telescope (their observations were unknown to their contemporaries). Lower wrote in a letter to his teacher that the moon reminded him of a tart with ham that his cook had baked the previous week. Leonardo da Vinci and Michael Mästlin (1580–1635), Kepler's teacher, had already spoken of the moon's ashen light.

Then, before Galileo's eyes, the Milky Way broke into separate stars: "All the disputes which have vexed philosophers through so many ages have been resolved... the galaxy is, in fact, nothing but a congeries of innumerable stars grouped together in clusters."

Finally, on January 7, 1610, Galileo turned his telescope towards Jupiter. Near Jupiter he discovered three stars. He did not doubt that he was seeing ordinary "fixed" stars, but something greatly attracted his attention. The following night, "unknowingly led by some sort of fate," he again looked at Jupiter. He had no reason to regret it! Again he saw the familiar stars, but... their position with respect to Jupiter had changed: Yesterday they were found on different sides of Jupiter but today all were on the same side. He could still assume the stars were fixed and explain the change in relative location by Jupiter's motion. On January 9th "the sky was then covered by clouds everywhere." On January 10th and 11th he found only two of the three stars, but on the 13th, to the contrary, a fourth appeared.

Galileo saw a new solution: The stars he was observing moved in relation to Jupiter, they were its satellites—moons—and they disappear because they are eclipsed by Jupiter. By the end of the month he was sure, "passing from the sensation of enigma to the feeling of rapture." He wrote to Belisario Vinta, the Florentine Secretary of State: "But the greatest miracle of all is that I discovered four new planets and observed their own distinctive motions, and the differences in their motions relative to one another, and relative to the motions of the other stars. These new planets move around another big star in the same way Venus and Mercury and, possibly,

other known planets move around the Sun." There is no doubt in what context Galileo viewed his discovery, but see how careful a formulation he still used in regard to "other known planets."

Until March 2nd, Galileo observed Jupiter's moons, taking advantage of each cloudless night, and as early as March 12th his famous *Sidereus Nuncius* (*The Starry Messenger*) appeared: "THE STARRY MESSENGER, Revealing great, unusual, and remarkable spectacles, opening these to the consideration of every man, and especially of philosophers and astronomers: AS OBSERVED BY GALILEO GALILEI, Gentleman of Florence, Professor of Mathematics in the University of Padua, WITH THE AID OF A SPY-GLASS *lately invented by him*, In the surface of the Moon, in innumerable Fixed Stars, in Nebulae, and above all in FOUR PLANETS swiftly revolving about Jupiter at differing distances and periods. . . ."

Everyday matters were superimposed on all this. It turned out that Galileo's salary would be increased only after a year and, moreover, his teaching duties had begun to weigh on him. He began to think about moving to Florence. The Grand Duke, Ferdinando I de' Medici had just died and Cosimo II, Galileo's former student, had ascended to the throne. The duke's patronage could be unmatchable for solving many problems, especially in the difficult matter of defending the Copernican system. There was already no doubt that this would be Galileo's main work. He wrote in a letter to Vinta, in connection with the possible move: "The works which I must bring to conclusion are these. Two books on the system and constitution of the universe—an immense conception full of philosophy, astronomy, and geometry."[21]

Soon Galileo proposed, through Vinta, to name the moons of Jupiter the Cosimean or the Medicean stars, in honor of Cosimo de' Medici. The second form was chosen. The number of moons fortunately coincided with Cosimo's having three brothers. *Sidereus Nuncius* is dedicated to Cosimo de' Medici: "And so, most serene Cosimo, having discovered under your patronage these stars unknown to every astronomer before me, I have with good right decided to designate them by the august name of your family. And if I am first to have investigated them, who can justly blame me if I likewise name them, calling them the Medicean Stars, in the hope that this name will bring as much honor to them as the names of other heroes have bestowed on other stars?"[22] Later, all four satellites received their own names (Io, Europa, Ganymede, and Callisto), and in order to distinguish them from the moons of Jupiter that were discovered later, they were called Galilean.

[21] For a translation of this letter, see Drake, *Discoveries*, pp. 60–65.—*Transl.*
[22] Ibid., p. 25.

Galileo set off for Florence for the Easter vacation. He took a tube with him so that the duke could see "his" stars for himself. Galileo was showered with respect, a medal with the image of the Medicean stars was to be struck in his honor, the conditions for his move were roughly set, and only the title of Galileo's position remained to be specified. The ruler was pleased to have his name immortalized in the heavens; no other royal personage could boast of that. On May 14th, Galileo received a letter from France dated April 20th, in which he was asked "to discover as soon as possible some heavenly body to which His Majesty's name may be fitly attached," meaning Henri IV. It was specified that the star was to be called "rather by the name Henri than Bourbon." It turned out that the author of the letter rushed Galileo for nothing: As soon as he sent it, "the sovereign attended by fortune" was assassinated. Later Galileo wrote to Florence that the House of Medici was in an exclusive situation: Neither Mars nor Saturn turned out to have moons (about fifty years later Huygens and Cassini discovered moons of Saturn, and then moons were also discovered around Mars).

Doubts plagued the Grand Duke. Rumors stubbornly spread that the stars given him were the fruit of Galileo's fantasies or the creation of his tube. Even Christophorus Clavius (1538–1612), the leading mathematician of the College of Rome, said it. The situation was made more complicated because no astronomer other than Galileo himself had seen the Medicean stars. Galileo paid for the fact that no one else had made such perfect tubes. Such an important discovery had to be verified by the three most famous astronomers: Kepler, Giovanni Magini (1555–1617), and Clavius. And for the time being the question of moving to Florence was put aside.

Kepler, Magini, and Clavius

Things were simplest with Magini. On the way from Florence to Padua, Galileo stopped at Bologna and showed Magini the stars he had discovered. Magini, equally famous for his computational abilities and his guile, gave the impression that he could see nothing around Jupiter but had certainly been forewarned. He did not argue and was prepared to attribute it to his poor vision, but this could not comfort Galileo.

Kepler responded immediately to the report of the discovery. As early as April 19th, he wrote an enthusiastic letter to Galileo.[23] It turned out that news about the new planets had already come to Germany in mid-March.

[23]This letter was published as *Dissertatio cum Nuncio Sidereo* (*Conversation with the Starry Messenger*). The following quotations are taken from the English translation by Edward Rosen, *Kepler's Conversation with Galileo's Sidereal Messenger*, Johnson Reprint Corporation, New York, 1965, pp. 9–39 passim.—*Transl.*

Kepler gently scolded Galileo about the lack of an answer to his *Astronomia Nova... de Motibus Stellae Martis* (*A New Astronomy... Commentaries on the Motions of Mars*), which contained his first two laws and which he had recently sent him: "Instead of reading a book by someone else, [my Galileo] has busied himself with a highly startling revelation... of four previously unknown planets, discovered by the use of the telescope with two lenses."

The first reports were unclear, Kepler was afraid that Galileo had discovered new (more than six) planets in the solar system. Kepler strongly held to the opinion that there were exactly six planets, and that the number six was not accidental but related to the five regular polyhedra. Kepler's fantasy led to still another possibility: All the planets similar to the earth have one moon in common, and this is what Galileo must have discovered. "The earth, which is one of the planets (according to Copernicus), has its own moon revolving around it as a special case. In the same way, Galileo has quite possibly seen four other very tiny moons running in very narrow orbits around the small bodies of Saturn, Jupiter, Mars, and Venus. But Mercury, the last of the planets around the sun, is so deeply immersed in the sun's rays that Galileo has not yet been able to discern anything similar there." Kepler sought numerical laws everywhere! Then he thought about the fact that one can speak of planets revolving around "fixed stars," rather than the sun. He recalled Bruno's innumerable worlds and even thought about the possibility "that countless others will be hereafter discovered in the same region, now that this start has been made."

At the same time, Emperor Rudolf II received *Sidereus Nuncius* (Kepler was the Imperial Astronomer). Kepler unhesitatingly believed Galileo's report: "I may perhaps seem rash in accepting your claim so readily with no support from my own experience. But why should I not believe a most learned mathematician, whose very style attests the soundness of his judgment? He has no intention of practicing deception in a bid for vulgar publicity, nor does he pretend to have seen what he has not seen. Because he loves the truth, he does not hesitate to oppose even the most familiar opinions...."

It seems that the regularity of the distribution of the number of planetary moons bothered Kepler: "I should rather wish that I now had a telescope at hand, with which I might anticipate you in discovering two satellites of Mars (as the relationship seems to me to require) and six or eight satellites of Saturn, with one each perhaps for Venus and Mercury." Kepler did not know if he had an arithmetic or a geometric progression!

Kepler pointed out some of Galileo's predecessors to him (Mästlin spoke of the ashen light of the moon and Giambattista Della Porta (1535–1615) predicted the possibility of constructing an optical tube). Kepler hoped

that the sun is brighter than the fixed stars, and wanted to believe in the uniqueness of our world: "This world of ours does not belong to an undifferentiated swarm of countless others." There is no limit to Kepler's fantasies: "It is not improbable, I must point out, that there are inhabitants not only on the moon but on Jupiter too.... Given ships or sails adapted to the breezes of heaven, there will be those who will not shrink from even that vast expanse."

Magini tried to draw Kepler to his side. Kepler was implacable: "We are both Copernicans—each is happy with his own kind." Critical remarks from Kepler's *Dissertatio* reassured Magini: "Now there remain only these four new servants of Jupiter to banish and destroy." Martin Horky, an astronomer in Magini's circle, initiated a series of pamphlets against Galileo in May 1610. In *Excursion Against the Starry Messenger*, he explained the moons of Jupiter as an optical illusion. Kepler's relations with Horky were equivocal. In a letter to Galileo he called the essay cheeky, and said he was "surprised at the impudence of this youth." To Horky himself, expressing surprise at his continuing doubts about "Galileo's stars," Kepler wrote: "...not surprised and do not accuse you; the opinions of those who philosophize must be free."

The absence of confirmation began to bother Kepler. He himself had no suitable tube. From Bologna came the university's conclusion (at Magini's instigation) that the stars could not be seen through Galileo's own tube. In August, the worried Kepler wrote to Galileo: "I cannot conceal from you that letters are arriving in Prague from many Italians that these planets cannot be seen with your optical tube... therefore I beg of you, Galileo, to send some witnesses as soon as possible.... The entire proof of the truth of the observations lies on you alone." Happily, the emperor, Rudolf II, known not only for his caprice but also for his love of science, became passionate about the tubes. Finally, a sufficient well-perfected tube appeared in Prague, and in September, Kepler saw the moons of Jupiter. The participants in the observation independently drew the positions of the stars and the drawings coincided. "Galileo, you have won!" exclaimed Kepler.

In September, Antonio Santini saw Jupiter's moons from Venice, and especially joyful news came in December: Clavius had seen the moons. True, he still was not "sure whether or not they are planets." In September, Galileo moved to Florence. He began a correspondence with Clavius (in the Republic of Venice, it had been forbidden to correspond with Jesuits). "In truth, you, your Excellency, deserve great praise since you are the first who has observed them," Clavius wrote to Galileo. Galileo found the way to Magini's heart. He recommended his work on burning lenses to the Grand Duke and enabled him to obtain the vacated chair at Padua. (Magini had

sought this position before, when Galileo moved to Padua from Pisa.) The careful Magini gave a positive opinion on Santini's testimony. One could not ask for anything more!

A Year of Great Discoveries

The year 1610, beginning with the discovery of the moons of Jupiter, was an unusually happy one for Galileo as an astronomer: He made almost all his remarkable astronomical observations in that one year. On July 25th, he again observed "Jupiter in the morning in the east, together with its retinue." After this he discovered "still another most unusual miracle." He communicated his discovery to Florence, asking that it be kept secret until he could publish it: "The planet Saturn is not merely one single star, but three stars very close together, so much so that they are all but in contact one with another. They are quite immovable with regard to each other... the middle star of the three is by far greater than the two on either side." Galileo sent a phrase to Kepler, encoded in the form of an anagram: "I have observed the most distant of the planets to have a triple form." Later, Galileo wrote: "So you see a guard of satellites has been found for Jupiter, and for the decrepit little old man [Saturn] two servants to help his steps and never leave his side."

Galileo did not reveal his secret for five months. Kepler and Rudolf had no patience to figure out the clue and made the most improbable conjectures: "As quickly as possible, satisfy our passionate desire to know what your new discovery consists of. There is no man whom you could fear as a rival." Galileo divulged the secret, adding that in the weakest tube Saturn looked like an olive. This is how Galileo's discovery first appeared in print, in the preface to Kepler's *Dioptice* (*Dioptics*),[24] with the obligatory references.

Within two years, Saturn unexpectedly stopped appearing as a triple. Galileo attributed this to Saturn's motion around the sun and predicted that it could soon be seen again in the form of three stars. The prediction came true, but Galileo did not guess Saturn's secret. The secret was unveiled in 1655, when Huygens, looking at Saturn through a 92× telescope, discovered that Saturn is surrounded by a ring that at lower magnification appears as adjacent stars. The ring becomes invisible when the observer happens to be in its plane, and Galileo was lucky to see this rare phenomenon. The visual impression of Saturn evolved as telescopes became stronger, from an

[24]The branch of geometrical optics dealing with lenses and images. Kepler's remarks are also contained in the translation of *Sidereus Nuncius* by E. S. Carlos, *The Sidereal Messenger of Galileo Galilei*, Rivingtons, London, 1880, pp. 90–91.—*Transl.*

olive to a sphere surrounded by a ring. Huygens also discovered Saturn's largest moon—Titan.

Soon after Galileo sent Kepler his letter with the anagrammatic clue, there was news of still other planets. For a long time, Galileo had been intently observing Venus both as a morning and as an evening star. There was a great deal of arguing over Venus and Mercury between the adherents of Ptolemy and Copernicus. The former could not agree where their "spheres" were—within the sun's "sphere" or outside it. For Copernicus' adherents it was clear that if these planets are dark bodies then, since they lie between the sun and the earth, they must at times be seen as partial discs (like the phases of the moon). This problem does not arise if we assume that the planets shine by their own light (Kepler's position, apparently) or that they are transparent (this possibility was seriously discussed). Perhaps the telescope could help see what had not been seen with the naked eye.

Castelli recalled this problem in a letter to Galileo on December 5, 1610: "Since (so I believe) Copernicus' position is correct that Venus revolves around the Sun, we clearly had to observe it sometimes with horns and sometimes not. . . , if, however, the small size of the horns and the emission of light do not prevent us from observing these differences." But Galileo hardly needed to be reminded of this. As early as December 10th, he sent a coded message to Kepler in Prague, via the Tuscan ambassador Giuliano de' Medici, about his discovery of the phases of Venus, together with an accompanying letter: "I am sending you a coded message about yet another of my unusual observations, which leads to the solution of the most important disputes in astronomy and which contains the most important argument in support of the Pythagorean and Copernican system." Kepler, as always, had no patience to figure out the clue: "You will see that you're dealing with a German among Germans!"

But the first to whom Galileo revealed his secret was Clavius. Galileo had just heard from Clavius that the astronomers of the *Collegio Romano* had also observed both the moons of Jupiter and the elongated form of Saturn. The support of the *Collegio* played a special role in Galileo's plans, and he hurried to surprise Clavius with his new discovery. Galileo described his observations of Venus after "its evening appearance," and talked about how its circular form unexpectedly became twisted to one side and turned towards the sun, so that Venus no longer looked like a semicircle but "became noticeably horned." He predicted the form Venus would take when it would be seen as a morning star, and concluded, "This is how, my lord, it is explained how Venus (and undoubtedly Mercury does the same) moves around the Sun, which without any doubt is the center of the greatest revolutions of all the planets. Moreover, we are sure that these planets by

themselves are dark and shine only when illuminated by the Sun, which, as I think, does not travel with the fixed stars according to certain of my observations. . . ." Clavius could have no doubt about where Galileo was heading! This brought Galileo's year of great astronomical discoveries to a close.

Galileo did not halt his astronomical observations, but they were basically a continuation of what he saw in 1610. He continued to observe the sunspots that had appeared in the summer of 1610, and by 1613 had discovered that the sun rotates on its axis; we have already spoken about how he saw the disappearance of Saturn's "appendages." At the end of his life, before he was finally blind, Galileo was fortunate to discover the appearance of the libration of the moon (as a result of which it became possible to observe more than half the moon's surface). But he would never again be able to devote as much time to perfecting the telescope or to astronomical observations. And the mysteries of the universe would not again reveal themselves to him as they had in that great year! Galileo's achievements were so great that it would be at least half a century before comparable discoveries would be made in observational astronomy (by Huygens and Cassini). Now other problems began to disturb Galileo, and to solve these problems it was important for him to go to Rome.

Subjugation by Rome

Galileo arrived in Rome on March 29, 1611. He enjoyed the special attention of the Grand Duke of Tuscany, arriving in the Duke's sedan-chair and staying at the Medici palace in Rome. The four astronomers of the *Collegio Romano*, Clavius, Christoph Grienberger, Odo van Maelcote, and Paolo Lembo, received him warmly. Galileo discovered that the Jesuit fathers systematically observed the Medici stars through optical tubes, trying to determine their periods. On April 21st, one of the leaders of the Holy Office, Cardinal Roberto Bellarmino, sent them an official query "about the new celestial observations of a leading mathematician" (his name was not mentioned) regarding the Milky Way, Saturn, the moon, and the moons of Jupiter. The answer came on April 24th, essentially confirming the observations. They noted minor discrepancies in the observations (the stars forming Saturn did not seem to them to be separate), and substantive ones in the interpretation of what was seen on the moon (not mountains but rather that the density of the "lunar body" was not uniform).

On April 14th, Galileo became the fifth member of the *Accademia del Lincei* ("lynx-eyed"),[25] founded eight years earlier by Federico Cesi, Mar-

[25]Referring to Lynceus the Argonaut, noted for his keen sight.—*Transl.*

quis of Montebello. The academy was devoted to the free study of nature, with no limitations. Cesi later wrote to Galileo, "Those whom we accept will not be slaves of Aristotle or of any other philosopher, but will be persons of noble and free thought in the study of nature." His friendship with Cesi played an important role in Galileo's later life, and he now put *Galileo Galilei Linceo* on his works. A demonstration of Galileo's surprising tube took place atop the Janiculum hill (that was then Cesi proposed to call it a telescope).

Galileo was also honored by the *Collegio Romano*. Odo van Maelcote read a report entitled, "The Starry Messenger of the Roman College." He called Galileo "the most remarkable and fortunate astronomer currently alive" and praised his discoveries, but gently said that Galileo's explanations of the phenomena he discovered are not the only ones possible. Galileo was given to understand the limits within which he must stay. This wish was expressed very precisely by Paolo Gualdo: "...you should be satisfied with the glory you have acquired thanks to observations of the moon, the four planets, and similar things, and not take on the defense of ideas so contrary to human reason...." Gualdo's advice also foreshadowed the path that Galileo would later take: "... many things can be uttered by way of disputation which it is not wise to affirm as truths, in particular, if against them you have general opinion, absorbed, if one can say, with the creation of the universe." Cardinal Bellarmino apparently also indicated to Galileo, during an audience, the limits of what was permitted. Bellarmino gave a more definite warning to the Tuscan ambassador Francesco Niccolini: "Galileo should stay within the indicated bounds, or else his works will be given to the theological experts for consideration," and the ambassador was to understand that nothing good could come of that.

The rest of Galileo's trip was successful. Cardinal Francesco del Monte wrote to the Grand Duke, "Galileo, during the days he was in Rome, gave much satisfaction and, I think, obtained much, for he was able to demonstrate his discoveries so well that all the worthy and leading people of this city recognized them not only as true and real, but also as staggering. If we now lived in the ancient Roman republic, I am convinced that they would erect a statue to him on the Capitoline, in order to pay respect to his exclusive prowess."

"Philosopher and First Mathematician to the Grand Duke"

Thus not even a year had passed before Galileo's surprising astronomical discoveries obtained recognition. Do no think, however, that the *Collegio Romano* stopped the accusations against him. As before, there were peo-

ple who opposed the existence of new planets. There was still suspicion of optical tubes. The argument was most absurd (perhaps from today's standpoint). For example, Francesco Sizi reasoned this way: An optical tube is similar to eyeglasses, eyeglasses cannot fit young and old people equally well, both see the planets through Galileo's tube at the same time, and so it is an optical illusion. Also, in Pisa, Giulio Libri simply refused to look through the tube. "I hope that when he goes to Heaven he will finally see my moons, which he did not want to see from Earth," Galileo said after Libri's death. Many of Galileo's opponents understood that claiming his statements contradicted Scripture were especially effective denunciations to the Inquisition.

But if this is how things stood with phenomena that could be observed directly, then what dangers threatened Galileo for his statements supporting the Copernican system! In *Sidereus Nuncius*, Galileo promised to write *Massimi Sistemi*, in which "by six proofs and arguments of natural philosophy" he confirms that "the Earth moves and surpasses the Moon in its light."[26] His reconnaissance in Rome showed clearly that at the present moment these arguments found no support among "the leading lights." Galileo did not abandon his plans, but a long siege had begun. He understood very well that accepting Copernicus was not an internal scientific question, that he first had to convince the best scientists in the world, and that this required all his effort and would take him away from his immediate scientific activities. Many scientists doubted the validity of Galileo's solution. Einstein's opinion on this is well known: "As regards Galileo, I imagine him differently. We should not doubt that he passionately sought after truth, more than anyone else. But is it hard to believe that a man of vision sees sense in joining the truth he has found to the ideas of the superficial masses, confused in their petty interests. Was such a problem important enough for him to give it the last years of his life.... Without needing to personally, he went to Rome to tilt there at priests and politicians. This picture does not correspond to my image of the aged Galileo's internal independence. I cannot imagine that I, for example, would have undertaken anything like that to defend the theory of relativity. I would have thought the truth was far stronger than myself, and it would have seemed to me ridiculously quixotic to defend it by the sword, astride Rosinante...." Galileo held to a different opinion, but he hardly seems a scientific Don Quixote. He did not tilt at "priests and politicians" as much as draw them to his side with the greatest art.

It is interesting to compare Einstein's statement with the view of the

[26] In other words, in the amount of sunlight it reflects.—*Transl.*

Pythagoreans, who were the first to accept a moving earth and fixed sun: "Let us only try to know something for ourselves, finding satisfaction just in this, and forget the desire and hope to rise in the eyes of the crowd or to strive for the approval of the bookish philosophers."

First of all, mathematicians traditionally were not supposed to discuss questions about the creation of the universe. To observe heavenly bodies, construct tables, use tables for horoscopes—these were the limits of a mathematician's duties. Galileo had no taste for making horoscopes (unlike Kepler, for instance) but at times he still had to do it. Thus, in anticipation of his move to Florence and at the insistence of the Duchess, he had made a horoscope for Grand Duke Ferdinando I (the father of Cosimo II, the present Duke), who had fallen ill. The horoscope promised a favorable turn of events, the Duke was pleased. Galileo's son-in-law obtained a position he desired, and within several days the Duke died.... You had to be at least a philosopher in order to discuss the creation of the universe (after all, even their salaries were noticeably higher than the mathematicians'), and if it conflicted with Scripture then you certainly had to be a theologian. Galileo could not become a theologian, but he could try to become a philosopher.

Galileo went through long negotiations over the title of his future position in Florence; he wanted the word "philosopher" in his title, for he had "studied more years in philosophy than months in pure mathematics." In the end, they agreed on the title "Philosopher and First Mathematician to His Highness the Grand Duke of Tuscany" (first mathematician, but not first philosopher!).

He began life in Florence in discussions with the conservative philosophers of the University of Pisa, followers of Aristotle, who assumed that the truth, "speaking in their own words, must be sought not in the universe and not in nature, but in the comparison of texts." Galileo was satisfied with his first successes: "How you, dear Kepler, would have laughed if you had heard how in Pisa, in the presence of the Grand Duke, the first philosopher of the local university came out against me, trying by arguments of logic, as if by bewitched incantations, to tear the new planets down from the heavens and destroy them!" His discussions concerned more than astronomy. In 1612, *Discorso Intorno alle Cose che Stanno in su l'Aqua* (*Discourse on Bodies Floating in Water*) appeared, devoted to hydrostatics and rather unpleasant for Aristotle's adherents. Within a year came *Istoria... intorno alle Macchie Solari...* (*Letters on the Solar Spots*), with barbs aimed in the same direction: "This news, I fear, will become the death knell or, rather, the death sentence for pseudo-philosophy.... I hope that the hilliness of the moon will become for the peripatetics a mere trickle compared to the torment of the

clouds, steam, and abundance of smoke that constantly arise, move about and disappear on the very face of the sun" (from a letter to Cesi; Aristotle's adherents were called peripatetics). Perhaps Galileo celebrated his victory prematurely. . . .

Galileo was increasingly pulled into discussions with people who were far removed from actual science. Sometimes doubts plagued him: "With unspeakable disgust, I have reached this point and, as if I were repenting for my deeds, understood how fruitlessly I have squandered time and effort." The struggle intensified. The Dominican monk Tommaso Caccini, directing his sermons against Galileo, proposed radical measures: "Mathematicians must be banished from all Catholic states!" At the same time, Galileo decided to discuss theological questions. In 1614, copies were circulated of a letter he wrote to Castelli, in which one can find words such as: "Hence it appears that physical effects placed before our eyes by sensible experience, or concluded by necessary demonstrations, should not in any circumstances be called in doubt by passages in Scripture that verbally have a different semblance, since not everything in Scripture is linked to such severe obligations as is every physical effect."[27] This very letter probably served as the means for Father Nicolò Lorini's denunciation of Galileo to the Inquisition. It turned out that Galileo was accurate enough. The ravenous qualificators[28] could find only "three evil-sounding places" in the letter, and two of these were not in the original, which the Inquisition could not obtain.

In February 1615, a book appeared in Naples by Paolo Foscarini, a member of the Carmelite order, giving an account of the Copernican system in the form of a letter to the order's general. Bellarmino used the book as a way to state his relation to the problem, in a letter to Foscarini: "It seems to me that your Reverence and Signor Galileo act prudently when you content yourselves with speaking hypothetically and not absolutely, as I have always understood that Copernicus spoke. To say that on the supposition of the Earth's movement and the Sun's quiescence all the celestial appearances are explained better than by the theory of eccentrics and epicycles is to speak with excellent good sense and to run no risk whatever. Such a manner of speaking is enough for a mathematician. But to want to affirm that the Sun, in very truth, is at the center of the universe and only rotates on its axis without going from east to west, is a very dangerous attitude and one calculated not only to arouse all Scholastic philosophers and theologians but also to injure our holy faith by contradicting the Scriptures."[29]

[27] For an English translation of this letter, see Drake, *Galileo at Work*, pp. 224–229.—*Transl.*

[28] Who examined cases and prepared them for trial.—*Transl.*

[29] English translations, in whole or in part, of Cardinal Bellarmino's letter and other relevant

We must give the head of the Inquisition his due—he expressed his opinion with the utmost clarity.

In December 1615, Galileo was again in Rome. He probably wanted to influence the path of the investigation forming against him, and had not yet lost hope of changing the Church's opinion about the Copernican system.

A "Salutary Edict"

Galileo was in every way a diplomat. He visited Bellarmino, and tried to bring Cardinal Alessandro Orsini over to his side. In a message to Orsini, he set forth his most secret argument in support of the earth's motion—the tides. He explained them by the mutual action of the daily and orbital motions of the earth, and saw no competing explanation. Galileo thought up this explanation in Venice, where he saw how the water in a boat moved when it sped up and slowed down. "This phenomenon is indisputable, easy to understand, and can be verified by experiment at any time." Simpler explanations make the Copernican system very plausible, but a definite proof of the earth's motion can only be discovered on the earth itself! The future showed that Galileo's trump card was erroneous, but the explanation came much later. Galileo was at the very center of Roman intrigue: "I find myself in Rome, where just as the weather constantly changes, instability always rules in affairs."

Everything came to an end on February 24th, when a commission of eleven theologians voted that the claim that the earth moves is "at least erroneous in faith." Galileo was told about this decision by the Commissary-General of the Inquisition in the presence of Cardinal Bellarmino. On March 5th, the Congregation of the Index "suspected" (but did not ban) Copernicus' book.[30] This act was almost symbolic. They planned to remove several phrases from the book about how the doctrine being presented did not contradict Scripture, and to correct those places where Copernicus calls the earth a heavenly body (the sun and moon were heavenly bodies!). The Tuscan ambassador, in a letter to home, regretted Galileo's persistence but expressed the hope that he would not suffer. Rumors spread that Galileo would be required to swear an oath of renunciation, and Galileo obtained reassurance from Bellarmino refuting the rumors: "...but that only the declaration made by the Holy Father and published by the Sacred Congregation of the Index has been notified to him, wherein it is set forth that the doctrine attributed to Copernicus, that the Earth moves around the Sun

documents appear in de Santillana, *The Crime of Galileo*, p. 99.—*Transl.*

[30]This was the "salutary edict" Galileo later referred to in his preface to *Massimi Sistemi*.—*Transl.*

and that the Sun is stationary in the center of the world, and does not move from east to west, is contrary to the Holy Scriptures and therefore cannot be defended or held."[31] This was in May before Galileo left Rome, and still earlier he had been received by Pope Paul V. What took place was not a sentence, but a stern warning. A violation of a clearly expressed ban was surely a crime.

Awaiting a Change

Galileo quit Rome, subject to the "salutary edict." But his obedience is not so apparent. Here, for example, is what he wrote while sending *Discorso sopra il Flusso e Reflusso del Mare* (*Discourse on the Tides*) to the Archduke Leopold of Austria, brother of the Grand Duchess: "Now, knowing as I do that it behooves us to obey the decisions of the authorities and to believe them, since they are guided by a higher insight than any to which my humble mind can of itself attain, I consider this treatise which I send you to be merely a poetical conceit, or a dream, and desire that your Highness may take it as such, inasmuch as it is based on the double motion of the Earth and, indeed, contains one of the arguments which I brought in confirmation of it."[32] It is hard to believe that this man will never say the earth moves. However, in order to return to this theme, Galileo needed not new arguments but a change in his everyday situation. And he waited for a change. Pope Paul V died, Giovanni Ciàmpoli, who was kindly inclined towards Galileo, became the influential secretary to the new Pope Gregory XV, and in 1621 the terrible Cardinal Bellarmino died. In 1623, Cardinal Maffeo Barberini, an educated man and patron of the sciences who did not hide his admiration for Galileo, became Pope Urban VIII.

At this time Galileo's pace quickened noticeably. In 1623 his book *Il Saggiatore* (*The Assayer*) appeared, a response to Orazio Grassi, an astronomer of the *Collegio Romano*, that was devoted to comets. Here he still did not speak directly about the earth's motion. But his following work, *Letter to Ingoli*, written in 1624, directly relates to this question. It was a reply to a 1616 essay by Francesco Ingoli, a highly educated clergyman, directed against the Copernican system. It is significant that Galileo waited eight years to respond. There are many brilliant and daring pages in this slim volume. There is even a poetic description of shipboard experiments that do not show the ship is moving, a remarkable explanation of the law of inertia; there are also arguments involving the fixed stars, comparing them

[31] De Santillana, *The Crime of Galileo*, p. 132.
[32] Ibid., p. 151.

to the sun, and even a free discussion of the question of the size of the universe.

As for the latter, there is not even a hint of a universe bounded by "eight heavens" of fixed stars. Galileo clearly explains that he sees no arguments that allow one to choose between the hypotheses of a finite or infinite universe, but completely admits that only a small part is accessible to us: "...I am not at all sickened by the thought that the world, whose boundaries are set by our external senses, may turn out to be as small in relation to the Universe as the world of worms is in relation to our world." Galileo very daringly admits the hypothesis that the universe is infinite! Recall how uncomfortable the great fantasizer Kepler felt in assuming an infinite number of worlds similar to the solar system in his *Dissertatio*: "If you had discovered any planets revolving around one of the fixed stars, there would now be waiting for me chains and a prison amid Bruno's innumerabilities, I should rather say, exile to his infinite space."[33]

Letter to Ingoli was written in the autumn of 1624, and in the spring of 1625 Galileo visited Rome again. It seems that his goal was to make contact with the new pope, and to judge how favorable the situation had become. Galileo met with the pope six times, was treated very well by Barberini's large family, and established favorable connections with many cardinals, including the influential German Cardinal Eitel Friedrich von (Hohen) Zollern. Relations with Galileo personally could not have been better, but his main hope was not justified: Urban VIII strongly supported the assertion of the "salutary edict" that the sun moves and the earth is stationary. Galileo discovered that in discussing this question, he and the pope had been speaking in different languages. Galileo claimed that the tides cannot be explained without assuming that the earth moves, but was told that what is unknown to people may be known to God. It is hard to argue with such reasoning! Galileo returned, and afterwards the pope sent a message to Grand Duke Ferdinando II (Cosimo had just died) expressing satisfaction with the Florentine scientist's visit and giving the most laudatory opinion of him.

Massimi Sistemi (World Systems)

Returning from Rome, Galileo finally decided to write a book setting forth all the arguments supporting the Copernican system. He had dreamt of this book in 1597 when he wrote to Kepler, had promised it in *Sidereus Nuncius*, and had considered it his main goal in moving to Florence. Galileo had

[33] *Dissertatio*, p. 36.

turned sixty, and his health left something to be desired. The journey to Rome had not been a complete success, but there was no point in waiting for a better time. It would seem that after the "salutary edict" which, as we have explained, was strongly supported by Rome's "leading lights," he could not think of openly supporting the heliocentric system, but Galileo was not accustomed to guile.

Even in theological disputes, one of the participants was permitted to defend a heretical point of view "conditionally," so as to unmask it more graphically. The Copernican system was not declared heretical, and even Bellarmino had allowed it to be spoken of "hypothetically," as a mathematical construction. Galileo devised an artifice. Three interlocutors, Salviati, Sagredo, and Simplicio,[34] meet at Sagredo's palace and "dispassionately" discuss both world systems over the course of six days. The first two heroes are named for Galileo's deceased friends, and the third—an adherent of Aristotle and Ptolemy—is imaginary.

For more than five years, Galileo anxiously worked on the book; it goes without saying that he thought of it as the major work of his life. By 1630, four of the six days were finished: On the first day they discussed the possibility of the earth moving, on the second day its daily motion, on the third its yearly motion, and, finally, on the fourth day the tides, Galileo's most beloved find. He decided to limit himself to four days, and to call the book *Dialogo del Flusso e Reflusso* (*Dialogue on the Tides*). In the spring of 1630, Galileo sent the manuscript to Rome.

Galileo's book, in today's terminology, should really be called popular science. He consciously addressed it to the public at large, not just to scientists; he wanted to convince everyone that there were irrefutable arguments in favor of Copernicus. Partly because of this and partly because of his own scientific tastes, Galileo dealt almost exclusively with qualitative phenomena, without linking the system to the numerical data of astronomical observations. His planets moved uniformly in circles around the sun, which had no chance of agreeing with the observational data. In this regard, Galileo was significantly inferior to Kepler and avoided discussing the problems that bothered Copernicus. Evidently, computational astronomy was not Galileo's strength.

Galileo obtained an audience with the pope, and met with the influential cardinals. Urban VIII was not against a book that would contain conditional arguments in support of a condemned system, but it could not create the feeling that the reader had a choice between two systems. The book must indicate unambiguously the finality of the assertion, sanctified

[34] Who later appeared in *Discorsi*.—*Transl.*

by the Church, that the sun moves and the earth does not. Moreover, the pope rejected the title *Dialogo del Flusso e Reflusso*. Galileo promised to satisfy the pope's wish for a yet-unwritten introduction and conclusion. The manuscript was given to Niccolò Riccardi, the Master of the Holy Apostolic Palace (the chief licenser), also known as Padre Mostro, for an opinion. Padre Mostro chose a delaying tactic; unlike Galileo he was in no hurry.

The rest sounds like a detective story, with Galileo and his supporters acting with wonderful ingenuity so that the book would see the light of day. Ciàmpoli, the former papal secretary, evidently resorted to fraud just to obtain preliminary consent, risking his career. The book was supposed to be printed in Rome. With enormous cunning, with references to Galileo's health, plague in Italy, etc., it was printed in Florence.

On February 22, 1632, Grand Duke Ferdinando received a present, the first copy of the book dedicated to him: "A Dialogue of Galileo Galilei Linceo, Extraordinary Mathematician of the University of Pisa and Philosopher and Chief Mathematician of His Highness the Grand Duke of Tuscany, where in four days of meeting the two Grand Systems of the World of Ptolemy and Copernicus are discussed, and the philosophical and physical reasons for one side and the other are indefinitely propounded." The preface, addressed to the "discerning reader," explains the author's motives in presenting arguments supporting the Copernican system. He recalls the "salutary edict which, in order to obviate the dangerous tendencies of our present age, imposed a seasonable silence upon the Pythagorean opinion that the earth moves."[35]

Galileo's "zeal could not be contained" by the spreading rumors "that this decree had its origin not in judicious inquiry, but in passion none too well informed." The book must refute these rumors. He wants to show "foreign nations that as much is understood of this matter in Italy, and particularly in Rome, as transalpine diligence can ever have imagined. Collecting all the reflections that properly concern the Copernican system, I shall make it known that everything was brought before the attention of the Roman censorship, and that there proceed from this clime not only dogmas for the welfare of the soul, but ingenious discoveries for the delight of the mind as well." Finally, "it is not from failing to take count of what others have thought that we have yielded to asserting that the earth is motionless, and holding the contrary to be a mere mathematical caprice, but... for those reasons that are supplied by piety, religion, and knowledge of Divine

[35]The quotations here and below are taken from the *Massimi Sistemi* section of Galileo Galilei, *Dialogue Concerning the Two Chief World Systems—Ptolemaic and Copernican*, translated by Stillman Drake, 2nd revised ed., University of California Press, Berkeley, 1967, pp. 5–6.—*Transl.*

Omnipotence, and a consciousness of the limitations of the human mind." In fact, his goals had to have been seen in Rome as worthy: to cut off talk about the rashness of the edict and to put the "foreign nations" in their place. Nevertheless, certain statements seem ambiguous today and may have also seemed so to some of the "leading lights." At the least, soon after copies of *Massimi Sistemi* appeared in Rome, there was a thunderclap.

Trial and Renunciation

The initiative for pursuing Galileo evidently came from Urban VIII himself. What angered the pope so and made him unappeasable? Perhaps he found the praise of the "seasonable salutary edict" insincere? *Massimi Sistemi* undoubtedly appeared at a very difficult time for Urban. The strong Spanish opposition in Rome was trying to remove the pope, and he could have been greatly threatened by accusations of supporting a scientist "suspected of heresy." It was said that the pope saw himself in the simpleton Simplicio, who defended the immobility of the earth. Galileo writes in the preface that this hero, unlike the two others, is not called by his proper name. What must Urban have thought if he really discovered something he had once told Galileo in Simplicio's verbiage?

In August 1632, the Papal Curia forbade the distribution of *Massimi Sistemi*. In September the matter was given to the Inquisition. A protracted game began. Galileo's supporters, including the Grand Duke, tried first to avoid consideration of the issue by the Inquisition, then to move the inquiry to Florence, and finally to procrastinate as much as possible, referring to Galileo's illness. All these attempts led nowhere—Urban VIII was implacable.

A threat to put Galileo in chains made him leave for Rome in January 1633. He arrived on February 13th, and on April 12th stood before Vincenzo Macolani, the Commissary-General of the Inquisition. An agonizing inquiry began, pressure was applied, and Galileo was apparently shown instruments of torture. An exhausting struggle to find a compromise took place. Three qualificators of the Holy Office concluded that the book at least violated the ban on holding and spreading condemned doctrines. Galileo admitted that, against his wishes, he had strengthened the arguments in favor of the Copernican system. On June 22nd, in the monastery of Santa Maria sopra Minerva, the kneeling Galileo, who would reach seventy in half a year, heard the verdict. Because he "believed. . . that an opinion may be held and defended as probable after it has been declared and defined to be contrary to the Holy Scripture,"[36] Galileo was declared to be "ve-

[36]Complete English translations of Galileo's sentence and renunciation appear in de Santil-

hemently suspected of heresy," and *Massimi Sistemi* was banned. Galileo was sentenced to "the formal prison of this Holy Office" (one suspected of heresy was not burned as a heretic!), and he must "for three years. . . repeat once a week the seven penitential Psalms." Then Galileo read out the text of renunciation: ". . .after it has been notified to me that the said doctrine was contrary to Holy Scripture—I wrote and printed a book in which I discuss this new doctrine already condemned and adduce arguments of great cogency in favor without presenting any solution of these. . . ." He swore to "fulfill and observe in all their integrity all penances" placed on him.

Perhaps at that moment Galileo regretted that he had abandoned the Republic of Venice, where he had been beyond the Inquisition's reach, and reassessed the Grand Duke's capabilities. But in Venice there had evidently been no chance of publishing his major work, which, regardless of the terrible consequences, he had succeeded in doing in Florence.

The Inquisition's prison sentence was replaced by exile, at first in the Medici palace in Rome; in two weeks, they sent him to Siena, to the archbishop Ascanio Piccolomini. In half a year they decided to move him again, to his villa in Arcetri, near the convent where his daughters were. Galileo live there for his eight remaining years, except for a short trip to Florence. Everywhere, he was under the vigilant eye of the Inquisition, which carefully controlled his contacts with the outside world. Urban VIII showed no pity to the disgraced scientist, even on the day of his death. His relative Cardinal Francesco Barberini wrote to Florence: ". . .it is not good to build a mausoleum for a corpse who was punished by the tribunal of the Holy Inquisition and died while under that punishment." The Grand Duke was unable to bury Galileo next to Michelangelo (this wish was fulfilled after many years).

Galileo's renunciation continues to disturb people even today. Did a scientist have the right to renounce a theory that he believed to be true without a doubt and to whose confirmation he had given a significant part of his life? Various explanations were put forth for Galileo's decision: The seventy-year-old ailing scientist's fear of torture and burning, the feeling that he had fulfilled his mission and that nothing could any longer interfere with the distribution of his book, and the possibility of preserving what proved to be his eight remaining years for scientific work (he returned to the studies he had interrupted for a quarter of a century, thanks to working out the ideas he now was forced to renounce). Constance Reid's book, *Hilbert*, tells what that great mathematician, with characteristic directness, said about Galileo: "But he was not an idiot. Only an idiot could believe

lana, *The Crime of Galileo*, p. 306 ff.—*Transl.*

that scientific truth needs martyrdom—that may be necessary in religion, but scientific results prove themselves in time."[37] Keep in mind that Galileo had also compromised before, and even after 1616[38] had formally acknowledged the immobility of the earth (and in *Massimi Sistemi* as well).

Galileo apparently never spoke the legendary phrase, *eppur si muove* ("still it moves"),[39] but regardless of his unquestionable faith his renunciation could not have been sincere. He must have been happy that *Massimi Sistemi* was not completely suppressed and that in 1635 a Latin translation appeared in Europe. Fulgenzio Micanzio, a Venetian acquaintance, wrote to him: "A remarkable thing—after your *Massimi Sistemi* came to light, people knowledgeable about mathematics immediately went over to the side of the Copernican system. This is what the bans have led to!" Galileo answered: "What you have written to me about *Massimi Sistemi* is most unpleasant for me, since it can cause great concern among the leading lights. After all, the permission to read *Massimi Sistemi* is so limited that his Holiness keeps it only for himself, so that in the end it may happen that they will forget about this book completely."

The disgrace of the trial and verdict was difficult for Galileo, but so was the ban on continuing his work on the problem of the universe. He had no doubt that he should refrain from such work, but what was left for him? He had every reason to regret this period: "Our times are unhappy, a firm resolve to eradicate every new thought, especially in the sciences, is now the rule, as if everything possible to know was already known!" He could comfort himself with the predictions of the like-minded Tommaso Campanella in his *Defense of Galileo*, written in a Naples prison in 1616: "The coming century will judge us, for the present always crucifies its benefactors, but they later rise on the third day or in the third century."

Several weeks after the verdict, Galileo remembered the treatise on mechanics that had been cut short, and writing this book became his main endeavor for the coming years, the goal of his life. He recalled his youthful discovery of the isochronic property of the pendulum and entrusted his son Vincenzio with making a pendulum clock. Galileo's blindness was inexorable. By the end of his work on the book he had lost his vision in one eye, but still looked at the sky from time to time through his telescope and described the libration of the moon, until at the end of 1637 he was totally blind: "...this sky, this world, and this universe, that with my startling observations and clear proofs I have extended a hundred and a thousand times compared to how the sages of all past centuries had usually seen it,

[37] Constance Reid, *Hilbert*, Springer-Verlag, New York, 1970, p. 92.
[38] The year of the "salutary edict."—*Transl.*
[39] For more on this story, see Drake, *Galileo at Work*, pp. 356–357.—*Transl.*

has now so shrunk and narrowed for me that it has become no larger a space than is occupied by my own body. Because it has so recently occurred I still cannot regard this unhappiness with patience and resignation, but the passage of time should accustom me to it." Nevertheless, in the last year he was given, he again looked at the Medicean stars, and his old friends drew him to the idea that took possession of him in his last days.

The Medicean Stars Revisited

This idea may have occurred to Galileo even earlier, at the end of 1635, when he gave the French commission created by Cardinal Richelieu an opinion on the method of Jean-Baptiste Morin (1583–1656) for determining longitude by observing the moon's motion. The method turned out to be unsound, but note how high-ranking a person was interested in it. The point is that finding the longitude aboard ship was one of the most pressing problems of the 17th century, the century of seafaring. Today it is hard to believe that at that time, sailors completed long voyages without any sort of reliable method for measuring the coordinates of a ship on the open sea. This of course did not apply to latitude, which, at least by the 16th century could be reliably measured (for example, by the sun's height at noon). Scientists could propose nothing workable for longitude. This problem worried the maritime powers, especially for economic reasons. The author of a method for measuring longitude with acceptable accuracy (say to half a degree) could at various times receive 100,000 écu from Philip II of Spain, or 100,000 livres from Louis XIV, or 20,000 pounds from the English Parliament, or 100,000 florins from the States General of Holland. Less accuracy decreased the prize proportionally. These numbers quite expressively demonstrate interest in the problem.

As long ago as the 2nd century B.C., Hipparchus had an idea for measuring longitude: It used the fact that the difference in longitude between two points on the earth's surface is proportional to the difference in local time at these points. Thus at points whose longitudes differ by, say, 15°, the difference in local time equals one hour ($\frac{360^\circ}{24} = 15^\circ$). Therefore the problem can be reduced to measuring local time aboard ship and the corresponding time at some fixed point, for example, at the port from which it sailed. It is practical to measure the local time at the point where the ship is now, but how can we know the local time at the port? For a long time, no one even thought of "keeping" it. An excellent example is the story of the twenty-fours "lost" when Magellan sailed around the globe! And there were no clocks that could have kept that time, especially with the rocking motion of the sea.

Another possibility was to use astronomical phenomena that could be observed aboard ship and for which the precise time it would be observed in port was known. But few phenomena were suitable! What could be used, aside from solar and lunar eclipses, which are very rare? Tables for the moon's motion were so imperfect that longitude could not be measured from daily lunar observations (an example of such attempts was Morin's method). Galileo described the situation with a characteristic sense of triumph: "In former times, heaven was generous on that count, but for current needs it is rather stingy, assisting us only with eclipses of the Moon: and not because the same heaven does not abound with frequent phenomena that are visible and more suitable for our needs, but it has been convenient for the rule of the world to conceal them right up to the present. . . ." The optimism we feel in these words is connected to the hopes Galileo placed on the Medicean stars he had discovered. Jupiter's satellites. Among their peculiarities, discovered at the time of the first observations in 1610, are partial eclipses. If the moon's orbit were not inclined towards the earth's, the moon would fall in the cone of the earth's shadow at each full moon. Jupiter's moons fall in the broad cone of its shadow at each revolution, and they revolve rather quickly (Io completes a full revolution in about 42.5 earth-hours). While observing the eclipses of Jupiter's moons, Galileo decided to develop his own method for measuring longitude aboard ship.

Galileo began negotiations, without waiting to work out the method definitively. First he thought of Spain (it was probably important that this was a traditional Catholic country) and of meeting the viceroy in Naples, but gradually switched to Holland, where his idea aroused great interest. In 1636 secret negotiations with the States General were in full swing, and in August it was decided to ask Galileo for the necessary materials. Galileo wrote a triumphant message to the States General, the "tamer and ruler of the ocean." The quote given above was taken from this message. Galileo considered it symbolic that the telescope, which plays a leading role in his method, was invented in Holland. He did not stint in his description of the preeminence Holland would obtain through his method: "I could name a collection of arts, but it suffices to limit myself to seafaring, which has been brought by your Dutchmen to such startling perfection, and if the only remaining thing—the determination of longitude in which, as we see, they have so far been unsuccessful—joins the list of the rest of their clever operations, thanks to their recent and greatest invention, then their glory would reach such extremes that no other nation could dream of surpassing it."

A competent committee was formed, including Admiral Laurens Reael, the astronomer and mathematician Martinus Hortensius, and later Con-

stantijn Huygens, member of the Council of State and father of the great scientist Christiaan Huygens. It was not easy for the practical Dutch to believe the proposed method was feasible. "Imagine to how many people of high position and wealth we were forced to preach a hitherto unknown truth, that was first taken to be unreasonable," lamented Huygens. Even the most supportive members of the committee were not sure that the project could be realized. In a letter to Galileo, Admiral Reael feared that his method might prove to be too refined, "for so coarse a people as the Dutch sailors." Doubts can be felt even in Huygens' words: "Our people will with difficulty consider themselves indebted for a grand gift that is more beautiful than profitable." Even Hortensius had difficulty in adapting to seeing Jupiter's moons. Good telescopes were not enough. At the end of 1637 Galileo sent his own telescope, which he could no longer use because of his blindness. Tables were necessary to observe the moons, and they were not easy to construct (for a long time not even the periods of revolution could be determined).

Computational astronomy was never Galileo's strong suit, and now blindness robbed him even of the ability to make observations. Galileo asked the Olivetan monk Vincenzio Renieri, an experienced computational astronomer, to find the ephemerides of Jupiter's moons, at least for the coming year. The calculations were delayed, and Renieri did not succeed in constructing the tables that were needed.

The States General instructed Hortensius to meet with Galileo, to firm up some necessary details, and to present him with a gold chain, as a gift. At that point, the Inquisition interfered with the negotiations. A complicated game began, and as a result Galileo either thought it wise not to meet Hortensius and accept the gift, or was directly prohibited to do so by the Inquisition. Discussions began over keeping priority for Italy. Castelli, who for a long time had not been allowed to see Galileo, even received permission to meet with his teacher and learn the details of the method. Hortensius and Reael unexpectedly died; Galileo's strength deteriorated. The Florentine inquisitor informed Rome that the scientist, "completely blind, will lie in his grave before he [again] studies mathematical constructions." Galileo did not lose hope, but it became clear that he would not live to see the realization of his idea. In fact, it was probably impossible to carry out the project. A long time went by before the problem of measuring longitude at sea was finally solved, but in a completely different way—using accurate maritime clocks.

One of Galileo's last statements shows that he never stopped thinking about the major question of his life, and speaks to his "incorrigibility": "And just as I deem inadequate the Copernican observations and con-

jectures, so I judge equally, and more, fallacious and erroneous those of Ptolemy, Aristotle, and their followers, when [even] without going beyond the bounds of human reasoning their inconclusiveness can be very easily discovered."[40] He was not allowed to argue against there being arguments inaccessible to man that refute Copernicus, but arguments accessible to man are enough to refute Ptolemy.

Epilogue

We see much about these events of three and a half centuries ago differently from Galileo. This involves both the difference between the Ptolemaic and Copernican systems and the question of the earth's "true" motion.

It is difficult to construct a consistent system of the universe that does not at heart rest on celestial mechanics. Paradoxically, Galileo's theory of celestial mechanics, as opposed to his "terrestrial" mechanics, was rather naive and close to Aristotle's view. First, he assumed that celestial bodies move because of inertia, rather than because of constantly acting forces. He had no acceptable notion of forces acting from afar and, for example, the idea of the solar or lunar attraction of terrestrial objects was considered an astrological anachronism. Second, according to Galileo, celestial bodies moving by inertia exhibit uniform rotary motion. This is a *prima facie* contradiction of the "terrestrial" principle of inertia!

The main question for Galileo was the true (absolute) motion of the earth, and its experimental proof. Since terrestrial phenomena must be used for the proof, the terrestrial and celestial principles of inertia inexorably collide. With the greatest insight, Galileo refuted Tycho Brahe's claim, repeated by Ingoli, that phenomena aboard a moving ship must reveal that motion. Galileo's refutation (essentially the first statement of the "terrestrial" law of inertia) was mostly based on experiment. Simultaneously, he claimed that there are phenomena (the tides) that do reveal the earth's motion. How the hypothetical motion of the earth differs from a ship's motion, which cannot be discovered internally, is not clarified.

We emphasize that these phenomena were supposed to have been consequences of the earth's own motion, occurring by inertia without the participation of forces acting from afar. Galileo saw no contradiction here. As we have already noted, Galileo's "decisive" argument turned out to be completely wrong.

Galileo's view of true (absolute) motion was incorrect. The author of the law of inertia was still far from understanding the relative nature of

[40] Ibid., p. 417.

motion and the role of a frame of reference. Christiaan Huygens did much to clarify this aspect of motion. Newton (unlike Huygens) assumed revolution was absolute. The Ptolemaic and Copernican systems use different frames of reference: The earth is fixed in one, and the sun in the other. The development of mechanics has shown that an opportunely chosen frame of reference is needed to reveal the laws of motion. The chief merit of the Copernican system was that it made the revelation of Kepler's laws (which, by the way, Galileo did not accept) possible. In Copernicus' system, the most massive body serves as a fixed origin, and as a first approximation in considering an individual planet we can restrict ourselves to the planet's interaction with the sun (its interaction with the other planets is negligible). This is the two-body problem, and Kepler's law immediately follow from Newton's law of universal gravitation, as Newton showed. In a frame of reference where the earth is stationary it becomes more complex to describe motion and, in particular, Kepler's laws do not hold.

Galileo's astronomical observations opened a new era in astronomy, and the moons of Jupiter played a special role. More than half a century passed before their periods were calculated, which Galileo himself and the astronomers of the *Collegio Romano*, who were experienced in computation, had tried to do. Calculating their distances from Jupiter was even harder, because of insufficiently developed measurement techniques. But in 1685, when Newton published *On the System of the World*, part of his *Principia Mathematica* (*The Mathematical Principles of Natural Philosophy*), he could already say that Jupiter's moons obeyed Kepler's third law $T^2 \sim R^3$, where T is the period of revolution and R is the distance from Jupiter, although the data needed to be made more accurate. This was in the section *Phaenomena*, listing the experimental facts on which Newton's "world system" relied.

Constructing a theory of motion for the moons of Jupiter, based on the law of universal gravitation, tested the ambitions of the founders of celestial mechanics for a long time. A sufficiently precise theory needed to account not only for Jupiter's attraction, but also for the sun's attraction and for the mutual attraction of the moons. In 1774 this problem was the theme for a prize given by the French Académie des Sciences.

Laplace constructed a rather precise theory in 1789. For a long time, the Medicean stars remained a goal that not one of the great astronomers could pass up. They presented the scientists with new and ever-surprising facts. Thus, for example, Laplace established that the time it takes for the first moon to revolve plus twice the time for the third is three times that of the second. But undoubtedly the most remarkable page in the study of Jupiter's moons is a discovery of Olaf Römer, which we will describe in detail.

Appendix: Olaf Römer's Conjecture

Cassini's Observations

Gradually, the telescope became a recognized astronomical tool. The power of telescopes grew: Christiaan Huygens' telescope gave 92-fold magnification, and in 1670 a telescope appeared in Paris that magnified objects 150 times. It is characteristic that this telescope was no longer at the disposal of a single scientist: It was installed at a new type of scientific institution, an observatory. The Paris Observatory, under the patronage of Louis XIV, was directed by Jean-Dominique Cassini (1625–1712), an Italian astronomer. Astronomy owes much to Cassini. He discovered that Saturn has four moons besides the one (Titan) discovered by Huygens, and the ring Huygens found around Saturn turned out, under Cassini's more careful observations, to consist of two rings separated by a gap (which came to be called Cassini's division). Cassini proved that Jupiter and Saturn rotate on their axes. He also did great work in the area of astronomical computation, measuring the astronomical unit, the distance from the earth to the sun, with an accuracy unheard of at the time. It is interesting to compare Cassini's value of 146 million kilometers with the true value of 149.6 million, and the 8 million that had been previously assumed.

As we have already noted, calculating the periods of revolution of Jupiter's moons became one of the central problems of astronomy in the second half of the 17th century. These values can be obtained by straightforward calculations if we know the successive times of their eclipses accurately. Conversely, knowing the periods of the moons, we can predict the times of their eclipses. In 1672 Cassini very carefully recorded the eclipses of Io, one of Jupiter's moons. He was surprised to find that his values for Io's period differed from time to time, as if the eclipse were sometimes a bit late and sometimes a bit early. The greatest difference he obtained was 22 minutes (for a period of 42.5 hours), and could not be explained by the accuracy of the measurement. Evidently, Cassini was already able to use Huygens' pendulum clock, which had begun to be used for astronomical observations. The observed effect had no reasonable explanation.

In 1672, the year Cassini systematically observed Jupiter's moons, a young Danish scientist named Olaf Römer (1644–1710) appeared at the Paris Observatory. He was intrigued by a striking coincidence (that Cassini may have also noticed): The greatest delay in Io's eclipse occurred at those times when Jupiter was furthest from the earth. It was possible that he noticed this phenomenon accidentally, but he must have had foresight not to explain it as an accident! Although at the time of Louis XIV the earth was

still at the center of the universe in the astronomical atlases, scientists were not prepared to explain a change in the revolution of Jupiter's moon by the earth's influence! Römer proposed a competing explanation that must have seemed no less fantastic. He suggested that there is a delay in seeing Io's eclipse because its light travels a greater distance when the distance between the earth and Jupiter is greater. In order to evaluate Römer's hypothesis, we must recall what his contemporaries thought about the speed of light.

Digression on the Speed of Light

The ancient scholars assumed that light travels instantaneously (the only exception may have been Empedocles). For many centuries, this opinion was reinforced by Aristotle's authority. In the East, Avicenna and Alhazen[41] assumed the speed of light is finite but very large. Of the later European scientists, Galileo was one of the first who was ready to assume the speed of light is finite. In *Discorsi*, the interlocutors Sagredo, Simplicio, and Salviati discuss the problem.[42] Sagredo raises the question, and Simplicio assumes the speed of light is infinite since we see the flash of an artillery shot "without lapse of time." For Sagredo, the fact that the sound arrives after a noticeable delay means only that sound travels significantly more slowly than light. In response, Salviati, representing Galileo's interests in this triumvirate, proposes an experiment with two observers supplied with lanterns, where each one uncovers his lantern when he sees the other's light. But this experiment, which the scientists of the Florentine *Accademia* in fact tried to carry out, does not have a real chance of convincingly showing that the speed of light is finite. (Einstein and Infeld remark that for this they would have had to determine an interval of time on the order of $\frac{1}{100,000}$ of a second.[43] Kepler assumed that light travels instantaneously; Robert Hooke thought its speed is finite but impossibly large to measure. Descartes and Fermat assumed it is infinite, which greatly complicated their research in geometric optics. Descartes assumed on the one hand that light travels instantaneously, but on the other hand decomposed its "speed" into components. Fermat, in trying to avoid talking about the speed of light while stating his famous principle of least time, resorted to every possible subterfuge, talking about "the antipathy of light towards matter" and introducing a formal coefficient that for all practical purposes

[41] An Egyptian physicist and mathematician, al-Hasan ibn al-Haitham (c.965–1039).—*Transl.*

[42] *Discorsi*, pp. 42–44.

[43] Albert Einstein and Leopold Infeld, *The Evolution of Physics*, Simon and Schuster, New York, 1938, p. 95.

is a ratio of speeds of light. Thus, most of Römer's contemporaries were not prepared to acknowledge the finiteness of the speed of light, let alone make it responsible for phenomena that were perfectly tangible but occurred on the astronomical scale. By comparison, we note that the speed of sound was only recently measured.

Römer's Calculations

Römer's calculations were of the utmost simplicity. He began with the fact that twenty-two minutes, the maximum delay in the onset of Io's eclipse, is exactly the time it takes light to travel a distance equal to the difference between the greatest and least distances between the earth and Jupiter. This difference is twice the distance between the earth and the sun. Compared to this, we may neglect the distance from Jupiter to Io.

We see that Römer had still another reason to be grateful to Cassini: a rather precise value for the distance from the earth to the sun (146 million kilometers). According to Römer, light thus requires 1320 seconds (22 minutes) to travel 292 million kilometers, and so he obtained 221,200 kilometers per second for the speed of light. One error lay in his value for the astronomical unit (the true value is 149.6 million kilometers), but the main error was a very large mistake in the maximum delay (the true value is 16 minutes and 36 seconds). For the correct values he would have obtained 300,400 kilometers per second for the speed of light, which is very close to the true value (299,792.5 kilometers per second). It is striking that Römer succeeded in obtaining a value of the correct order of magnitude.

Römer carried out these calculations in September 1676. To convince scientists he was correct, he thought of a stunt worthy of the ancient Egyptian priests. He carried out his calculations and predicted that in November, Io's eclipse would be ten minutes late. Observations, in which Cassini took part, proved that Römer had accurately predicted the time to within a second. But this agreement did not make too great an impression on those around him, at least not the Parisian *Académie*, among whom the Cartesians (adherents of Descartes) predominated. After all, their teacher had written about astronomers that "although their propositions are always wrong and unreliable, they draw quite correct conclusions, relying on the various observations they make." Even Cassini refused to support Römer! This sort of thing is not at all rare in the history of science. Römer did have his adherents, including the English astronomer Edmund Halley (1656–1742).

Römer's theory was finally accepted when, in 1728, James Bradley (1693–1762) studied a visible annual motion of the stars—aberration. It had a natural explanation as the result of adding the speed of light leaving

the stars to the speed of the earth in its orbit. Bradley found that the speed of light was 10,000 times that of the earth, which agreed well with Römer's figure. The fact that two essentially different paths led to the same answer convinced many people. The first measurement of the speed of light as the result of a "terrestrial" experiment was made by Armande Fizeau in 1849.

In telling about Galileo's discoveries today, we should not forget that the space probes Voyager 1 and Voyager 2 have enabled us to learn about the surfaces of Jupiter's Galilean moons. The probe that was launched in 1989 especially to study Jupiter carried Galileo's name.[44] The science writer Jonathan Eberhart wrote about what scientists saw in the pictures that Voyager transmitted to earth: The Galilean satellites are "no mere collection of rockballs. Callisto, farthest of the four from Jupiter, presented perhaps the oldest, most heavily crated surface yet studied. Ganymede... a whole gamut of tectonic thrashings, twistings, turnings, and slippings. Europa amazed onlookers... smoother than a billiard ball, yet crisscrossed with myriad linear features that may be cracks left by global wrenchings but which somehow survived through the eons in the icy crust. And finally, stunning Io, bedecked in red and gold, silver, black and white, seething with sulfurous volcanic activity that is one of the major discoveries in the history of planetary exploration. A whole, previously unimagined family of exotic worlds, each radically different not only from its companions, but also from everything else in the planet-watcher's experience."[45]

[44] See http://www2.jpl.nasa.gov/galileo/ for photos and other information about the mission.—*Transl.*

[45] *Science News*, Science Service, Inc., April 19, 1980, p. 251. Reprinted with permission.

Christiaan Huygens and Pendulum Clocks

> [The cycloid] pendulum was invented by Christiaan Huygens, the most ingenious watchmaker of all time. Sommerfeld, *Mechanics*[1]

We have told how Galileo laid the foundation for classical mechanics almost at the beginning of the 17th century. Christiaan Huygens (1629–1695) was Galileo's immediate scientific successor. In Lagrange's words, Huygens "was destined to improve and develop most of Galileo's important discoveries."[2] There is a story about how Huygens, at age 17, first came into contact with Galileo's ideas: He planned to prove that a projectile moves horizontally along a parabola after launch, but discovered a proof in Galileo's book and did not want "to write the *Iliad* after Homer." It is striking how close Huygens and Galileo were in scientific spirit and interests.

It sometimes seems that a rejuvenated Galileo was again perfecting his optical tubes and continuing the astronomical observations he had interrupted forty years before. He tried to use the most powerful telescope to guess the secret of Saturn, which appears as a trio of joined stars, and finally, looking through a 92× telescope (Galileo's was 20×), discovered that the adjacent stars were Saturn's rings.[3] He returned again to the problem that was of such keen interest in 1610: Do any planets besides the earth and Jupiter have moons? At that time Galileo wrote the Medicis that no

[1] Arnold Sommerfeld, *Lectures on Theoretical Physics*, Vol. 1, translated by M. O. Stern, Academic Press, New York, 1964, p. 94.

[2] *Mécanique Analytique*, p. 207—*Transl.*

[3] Huygens proposed a solid ring.—*Transl.*

Christiaan Huygens.

moons had been discovered around the other planets, and that no royal house except for the Medicis, in whose honor he had named the moons of Jupiter, could claim its "own" stars. Huygens discovered Titan, a moon of Saturn, in 1655. Times must have changed, because Huygens did not offer the moon he discovered to anyone as a gift.

Huygens turned next to mechanics, where he was concerned with the same problems as Galileo had been. He developed his principle of inertia, stating that not only is it sometimes impossible to discover motion by internal means, but that the very assertion that the body moves has no absolute meaning. Huygens understood every motion as relative, which was quite unlike Newton's view. At one point Galileo, reflecting on why a body stays on the earth's surface during its rotation, almost obtained the formula for centripetal acceleration, literally not taking the last step (see p. 42). Huygens completed Galileo's argument and obtained one of the most remarkable formulas in mechanics.

Huygens then turned to studying the isochronous nature of the oscillations of a mathematical pendulum. This was probably Galileo's first discovery in mechanics, and here too Huygens was able to add to what Galileo had done: A mathematical pendulum turns out to be isochronous

(the period of oscillation of a pendulum of fixed length is independent of the amplitude of its swing) only approximately, for small angles of swing. Finally, Huygens brought to fruition the idea that occupied Galileo in his last years: he constructed a pendulum clock.

Christiaan Huygens worked on the problem of creating and perfecting clocks, especially pendulum clocks, for nearly forty years, from 1656 to 1693. One of Huygens' fundamental memoirs, containing his results in mathematics and mechanics, appeared in 1673 under the name *Horologium Oscillatorium* (*Pendulum Clocks, or Geometric Proofs Relating to the Motion of Pendula Adapted to Clocks*).[4] Huygens invented much in trying to solve one of the most fundamental problems of his life, creating a clock that could be used as a marine chronometer; he thought through many things from the standpoint of their application to this problem (the cycloid pendulum, the theory of developments of curves, centrifugal forces, etc.). We will talk about Huygens' research in chronometry, but first of all we will explain why the problem of making a clock attracted the great scientist.

Clocks are among the most ancient human inventions. At first there were solar, water, and sand clocks; in the Middle Ages, mechanical clocks appeared. The measurement of time played different roles in peoples' lives in different eras. The German historian Oswald Spengler, noting that mechanical clocks were invented at the emergence of the Romanesque style and of the movement leading to the crusades, writes: "...the mechanical clock, the dread symbol of the flow of time, and the chimes of countless clock towers that echo day and night over West Europe are perhaps the most wonderful expression of which a historical world-feeling is capable. In the timeless countrysides and cities of the Classical world, we find nothing of the sort.... In Babylon and Egypt water-clocks and sundials were discovered in the very early stages, yet in Athens it was left to Plato to introduce a practically useful form of clepsydra [a variety of water-clock—S.G.], and this was merely a minor adjunct of everyday utility which could not have influenced the Classical life-feeling in the smallest degree."[5]

It is typical that in the first steps of the new mechanics and mathematical analysis, time did not immediately take the role of a fundamental variable quantity in the description of motion (in his search for the law of free fall, Galileo began with the hypothesis that velocity is proportional to distance, rather than time).

For a long time, mechanical clocks were inconvenient and imperfect.

[4]This appears, with a French translation, in Huygens' collected works, *Oeuvres Complètes*, Vol. 18, Société Hollandaise des Sciences, ed., Nijhoff, The Hague, 1938.—*Transl.*

[5]*The Decline of the West*, Vol. 1, translated by Charles F. Atkinson, Alfred A. Knopf, New York, 1926, pp. 14–15.

Several methods had been invented for transforming the accelerated fall of a weight into the uniform motion of a pointer, but even Tycho Brahe's astronomical clock, known for its accuracy, had to be "adjusted" with a hammer every day. Not a single mechanical phenomenon was known that would periodically repeat itself in a relatively small, fixed amount of time.

Pendulum Clocks

Such a phenomenon was discovered at the dawn of Galileo's creation of the new mechanics. Namely, Galileo discovered that the oscillations of a pendulum are isochronous, i.e., the period, in particular, does not change as the oscillations are damped. We have earlier described Viviani's story of Galileo's discovery.

Galileo proposed to use a pendulum to make a clock. In his letter of June 5, 1636 to the Dutch admiral Laurens Reael, he wrote of combining a pendulum with an oscillation counter. However, he began work on making a clock in 1641, a year before his death, and did not finish. His work was supposed to have been continued by his son Vincenzio, who was slow in renewing it and began only in 1649. This was not long before Vincenzio's death, so he too did not succeed in making a clock. Various scientists had already used the isochronous property of the pendulum in laboratory experiments, but it was not an easy path from there to creating a pendulum clock.

This was accomplished in 1657 by the twenty-seven-year-old Christiaan Huygens, already a well-known scientist because of his discovery of Saturn's rings. On January 12, 1657, he wrote: "During these days I have found a new way of making clocks, with which time can be measured so precisely that there is no little hope of being able to use it to measure longitude, even if this is to be done at sea." The first example of a pendulum clock was made by the Hague watchmaker Salomon Coster, and on June 16th the States General of Holland issued a patent, strengthening Huygens' priority. In 1658, he published a description of his invention in *Horologium* (*The Clock*).

Learning of Huygens' clock, Galileo's students undertook an energetic attempt to establish their teacher's priority. In order to appreciate the situation, it is important to understand that in the 17th century the problem of making accurate clocks was associated, first of all, with the possibility of using them to measure longitude aboard ship. Galileo understood this, and Huygens put it foremost from the very beginning (as the above quotation shows).

We have already discussed the problem of measuring longitude. Gali-

leo's students knew that at the end of his life he had carried out secret negotiations with the States General, proposing his method for measuring longitude. The contents of the negotiations, interrupted after the Florentine Inquisitor's interference, were not reliably known. One could assume that they included using pendulum clocks. Recall that the idea for this method consists of the fact that clocks "remember" the time at the port of departure, and the difference between this time and local time aboard ship determines the difference in longitude. It was important that a clock keep time correctly when being tossed about by the waves. A pendulum's isochronous oscillations would have been essential because of both damped oscillations and rough seas.

Galileo suggested to Holland another way to measure longitude, based on observations of the eclipses of Jupiter's moons. Although pendulum clocks may have been mentioned in the negotiations (the letter to Reael), undoubtedly no design for clocks or detailed information about them was given to Holland. When Galileo began working on making a clock (1641), the negotiations with the States General had practically ceased.

Huygens was not accused of plagiarism, although people may have been aware that pendulum clocks were being made in Holland, by the son of an influential member of the Council of State who had been involved in the negotiations with Galileo. Leopoldo de' Medici wrote a letter to the French astronomer Ismael Boulliau (1605–1694), who protected Huygens, with a commission to make a working mechanism following Galileo's idea. Viviani's story, mentioned earlier, and drawings of Galileo's clock were attached to the letter, to be given to Huygens. Huygens, familiarizing himself with the drawings, said that the basic idea was there but not its technical realization. In 1673, Huygens wrote: "...if they say that [Galileo] tried to find this machine but without being able to reach his goal, it seems to me that they diminish his glory more than mine, since in this case I searched for the same thing as he with more success."[6] Here it is unnecessary to recall that Galileo worked with clocks when he was blind and more than 50 years older than Huygens was when he worked on the same problem.

Huygens' first clocks mostly employed the design of those in use at the time (he had in mind being able quickly to remake the clocks he already had into pendulum clocks). From that time on, perfecting clocks became one of Huygens' chief concerns. His last work on clocks was published in 1693, two years before his death. If in his first work Huygens appeared most of all as an engineer, knowing how to realize the already-known isochronous property of the pendulum in a clock mechanism, then gradually Huygens

[6] *Oeuvres*, Vol. 18, p. 90.

the physicist and mathematician came to the foreground.

Incidentally, his engineering achievements were outstanding in number. Max von Laue[7] highlighted the idea of feedback in Huygens' clocks: The initial energy is communicated to the pendulum without altering its period of oscillation, "and then the very source of the oscillations determines the moments of time when additional energy is required." For Huygens this role was fulfilled by a simple and clever construction in the form of an armature with teeth cut slantwise, rhythmically nudging the pendulum.

While still beginning his work, Huygens discovered an inaccuracy in Galileo's assertion that the oscillations are isochronous. A pendulum has this property only for small angles of deviation from the vertical, but not, say, for an angle of 60° (Galileo may have considered this in the experiments described by Viviani). In 1673 Huygens noted that the ratio of the period for 90° to that for small angles is $\frac{34}{29}$. In order to compensate for deviations from isochronicity, Huygens decided to decrease the length of the pendulum when the angle was increased. For this purpose, in his first clocks he employed restrictors in the form of "cheeks" on which he partially wound a string attached to a weight. Huygens did not establish an empirical method for choosing the form of the cheeks. In 1658 he removed them completely from the design, introducing instead restrictions on the amplitude. But this did not mean he had stopped searching for an isochronous pendulum. Correcting discs appeared again in the clocks of 1659, but by this time Huygens was able to determine the form of the cheeks theoretically: it turned out that they had to take the form of a *cycloid*, a curve that played a major role in the development of seventeenth-century mathematics.

The next chapter of our book is wholly devoted to this curve. There, the reader will be able to learn just how Huygens arrived at his discovery.

Huygens attributed the greatest significance to the invention of the cycloid pendulum: "[To prove this] it was first of all necessary to corroborate and amplify the doctrine of the great Galileo on falling bodies, a doctrine whose most desirable fruit and highest peak, as it were, is the property of the cycloid that we have discovered."[8]

Centrifugal Force and a Clock with a Conical Pendulum

The cycloid pendulum was not Huygens' only invention in the course of perfecting the clock. Another direction of his work in chronometry is associated with the theory of centrifugal force. Huygens created this theory

[7] The German physicist Max von Laue (1879–1960) won the Nobel Prize for his 1912 discovery of X-ray crystallography.—*Transl.*

[8] *Oeuvres*, Vol. 18, p. 88.

and it is significant that he first published it in *Horologium Oscillatorium*. In the fifth section of this book, he presents, without proof, theorems on centrifugal force and describes the design of a clock with a conical pendulum (it is known that Huygens invented such a clock on October 5, 1659). The proofs of the theorems are contained in *Di Vi Centrifuga* (*On Centrifugal Force*), which was written in 1659 but came to light only eight years after Huygens' death. Aristotle had known about centrifugal force, and Ptolemy assumed that if the earth rotated on its axis then, because of centrifugal force, objects could not remain on its surface. Kepler and Galileo refuted this point of view, explaining that in this case weight counterbalances centrifugal force, essentially proposing that centrifugal force decreases with the distance from the center of rotation. However only Huygens obtained the remarkable formula for centrifugal force, $F_{\text{cf}} = \frac{mv^2}{R}$, to which Galileo had come very close. We present Huygens' original text in an appendix and the reader can see in what form (perhaps not the most economical from today's point of view) Huygens first reported his results.

Regardless of what problem Huygens studied, he always thought about possible applications of his results to clocks. And here he wanted to use the conical pendulum. A conical pendulum is a string with a weight that revolves around an axis through its point of suspension. Let l be the length of the string, α be the angle between the string and the vertical, and R be the distance from the weight to the axis. If the pendulum moves in a circle and the angle α remains constant, then $\frac{mv^2}{R} = mg\tan\alpha$. Thus $v = \sqrt{gR\tan\alpha}$. For the period, the time for one revolution, we obtain (since $T = \frac{2\pi R}{v}$)

$$T = 2\pi\sqrt{\frac{R}{g}\cot\alpha} = 2\pi\sqrt{\frac{l\cos\alpha}{g}} = 2\pi\sqrt{\frac{u}{g}}.$$

Here $u = l\cos\alpha$ is the length of the projection of the string onto the axis of the pendulum.

Huygens' text extensively discusses the formula for the period of a conical pendulum. The motion of a conical pendulum is compared to two motions that had been studied thoroughly by that time: free fall and the oscillations of a simple (or mathematical) pendulum (Huygens called the latter lateral oscillations, as opposed to the circular oscillations of a conical pendulum).

Thus, the period is determined by the projections of the string onto the axis. The difficulty of making an isochronous conical pendulum lies in the fact that its angle with the axis gradually decreases and the period increases. Huygens calculated that for the period to remain constant as the angle decreases, the length of the string must decrease so that its end lies

on a paraboloid of revolution.

Indeed, suppose we have a surface of revolution. (Huygens took a paraboloid, the surface of revolution of the parabola $py = x^2$ around the y-axis). A point mass revolves stably around a horizontal cross section (a circle) if the sum of the gravitational and centrifugal forces is directed along the normal to the surface (perpendicular to the tangent plane), and thus the formula for a conical pendulum can be applied. In this case, α is the angle of the normal to the axis, l is the length of the section of the normal between the axis and the surface, and u is the projection of this segment onto the axis. The passage here from a conical pendulum to the revolution of a point mass is somewhat analogous to Galileo's passing from a mathematical pendulum to the motion of a point mass along a circular trough. Here Huygens remarks that for the parabola $py = x^2$ the quantity u (the projection of the section of the normal onto the axis) is independent of the location of the point and equals $\frac{p}{2}$. From this he draws the conclusion that the period of revolution of a point mass along any horizontal section of the paraboloid is the same:

$$T = 2\pi\sqrt{\frac{p}{2g}}.$$

This gives a new way to obtain isochronous oscillations which, in Huy-

gens' opinion, was important for making clocks. If we suspend a conical pendulum so that its end moves along the surface of the paraboloid obtained by revolving the parabola $py = x^2$, independent of the angle of inclination α of the string to the axis, then the period of revolution will not depend on α. In other words, we must arrange things so that when α changes, the length l changes in a way that guarantees that the projection u onto the axis remains constant. Huygens thought of an extremely clever method of suspension. He proposed to make a plate in the form of a semicubical parabola $y^2 = ax^3 + b$ and to attach one end of the string at some point of the plate. It then turns out that we can choose a, b, and the length of the string so that no matter how we stretch the string and wind part of it onto the plate, its other end will be on the parabola. The secret of this clever method of suspension relies on the same mathematical considerations as the method for supporting a cycloid pendulum.

We note that in 1687 these calculations helped Huygens quickly solve Leibniz's problem about a curve along which a point mass moves so that the segments it traverses in equal time intervals have equal projections onto a vertical line. The semicubical parabola has this property.

The Physical Pendulum

One of Huygens' major achievements involves the theory of the physical pendulum, i.e., not the oscillations of a point mass but those of a configuration of weights or of a plate. This problem arose in connection with the idea of having, besides the basic weight at the end of a pendulum, a moving weight that allowed its period to be regulated. Huygens got this idea from Simon Douw, a craftsman from the Hague, who took out a patent in 1658 on his version of the pendulum clock, differing only slightly from Huygens'. Problems on the oscillations of a physical pendulum had arisen earlier. For mechanics, it was essential to be able to pass from the motion of a point mass to that of extended configurations. The first series of such problems involved the center of gravity, and here important results were known. But for a long time, no real progress had been made on problems about the oscillations of a physical pendulum.[9]

Huygens learned about problems involving the physical pendulum from Mersenne: "When I was still practically a child [less than seventeen—S.G.], the very scholarly Mersenne once suggested to me, as to many others,

[9]Recall that the effective length of a physical pendulum is the length of the mathematical pendulum that oscillates with the same period about a point on the line between the point of suspension and the center of gravity. The distance from the center of the oscillation to the suspension point is the effective length.

the study of the center of oscillation or perturbation. That was then a famous problem among the geometers of the time, as I conclude from the letters he wrote to me as well as the recently published writings of Descartes, which contain a response to Mersenne's letter on this subject.... At the same time he promised me a great and enviable reward for my work if by chance I managed to satisfy his request. But he did not receive what he wanted from anyone. As for me, as I found nothing that opened the way towards this contemplation but was turned away even at the entrance, I refrained from further study. Those who had hoped to succeed, illustrious men such as Descartes, Honoré Fabry, and others, did not achieve their goal at all, except in the easiest cases, or else they gave, in my opinion, no valid proof.... The manner of adjusting the pendulum of our *automate* [clock] by applying in addition to the lower weight, a moveable weight as explained in the description of the clock, gave us an occasion to undertake this research again. Taking up the question under better conditions and from the beginning, I finally triumphed over all difficulties and solved not only Mersenne's problems, but also more difficult ones; I even found at last a general method for finding the center [of oscillations] of lines, surfaces, and solid bodies. From this I had, beyond the pleasure of finding what others had so long searched for and of learning the laws and decrees of nature in these matters, the advantage of knowing henceforth an easy and sure method for adjusting a clock. A second result which seems to me the most important is that I can, based on this theory, give a very precise definition of length, well defined and invariable over the centuries...."[10]

This last idea of Huygens is that, just as the day is a natural unit for measuring time, a unit for measuring length should be $\frac{1}{3}$ the length of a pendulum whose period is one second.

Problems on the center of oscillation were beyond the reach of the methods of mathematical analysis that had been worked out at the time. Huygens noted that a whole series of difficulties could be overcome by beginning with energy considerations: A moving center of gravity cannot be raised higher than it was at the beginning (otherwise there would be perpetual motion). This method of proof drew comments from many leading scientists, and a great deal of effort was spent on it before Jakob Bernoulli succeeded in establishing analogous results by other means.

Maritime Clocks

The year 1673 was the acme of Huygens' work on pendulum clocks. *Horologium Oscillatorium* came out that year, and the Parisian clockmaker Isaac

[10] *Oeuvres*, Vol. 18, pp. 242–244.

Thuret made a model of a clock that incorporated every improvement. Pendulum clocks firmly caught on, but hopes for a maritime pendulum clock were unwarranted. The first models of such clocks had been made in 1661, and sea trials began in 1663. First, Count Alexander Bruce took a clock with him on a voyage from Holland to London, but the clock was slow; Captain Robert Holmes' experiments in sailing from London to Lisbon were more successful. In *Horologium Oscillatorium*, Huygens describes dramatic events involving clock experiments during the English fleet's voyage to Guinea. Experiments with varying success took place until 1687, although it had become clear that pendulum clocks did not give the hoped-for way to measure longitude. The demand for a maritime clock gradually subsided, and in 1679 Huygens himself was inclined to think that a spring clock with a balance wheel would have to serve as a maritime chronometer. In 1735, John Harrison succeeded in making such a chronometer, and received a prize of 20,000 pounds from the English government.

Three hundred years have passed. People have been well served by pendulum clocks, although they have rarely known the name of the man who invented them. The dramatic story of Huygens' work is very instructive. In some sense, his chief ambitions were not realized: He never succeeded in making a maritime chronometer, and the cycloid pendulum, which Huygens considered to be his principal invention, did not survive in clocks used on land (amplitude restrictors were quite sufficient). The conical pendulum suffered the same fate. But his mathematical and physical results that were motivated by problems on perfecting clocks have lasted to this day in infinitesimal analysis, differential geometry, and mechanics, and one cannot overestimate their significance.

Appendix

Part Five of Horologium Oscillatorium

Containing Another Construction Based on the Circular Motion of Pendulums, and Theorems on Centrifugal Force[11]

. . .At first, I intended to publish a description of these clocks with the theory of circular motion and centrifugal force—as I wish to call it—a subject on which I had more say than I had time for at the moment. But so that those interested in these matters could sooner enjoy this new and in no way useless theory and so that publication would not be hindered accidentally,

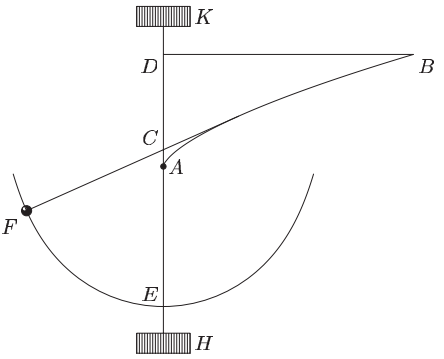
[11] Ibid., pp. 360–367.

I added this part to the others, against my plans. I briefly describe the construction of the device and at the same time state the relative theorems on centrifugal force, saving their proofs for later.

Construction of the Second Clock

I have not deemed it necessary to set out here the disposition of the gears that form the interior of the clock, since they can easily be arranged to suit the craftsmen and the disposition can be changed in various ways. It suffices to explain the part that controls the movement in a well-determined way.

The following figure represents this part of the clock. We must imagine the axis *DH* as being vertical and moveable on two poles. A rather large curved plate is attached to it at *A*; it is curved along *AB*, which is the [semicubical parabola] whose evolution describes a parabola, after a certain line has been adjoined to it, as we have proved in Proposition VIII of Part Three. This line is here *AE*, and the curve *EF* represents the parabola described by the evolution of the entire curve *BAE*. The string applied to *BA* and whose end describes the parabola is *BCF*. The weight attached to the string is *F*. But, while the axis *DH* rotates, the stretched string *BCF* makes the ball *F* move so that it travels in horizontal circles, which are larger or smaller according to the force placed on *DH* by the clock gears acting on the small membrane *K*. But these circles lie on the surface of a parabolic conoid, and by this means the revolution times will always be equal, as will follow by what we say below about this motion.



If we wish the clock to show half-seconds, the latus rectum of the parabola *EF* must be $4\frac{1}{2}$ inches of our hour-foot, i.e., half the length of the pendulum whose simple oscillation take half a second. But the length of the latus rectum of the [semicubical parabola] *AB* depends on that of

the parabola: The former equals $\frac{27}{16}$ times the latter. Similarly, the length of AE is half that of the latus rectum of the parabola. But if we wish each revolution to last one second, the latera recta and AE must be four times as large as before. . . .

Theorems on the Centrifugal Force Resulting from Circular Motion[12]

I. If two identical bodies travel unequal circumferences in equal times, the centrifugal force corresponding to the larger circumference is to that of the smaller as the circumferences themselves or their diameters.

II. If two identical bodies move with the same speed around unequal circumferences, their centrifugal forces will be inversely proportional to the diameters.

III. If two identical bodies move around equal circumferences with unequal speeds, each of which is constant, as we assume throughout, then the centrifugal force of the faster one is to that of the slower one as the squares of the speeds.

IV. If two identical bodies moving around unequal circumferences have the same centrifugal force, the time for one revolution around the larger circumference is to that around the smaller as the square roots of the diameters.

V. If a body moves around the circumference of a circle with the speed it acquires in falling from a height of one-fourth the diameter, it will have a centrifugal force equal to its weight, i.e., it will pull the string attaching it to the center with the same force as if it were suspended.

[If the height is $H = \frac{R}{2}$, then the final speed for free fall is $v = \sqrt{2gH} = \sqrt{Rg}$ and the centrifugal force is $F = \frac{mv^2}{R} = \frac{mRg}{R} = mg$.]

VI. On the concave surface of a parabolic conoid [paraboloid] with vertical axis, all revolutions of a body moving around horizontal circumferences, small or large, are accomplished in equal times. Each of these times equals that of one double oscillation of a pendulum whose length is half the latus rectum of the generating parabola. . . .

[12]Comments on the text are given in square brackets, using the following notation: m is the mass of the body, F the centrifugal force, T the period, R the distance from the center, and v the speed.

Secrets of the Cycloid

> The roulette is so common a curve that after the line and circle no other is so often met; it is described so often before everyone's eyes that it is surprising it was not at all considered by the ancients... because it is nothing other than the path made through the air by the nail on a wheel, rolling as usual, from when the nail begins to leave the ground until the continuous rolling of the wheel has brought it back to the ground after a complete revolution. Pascal[1]

I The Cycloid and the Isochronous Pendulum

The curve "described so often before everyone's eyes" was first noticed by Galileo in Italy and Marin Mersenne (1588–1648) in France. It was called a cycloid in Italy and a roulette in France. The term cycloid means "coming from a circle" and is due to Galileo. It prevailed, and the term roulette now denotes a more general kind of curve that we will discuss later. Seventeenth-century mathematicians, creating general methods of studying curves, were very interested in new "experimental" curves. The cycloid occupied a special place among these curves. It turned out to be one of the first *transcendental* curves (curves that cannot be derived algebraically), for which the problems of constructing tangents and calculating areas were solved beautifully and explicitly. But what was most striking was that the cycloid appeared again and again in the solution of very different problems where it was not a part of the original formulation. All this made the cycloid the most popular curve of the 17th century: the most powerful scientists in Italy and France (Torricelli, Vincenzo Viviani (1622–

[1] *Histoire de la Roulette*, in *Oeuvres Complètes*, Jacques Chevalier, ed., Gallimard, Paris, 1954, p. 194.

1703), Pierre de Fermat (1601–1665), René Descartes (1596–1650), Gilles de Roberval (1602–1675)) solved a variety of problems about the cycloid, and in 1673 Huygens said that "the cycloid is studied more carefully and thoroughly than all other curves."

From a Kinematic Definition to an Analytical One

The kinematic definition of the cycloid is found in the epigraph that begins this chapter. Let us try to interpret it.

Choose a system of coordinates in the plane so that the line along which the circle rolls (the *direction* line) coincides with the x-axis, and let the circle (called the *generating* circle) roll in the positive x direction. Suppose that at time $t = 0$ the point that we are observing on the boundary of the circle is at position $A_0 = (0, 0)$ (Figure 1).

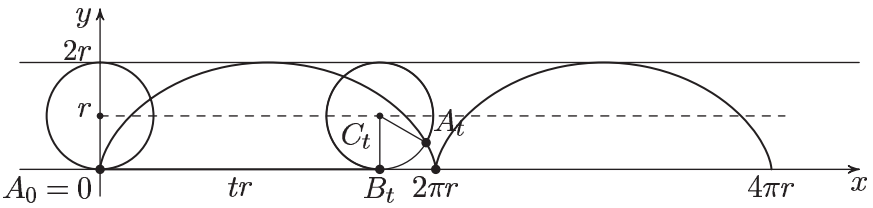


Figure 1.

If r is the radius of the generating circle, then the center of the circle of the circle will move along the line $y = r$. In order to characterize completely how the circle rolls, it suffices to describe the motion of its center if we just make the additional assumption that *the circle rolls without sliding*.[2] It will be convenient to choose the unit for measuring time so that the center of the circle moves uniformly with speed r. At time t the center of the circle is then at the point $C_t = (tr, r)$ and the generating circle will be tangent to the direction line at the point $B_t = (tr, 0)$. Let us first find the position A_t of the observed point at time t (by the definition of A_t this is a point on the cycloid). In order to do this, we need to give a precise formulation of the condition that the circle rolls without sliding: it is that the length of the segment between the points where the generating circle is tangent to the direction line at times 0 and t (the segment OB_t in Figure 1) equals the length of the arc B_tA_t "rolled out" by this segment (here the arc can exceed a complete circle). Therefore, at time t angle $B_tC_tA_t$ equals t radians, since the length of arc B_tA_t equals tr. Let D_t denote the projection of A_t onto the

[2]This is probably what Pascal meant when he wrote that the wheel is "rolling as usual."

line passing through the center C_t and parallel to the x-axis, and let E_t be the projection of A_t on the line through C_t and parallel to the y-axis (Figure 2).

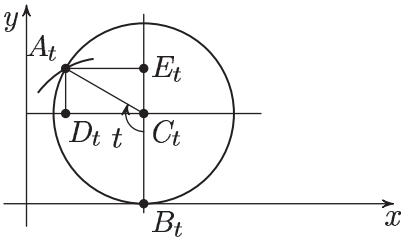


Figure 2.

Taking into account the directions of the axes we obtain

$$C_tD_t = -r\sin t, C_tE_t = -r\cos t$$

(check what will happen when $t > \frac{\pi}{2}, t > \pi$). Therefore, the coordinates of A_t on the cycloid are correspondingly equal to

$$x = rt - r\sin t, y = r - r\cos t.$$

Note that when $t = 2\pi$ the length of segment OB turns out to be equal to the circumference of the circle, the observed point again falls on the x-axis, and the picture begins to repeat itself. So the period of the cycloid equals $2\pi r$.

Thus, we can define the cycloid as the set of points with coordinates $(rt - r\sin t, r - r\cos t)$ and agree to forget about the original kinematic definition. We have found what is called a parametric representation of the cycloid: the x- and y-coordinates of a point A_t on the cycloid are functions of some auxiliary parameter t.

We will call the points of the cycloid that lie on the x-axis its *cusps*, the points lying on the line $y = 2r$ its *vertices*, and the sections of the cycloid between adjacent cusps its *arches*. In the coordinate system we have chosen the cycloid is characterized by the single parameter r (the radius of the generating circle). All cycloids for which $(0, 0)$ is a cusp can be obtained from one another by a homothetic transformation. For each point (x, y), $x \neq 0$, we can choose r uniquely so that this point will lie on the first arch of the corresponding cycloid starting at $(0, 0)$ (prove this).

Tangents to the Cycloid

We will construct the tangent to the cycloid using a method worked out by Torricelli and Roberval, based on the addition of velocities. Viviani was

probably the first to construct the tangent to the cycloid. However, since the cycloid was defined kinematically, it would have been natural to find a method for constructing its tangent based on kinematic considerations. This was indeed done by Roberval and Torricelli.

Consider the motion of a point mass. If the forces acting on it cease at time t_0, then the point will either stop moving or will begin to move uniformly along the tangent to its trajectory (the velocity of this uniform motion is called the instantaneous velocity vector of the original motion at $t = t_0$). This assertion follows from Newton's laws. But we may, as mathematicians often did in the 17th century, take it as the kinematic definition of the tangent, being convinced that it agrees with the observations of the simplest motions (and foremost with rotary motion). Taking this point of view, we can construct tangents to many interesting curves, using only the most simple facts about velocity.

We will consider only motion in a plane. Fix a point O in the plane, the origin from which we begin. If a moving point is at position A_t at time t, then we let $\mathbf{r}(t)$ denote the vector $\overrightarrow{OA_t}$. The motion is completely defined by the vectors $\mathbf{r}(t)$ for all values of t. We denote the instantaneous velocity vector at time t by $\dot{\mathbf{r}}(t)$. Recall that $\dot{\mathbf{r}}(t)$ is directed along the tangent to the path of motion. Its length $|\dot{\mathbf{r}}(t)|$ is called the speed. If the motion occurs along a coordinate line, then the vectors $\mathbf{r}(t)$ and $\dot{\mathbf{r}}(t)$ are directed along the line and we can describe their coordinates by $s(t)$ and $\dot{s}(t)$.

Example 1. Galileo showed that for the rectilinear motion $s(t) = \frac{gt^2}{2}$, the velocity will be $\dot{s}(t) = gt$.

Example 2. Let a point revolve uniformly about O at a distance R. Then the vector $\dot{\mathbf{r}}(t)$ is directed along the tangent to the circle along which the point moves, and $|\dot{\mathbf{r}}(t)| = 2\frac{\pi R}{T}$, where T is the period of revolution (the time for a full orbit). In particular, when $T = 2\pi$, we have $|\mathbf{r}(t)| = |\dot{\mathbf{r}}(t)| = R$.

The Law of Addition of Velocities

Suppose we have two motions $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$. We call their sum the motion for which $\mathbf{r}(t) = \mathbf{r}_1(t) + \mathbf{r}_2(t)$, where the right side is a vector sum. The law of addition of velocities asserts that the velocity $\dot{\mathbf{r}}(t)$ equals $\dot{\mathbf{r}}_1(t) + \dot{\mathbf{r}}_2(t)$, that is the vector sum of the velocities of the component motions. It is easy to establish the law of addition of velocities for the sum of motions with constant velocities. We obtain the general case from this special case by passing to limits.

Each motion $\mathbf{r}(t)$ can be represented as the sum of two rectilinear motions. For this it suffices to introduce any Cartesian coordinate system

such that $O = (0, 0)$ and to consider the change over time of the coordinates $x(t)$, $y(t)$ of the vector $\mathbf{r}(t)$. Obviously, the original motion $\mathbf{r}(t)$ is the sum of the motions $x(t)$ and $y(t)$ along the coordinate axes. The velocities of $x(t)$ and $y(t)$ are the components of $\dot{\mathbf{r}}(t)$, by the law of addition of velocities. In Example 2, if $R = 1, T = 2\pi$ and the vector $\mathbf{r}(0)$ is directed along the positive x-axis, then $\mathbf{r}(t) = (\cos t, \sin t)$ and $\dot{\mathbf{r}}(t) = (-\sin t, \cos t)$. Thus if $s(t) = \cos t$, then $\dot{s}(t) = -\sin t$, and if $s(t) = \sin t$, then $\dot{s}(t) = \cos t$. The reader acquainted with differentiation will note that this gives a very simple kinematic meaning to the formulas for the derivatives of $\sin t$ and $\cos t$.

Kinematic Definition of the Tangent to a Parabola

Galileo (1564–1642) had discovered that if a body is thrown at an angle to the horizontal, then it flies along a parabola. To prove this fact Galileo started with the assumption that such a motion is the sum of a uniform motion due to inertia and the motion of free fall. However Galileo did not use his calculations when he constructed the tangent to a parabola. That was done by Torricelli. This result is formulated in Problems 1 and 2 below.

Problem 1. Prove that when a body is thrown horizontally, the tangent to its trajectory at $A_t = (x(t), y(t)) = (vt, -\frac{gt^2}{2})$ joins A_t to $(0, -y(t)) = (0, \frac{gt^2}{2})$.

All the calculations are easy to generalize to the case of a body thrown at an angle to the horizontal with velocity (u, v). In this case the motion is partitioned into $\{x_1(t) = ut, x_2(t) = vt\}$ with constant velocity (u, v) and free fall motion $\{x_2(t) = 0, y_2(t) = -\frac{gt^2}{2}\}$. Therefore, the resulting motion can be described as $\{x(t) = ut, y(t) = vt - \frac{gt^2}{2}\}$ (when we add vectors that begin at O, we add the coordinates of their endpoints).

Problem 2. Prove that when a body is thrown with velocity (u, v), the tangent to the parabola along which it flies connects the point of tangency $(x(t), y(t))$ to the point $(0, -y_2(t)) = (0, \frac{gt^2}{2})$.

Note that the method given by Torricelli for constructing the tangent to a parabola was known beforehand, but his kinematic interpretation was undoubtedly instructive.

Let us return to cycloids. The motion of a point describing a cycloid can be thought of as the sum of a rotation $\mathbf{r}_1(t)$ around O and a translation $\mathbf{r}_2(t)$ along a line l, where these take place in a such a way that the paths traversed turn out to be the same $(s(t))$. We can express $s(t)$ in terms of t in different ways. The nature of the motion will depend on how this is done, but the

trajectory (the cycloid) and the tangents that interest us will not change. We will take the simplest example: $s(t) = ct$. Then both motions, the rotation and translation, will be uniform with the same velocity $\dot{\mathbf{r}}_1(t) = \dot{\mathbf{r}}_2(t) = c$ (see Example 2).[3]

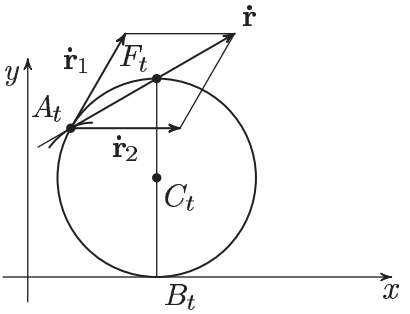


Figure 3.

We will find the velocity of the resulting motion. Suppose the point is at position A_t at time t (Figure 3). The vector $\dot{\mathbf{r}}_1(t)$ is directed along the tangent to the boundary of the generating circle, the vector $\dot{\mathbf{r}}_2(t)$ is horizontal, and they have the same length. We find the desired velocity $\dot{\mathbf{r}}(t)$, and thus the tangent to the cycloid, by the parallelogram law (a rhombus in this case).

Problem 3. Prove that the tangent to the cycloid at A_t joins this point to the point F_t of the generating circle that is highest when the circle is in this position.

To solve the problem we only have to prove a simple geometric fact: the vector $\dot{\mathbf{r}}(t)$ is directed along the line A_tF_t.

We note that the speed $|\dot{\mathbf{r}}(t)|$, the magnitude of the velocity, is not constant: it is maximal when the point is at its highest position (then the vectors $\dot{\mathbf{r}}_1(t)$ and $\dot{\mathbf{r}}_2(t)$ lie along the same line and have the same direction) and it is zero when the point falls on the line l (in this case the vectors $\dot{\mathbf{r}}_1(t)$ and $\dot{\mathbf{r}}_2(t)$ are opposites of one another). See Figure 4.

It can be shown that the property that the velocity is zero at all moments of time where the circle and line touch is equivalent to our earlier definition that the circle rolls without sliding.

Thus, we find that where the cycloid has cusps the velocity of the observed point is zero. It turns out that, in general, regardless of the trajectory

[3]We have $c = 2\frac{\pi R}{T}$, where R is the radius of the generating circle and T is the time for a complete orbit. In particular, if $T = 2\pi$, then $c = R$.

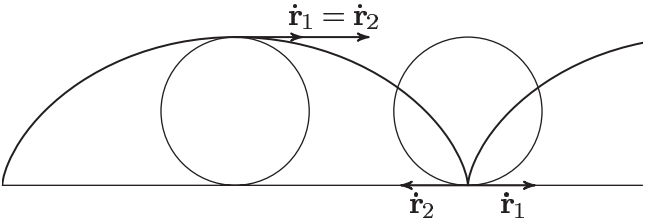


Figure 4.

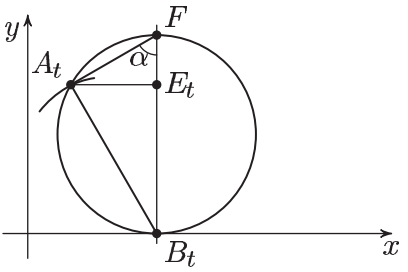


Figure 5.

the velocity is always zero at a cusp. We sometimes say that the trajectory cannot "break down" with nonzero velocity. We usually say that a curve has no tangent at a cusp. All this needs to be made more precise, but we will not do that.

Normals to the Cycloid

Thus, the tangent to the cycloid at A_t passes through the highest point of the generating circle, the point F in Figure 5. Let B_t be the lowest point of the circle and α be the angle between the tangent and the vertical FB_t. Then A_tB_t is *normal* to the cycloid, i.e., perpendicular to the tangent (an inscribed angle subtended by a diameter is always a right angle), and the y-coordinate of A_t is $y = E_tB_t = 2r\sin^2\alpha$. From this we obtain the following relation:

$$\sin\alpha = \sqrt{\frac{y}{2r}}. \tag{1}$$

This equation will play an important role in the sequel. One can show that the cycloid with parameter r is the unique curve passing through $(0, 0)$ that satisfies equation (1).

Areas of Curved Figures

Even in ancient Greece it was known how to calculate the areas of certain curved figures. At first only the quadratures of these figures were of interest, i.e., constructing with straightedge and compass a line segment whose length was equal to the area of the given figure. As was clarified later, this can be done for those figures whose areas are calculated using arithmetic operations and extracting square roots. Gradually, figures whose areas are calculated using arbitrary algebraic operations (algebraic quadrature) came to be of interest, and then even figures whose areas involved the number π. The basic method for computing areas consisted of approximating a given figure with polygons and passing to the limit. But one had to be very lucky for these calculations to lead successfully to an explicit answer.

Sometimes the calculation of an area could be simplified by using some general properties of areas. Here are some such properties:

1. Under a homothety with coefficient k the area of a figure is multiplied by k^2, and under a dilatation with coefficient k relative to some axis the area is multiplied by k.

2. Equidecomposable figures (i.e., figures that can be cut up into pairwise equal parts) have equal areas.

3. If whenever two figures are intersected with a line parallel to a fixed line we obtain equal line segments, then the figures have equal areas (this principle was formulated in 1635 by Bonaventura Cavalieri (1598–1647)).

Imagine that the contour of a figure is a flexible tape, and that the figure itself is made up of very thin, rigid fibers, parallel to a line l ("indivisible," in Cavalieri's terminology). Consider transformations that preserve these fibers but displace them relative to one another. All the figures that we obtain this way will be equivalent, by Cavalieri's principle.

The properties of area that are listed above require proof (the basic difficulty in these proofs is to give a strict definition of area), but they are easy to verify. Now we will discuss how they can be applied elegantly to compute the area under an arch of the cycloid.

A Companion of the Cycloid, Roberval's Leaves, and the Area Under the Cycloid

Since all cycloids are similar, we will restrict ourselves to the case $r = 1$. Following Roberval, we connect each point A_t on the cycloid to its

projection E_t on the vertical diameter of the generating circle (see Figure 5). Then E_t has coordinates

$$x = t, \qquad y = 1 - \cos t = 1 + \sin\left(t + \frac{\pi}{2}\right).$$

In Roberval's terminology, the "companion of the cycloid" is the curve made up of the points E_t for all possible t. It is easy to see that the "companion" is a translated sine curve (shifted one unit up and $\frac{\pi}{2}$ to the right).

An historical curiosity is associated with this. Since time immemorial, mathematicians have studied trigonometric functions but sinusoids first appeared only in the 17th century, and not as the graph of the sine function but as the "companion of the cycloid" (this can be partially explained by the fact that for a long time functions of nonalgebraic origin were not considered).

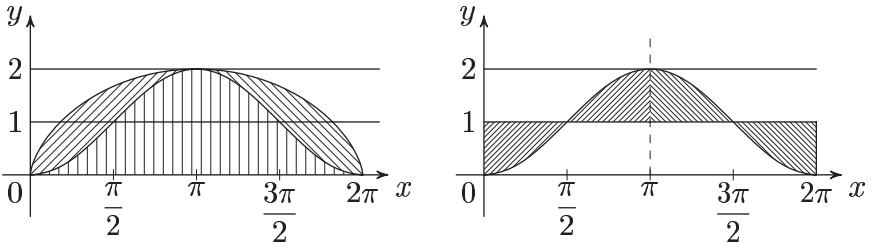


Figure 6.

The "companion of the cycloid" divides it into three parts (see Figure 6(a)): the figure under the sinusoid and two symmetric figures which we call "leaves of Roberval." By property 2, the area under the sinusoid equals 2π. This figure and the rectangle in Figure 6(b) with the same area are homogeneous. Consider one "leaf." The horizontal line at height $y = 1 - \cos t$ intersects it in the interval A_tE_t with length $|\sin t|$ (Figure 2). Moving these horizontal segments horizontally (for all possible t) so that their left endpoints fall on the same vertical line, we obtain a semicircle of the unit circle (Figure 7). By Cavalieri's principle, the area of the "leaf" equals the area of the semicircle, i.e., $\frac{\pi}{2}$. This means that the area under an arch of a cycloid with $r = 1$ equals $2\pi + 2(\frac{\pi}{2}) = 3\pi$ (and thus $3\pi r^2$ when $r \neq 1$).

The question of calculating the areas of segments of a cycloid is less elementary. Huygens wrote, not without pride, "I first measured the area of the part of the cycloid that is obtained by counting off $\frac{1}{4}$ of the axis from

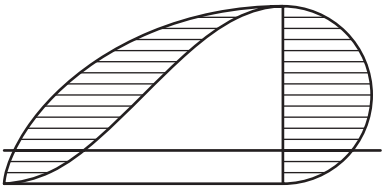


Figure 7.

the vertex and drawing a parallel to the base. This part constitutes half the area of the regular hexagon inscribed in the generating circle."

The Tautochrone

Galileo asserted that the period of oscillations of a mathematical pendulum is determined only by its length l and does not depend on the angle φ of its maximal amplitude. Huygens, explaining that this is valid only for small angles φ, decided to construct a pendulum whose period in fact did not depend on φ. Such a pendulum is called a tautochrone or isochrone.

Huygens divided the construction of an isochronous pendulum into two steps:

1. finding a curve along which the end of the pendulum (tautochrone) must move;

2. finding a way to suspend the pendulum that guarantees its end will move along the tautochrone.

We will begin with the search for a tautochrone, whose existence is not obvious in advance.

The end of a mathematical pendulum moves along a circular arc just like a point mass, along a trough whose contour coincides with the circle. If we disregard the forces of friction and air resistance, then a point mass starting with no initial velocity from height H along a circular trough will pass through its lowest point and rise again to height H. It will then oscillate periodically, climbing to height H first in one direction and then in the other. Galileo's incorrect claim was that the period $T(H)$ of oscillation does not depend on H. Our problem is to determine the form that the trough must have in order for Galileo's claim to be true.

Thanks to a happy coincidence (these are not the last word in the history of science) Huygens studied the cycloid (in connection with Pascal's contest in 1658) at the same time that he was searching for an isochronous

pendulum. And the cycloid turned out to be a tautochrone! It is likely that Huygens himself did not anticipate this. We can understand his words in this way when he wrote, "I discovered its (the cycloid's) suitability for measuring time, investigating it according to strict scientific principles and without suspecting its applicability."

Consider a trough in the form of an inverted cycloid made by a point mass (see Figure 8, where r is the radius of the generating circle). Let the point mass be at height H at the start (at the point C_0 in the sketch). We will try to find the time τ when it is at the lowest point B (the vertex of the cycloid). Then at 2τ it will be at the point $C_{2\tau}$ symmetric to C_0 with respect to the vertical axis, and at time $T = 4\tau$ (a complete period) it will return to C_0. We are interested in how τ depends on H.

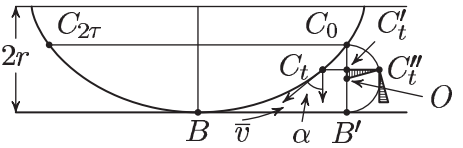


Figure 8.

At time t, let the point mass be at height $h = h(t)$ and position C_t. The velocity vector $\dot{\mathbf{r}}(t)$ at time t is directed along the tangent to the cycloid at C_t. Its length $|\dot{\mathbf{r}}(t)|$ (the magnitude of the velocity) is determined from the principle of conservation of energy:

$$\frac{m|\dot{\mathbf{r}}(t)|^2}{2} = mg(H - h(t)),$$

i.e.,

$$|\dot{\mathbf{r}}(t)| = \sqrt{2g(H - h(t))}.$$

We will look at how the projection of our point onto the vertical line C_0B' moves. At time t this projection is at position C'_t and at time τ it is at B' (see Figure 8), traversing the segment C_0B' of length H. The velocity $w(t)$ of this rectilinear motion at time t (at position C'_t in Figure 8) is the projection of the velocity vector $\dot{\mathbf{r}}(t)$ onto the vertical: $w(t) = |\dot{\mathbf{r}}(t)| \cos\alpha$, where α is the angle between the tangent to the cycloid and the vertical. Since (see (1)) $\cos\alpha = \sqrt{\frac{2r-y}{2r}}$ and $y = 2r - h(t)$, we have $\cos\alpha = \sqrt{\frac{h(t)}{2r}}$, which means

$$w(t) = \sqrt{\frac{g}{r}} \cdot \sqrt{h(t)(H - h(t))}.$$

Thus, the law by which the velocity of our rectilinear motion changes is rather complex. But Huygens noticed (a decisive guess!) that for uniform rotary motion around a circle of diameter H the vertical component of the velocity has the same form as $w(t)$. Indeed, consider the segment C_0B' to be the diameter of a semicircle and let C_t'' be the point on the semicircle at height $h(t)$. The length of $C_t'C_t''$ equals $\sqrt{h(t)(H-h(t))}$.

The shaded right triangles in Figure 8 are similar, since their sides are pairwise perpendicular (OC_t'' is a radius of the semicircle and at C_t'' we draw the tangent to the semicircle and a vertical line). It follows from this that the vector of length $(\frac{H}{2})\sqrt{\frac{g}{r}}$ that is tangent to the circle at C_t'' has a vertical projection of length $w(t)$. This means that as our point C moves along the cycloid the corresponding point C'' revolves uniformly with angular velocity $\sqrt{\frac{g}{r}}$ radians per second (independent of H!). C'' traverses the semicircle C_0B' in time $\tau = \pi\sqrt{\frac{r}{g}}$, the same time in which C' traverses the line segment C_0B', and C traverses the cycloid arc C_0B. Thus, we have not only proved that the cycloid is a tautochrone (i.e., that τ does not depend on H), but we have also found the period of the oscillations:

$$T = 4\tau = 4\pi\sqrt{\frac{r}{g}}. \tag{2}$$

We have actually shown that the motion of a point mass in a cycloidal trough can be represented as the sum of a uniform rotary motion with angular velocity independent of the initial height H of the point, together with some (in general, nonuniform) translation motion. For $H = 2r$ it is easy to deduce this from the kinematic definition of the cycloid and relation (1) on p. 99.

Equation (10) is reminiscent of Galileo's conjecture for the period of a mathematical pendulum of length l ($T = 2\pi\sqrt{\frac{l}{g}}$), so it was natural to try to use (10) to establish the latter. And indeed, Huygens used (10) to obtain the first rigorous proof of Galileo's formula for small swing angles φ. He noted that for small angles, a circular trough is almost the same as a cycloid, and it remained only to find the relation between the length l of a mathematical pendulum and the parameter r of a cycloid that minimizes this difference. This turned out to be $l = 4r$ (this is not obvious, and we will return to it later). Substituting $r = \frac{l}{4}$ into (10), we obtain the famous formula for the period of a mathematical pendulum: $T = 2\pi\sqrt{\frac{l}{g}}$ (for small φ).

The Cycloid Pendulum

In creating his first clock models, Huygens hoped to compensate for the deviation of a simple (mathematical) pendulum from isochronicity by decreasing its length as it swings. The length of a pendulum can be regulated by using "cheeks" (Figure 9(a)) on which the weight's string would be wound as it swings. Experiments to determine the needed dependence of the length of the pendulum on the angle of deflection were not successful, and in his next clock constructions Huygens used limitations on the amplitude instead of cheeks. When it became clear that the cycloid was a tautochrone, it was understood that the form of the cheeks should be such that the end of the pendulum moves along a cycloid.

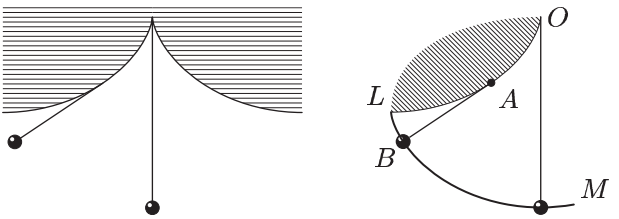


Figure 9.

Huygens searched for the form of the cheeks, reasoning roughly as follows, in a rather free retelling. Suppose we have an obstacle bounded by a curve L, and attach a stretched string of length l at some point O of L (Figure 9(b)). We wind the string around the obstacle, keeping it stretched, and observe the curve M described by its free end. Huygens called the curve M the *development*[4] of L; now M is called the *involute* or *evolvent* of L and L is called the *evolute* of M (with one evolute we associate many involutes, corresponding to different lengths l). We must find the evolute of the cycloid.

The curve M consists of those points B for which the sum of the lengths of the tangent BA to L at A and of the arc AO of L equals l (see Figure 9(b)). This corresponds precisely to stretching the string that is partially wound on L. Huygens first conjectured that the tangent to M at B is perpendicular to AB, i.e., that AB, the tangent to L at A, is simultaneously normal to M at B. The simplest way to explain this is from the kinematic definition of M. Recall that the velocity vector is tangent to the trajectory of motion and that as the action of the force changes, the velocity vector cannot change

[4]Pascal, *Oeuvres Complètes*, p. 188.

instantaneously (see below for the details). We "chop off" the obstacle at A, but continue the motion of the stretched string (Figure 10); then the end of the string begins to move along a circle with center A. Its velocity vector at B does not change, so at B the curve M and the circle with center A will have the same tangent, perpendicular to the radius BA.

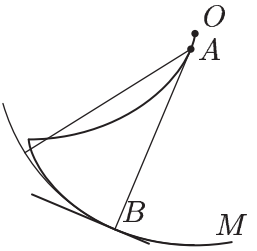


Figure 10.

When you read the section in this chapter about the roulette, you will notice that if we consider a string of any length, then the motion described for the end of the string extends to a motion of the entire plane as a rigid plate, where the points of the curve L are instantaneous centers of rotation and the different involutes are trajectories of the points of the plane. It follows immediately from this remark that AB is perpendicular to the tangent to M at B.

Huygens' next conjecture was that in a "good" situation the evolute of a curve can be reconstructed uniquely (recall that one curve has many involutes). The normals to M at various points are tangent to its evolute L. A "good" curve can be reconstructed from its tangents: taking many tangents, we construct the polygonal line they describe. Taking more tangents gives better approximations to the curve (we say the curve is the envelope of the set of its tangents).

We have to find a curve whose tangents are normal to the given cycloid. Huygens conjectured that this curve will be the same cycloid, but shifted up by $2r$ and translated by half a period. Its vertices will coincide with the cusps of the original cycloid as shown in Figure 11.

Indeed, let $r = 1$, let l, l' be the directrices of the lower and upper cycloids, respectively, and let O, O' be their points of origin (l' is two units above l and O' is π units to the right of O). We take a point C on l and consider the generating circles of the two cycloids when they are tangent to l at C. Let C', C'' be the points diametrically opposed to C on the upper and lower circles, respectively, and let A, A' be the corresponding points on the cycloids. The arc $CC''A$ is equal in length to the segment OC, so

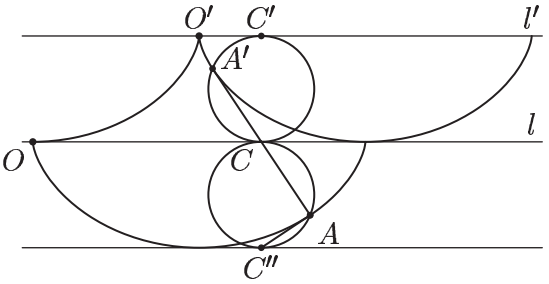


Figure 11.

it is π units longer than the arc $C'A'$, which has the same length as the segment $O'C'$. Thus $\angle C'CA' = \angle C''CA$ and the points A', C, A are colinear. It remains to note that CA' is tangent to the upper cycloid and CA is normal to the lower one (AC'' is its tangent).

Now we know that the cheeks of a tautochronous pendulum must be cycloidal, and that the length l of the string must equal $4r$ (it is precisely for this value of l that we will obtain the lower cycloid as the involute). For small amplitude angles φ the regulating cheeks will barely affect the length of the pendulum, and the cycloid will be nearly an arc of a circle of radius $4r$ (see the end of the preceding subsection).

Christopher Wren's Theorem, Evolutes, and Arc Length

Huygens did not stop after solving the problem of the cycloid pendulum, understanding that he had created a remarkable mathematical theory. He wrote, "...in order to apply this property [that the cycloid is a tautochrone] to pendulums, we have had to establish a new theory of curves, namely the theory of curves that are generated from others by evolution. This leads to comparing the lengths of curves and straight lines, which I pursued beyond what my subject required: I did it because of the beauty and apparent novelty of this theory."[5]

Huygens first noted that when the string is completely wound on a cheek, its end is at a vertex of the cycloid. This means that the length of the string ($4r$) coincides with the length of half an arch of the cycloid, so that one arch of the cycloid has length $8r$. This theorem was stated and proved in 1658 by Christopher Wren (1632–1723). Huygens, as we will see, found a very natural proof of the theorem.

[5]Huygens, *Oeuvres Complètes*, Vol. 18, p. 88.

Christopher Wren's theorem made a great impression on his contemporaries. Calculating the lengths of curves was of no less interest to mathematicians than calculating areas. At first, by analogy with quadrature (see p. 100 above), they were interested in "rectification," i.e., constructing an interval of corresponding length with straightedge and compass. Later they became interested in algebraic rectification, i.e., expressing the length in terms of any algebraic operations. We have said that the quadratures of certain figures had already been found by mathematicians in ancient times. But up to the second half of the 17th century, mathematicians had searched unsuccessfully for a curve for which even an algebraic rectification was possible. They had begun to think that there was no such curve (Descartes' words that "we, human beings, cannot find the relation between lines and curves" are sometimes interpreted in this way). Wren's rectification of the cycloid refuted this point of view. Then Fermat obtained rectifications of several other curves. However, all these examples involved nonalgebraic curves, and the skeptics "refined" the hypothesis, proposing that algebraic rectifications of algebraic curves are impossible (they correctly said that it was of course possible to construct a rectifiable curve synthetically). However, even in this form the hypothesis turned out to be false. The first counterexample to this hypothesis was constructed as early as 1657 but was not widely known. William Neile (1637–1670), Hendrick van Heuraet (1634–c.1660), and Fermat independently presented the same example of an algebraic curve admitting an algebraic rectification: the *semicubical parabola* $y^2 = ax^3$. This coincidence seemed mysterious until Huygens revealed the reason why this little-known curve has this remarkable property: it is the evolute of a parabola. More precisely, the evolute of the parabola $y = x^2$ is the curve

$$y = \frac{1}{2} + 3\left(\frac{x}{4}\right)^{2/3}.$$

Huygens' theory completely clarified the question of rectification. The results on the cycloid pendulum and related questions composed the greater part of his book *Pendulum Clocks*, which appeared in 1673.

In conclusion we present the reader with several venerable problems.

Two Problems of Galileo

1. Prove that, under the force of gravity, a mass point travels along all chords of a circle that terminate at its lowest point in the same length of time (analogously for all chords originating at the uppermost point of the circle).

2. Let L be a (sufficiently "good") curve and let A be a point not on L. Find the point B on L for which a mass point travels along the segment AB, under the force of gravity, in minimal time.

Problems of Newton

Consider a central force field in which the force is proportional to the distance r from the center: $F(r) = kr$, $k > 0$.

Newton noted that in such a field hypocycloids (see about them later in this chapter) play the same role as cycloids do in a gravitational field. Hypocycloids are tautochronous in this field (Newton called them isochrones), and evolutes of hypocycloids are similar hypocycloids. This is a purely geometrical fact, not related to mechanics, but it allows one to construct a hypocycloidal pendulum and at the same time compute the length of a hypocycloid.

Try to prove these assertions.

II Roulettes and Their Tangents

> Certain things first became clear to me by a mechanical method, although they had to be demonstrated by geometry afterwards because their investigation by the said method did not furnish an actual demonstration. But it is of course easier, when we have previously acquired by the method, some knowledge of the questions, to supply the proof than it is to find it without any previous knowledge. Archimedes[6]

Shortened Cycloids

So far, we have only followed a single, fixed, point on the boundary of the generating circle. It is clear that other points on the boundary will move along the same cycloid but will be shifted along the curve. Now we will follow the trajectories of points interior to the circle. The curves that arise are called shortened cycloids (Figure 12). They are characterized by the relation $k = \frac{\rho}{r}$, where R is the radius of the generating circle and ρ is the distance from the center of the circle to the point being observed. For $k = 0$ we obtain the line along which the center of the circle moves, and for $k = 1$ we obtain the cycloid.

[6] *The Method*, in *The Works of Archimedes*, translated by T. L. Heath, Cambridge University Press, Cambridge, UK, 1912.

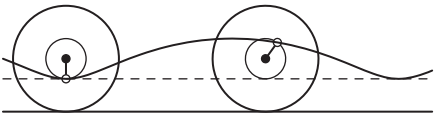


Figure 12.

Problem 4. Prove that the normal to a shortened cycloid passes through the lowest point of the generating circle.

Note that a point moving along a shortened cycloid never has zero velocity. At the lowest point the velocity vector is horizontal and its magnitude is $R - \rho$. This means that to the rolling motion of the circle of radius ρ we add a sliding motion with velocity $R - \rho$ (a translation).

Extended Cycloids

We now consider points that are exterior to the circle as it rolls (imagine that we attach a rim to a wheel moving on a rail). These points move along curves that are called extended cycloids (Figure 13). The arguments given earlier for shortened cycloids carry over verbatim to the extended case. Here, though, $k = \frac{\rho}{r} > 1$. We only note that at a lowest point of an extended cycloid the velocity is directed opposite to the motion of the circle ($|\dot{\mathbf{r}}_1| = \rho$, $\dot{\mathbf{r}}_2 = r$, $\rho > R$).

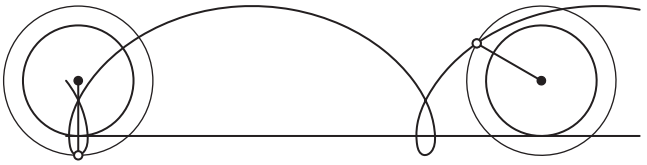


Figure 13.

Did you ever notice that the lowest points on the rim of a train wheel move backwards?

Instantaneous Center of Rotation

We have now incorporated all points in the plane as the circle rolls along the line. Each point moves along its own trajectory, but all the trajectories are coordinated since the moving points constitute a solid body. From the point of view of kinematics, the characteristic property of a solid body in motion is that the distances between its points remain the same. Here we

will limit ourselves to considering motions of solid plates that can be carried out without taking the plate out of the plane (for example, it is forbidden to turn the plate over). We will be interested in how the plate being solid leads to restrictions on the velocity of a point on the plate. Note that the motion of three-dimensional solid bodies is much more complicated than the planar problem that we consider.

Here are some of the laws governing the motions of solid plates.

Principle of Incorporation. *The motion of a solid plate is uniquely determined by the motion of any two of its points. A motion of two distinct points that preserves the distance between them can be uniquely extended to a motion of the entire plane as a solid plate.*

This assertion has a purely geometric character. We will not present its proof but will limit ourselves to visual explanations. First, the motion of a linear rod is completely characterized by the motion of two of its points, and second, if a triangle consists of rigid rods then the motion of one of them uniquely leads to the motion of the entire triangle. So from the motion of two points A, B we can incorporate the line AB and then any point C outside AB.

Principle of Inertia. *If no external forces act on a solid plate and any internal forces ensure solidity, then its motion is uniformly linear or rotational.*

In order to consider arbitrary motions of a plate we need one more fundamental principle of mechanics: *velocity cannot change instantaneously* (a change in velocity requires nonzero time). In particular, if the forces acting on a moving point change at time t_0, then the velocity $\dot{\mathbf{r}}(t_0)$ does not change, and so if $\dot{\mathbf{r}}(t_0) \neq 0$, then the tangent to the trajectory at t_0 does not change (although the trajectory itself can change starting from that moment).

Suppose that at time t_0, external forces cease to act on a moving plate. Then, on the one hand, the velocities of points at t_0 remain as before and, on the other hand, the motion should obey the above principle of inertia. Therefore, at each moment t of time, only one of two possibilities can hold for the motion of a plate:

(a) The velocities of all points are equal (as vectors).

(b) There is a unique point O_t at which the velocity is zero; at any point A of the plate the velocity is perpendicular to the vector $\overline{O_tA}$ and its magnitude is proportional to the distance from A to O_t. The proportionality coefficient depends only on t.

From the fact that the velocity cannot change instantaneously, it is not hard to deduce that we can pass from case (a) to case (b) and vice versa only at those times when the plate is at rest (the velocities of all points are zero). Therefore, in the intervals between rest times either (a) holds throughout or (b) holds throughout. In case (a), starting with the trajectory of a point B we can obtain the trajectory of any point A through a parallel translation by the vector $\overline{BA}$. Now we will consider case (b), i.e., we assume that at each moment of time there is a unique point O_t with zero velocity. We will call O_t the *instantaneous center of rotation* at time t. In the example of a circle rolling along a line the instantaneous center of rotation is the point where the circle is in contact with the line.

If the instantaneous center of rotation O_t is known, the normals to the trajectory at time t (the lines O_tA_t) and thus also the tangents, are automatically known. Conversely, if the velocity vectors of two points of the plate are known at t then by taking the point of intersection of the normals to these vectors we obtain the instantaneous center of rotation O_t.

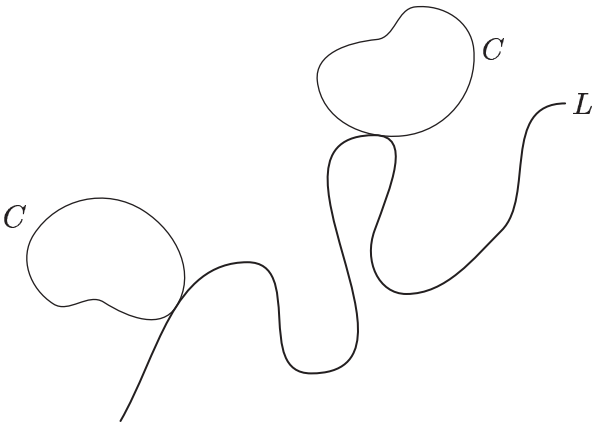


Figure 14.

Now let a solid plate move in a stationary plane. We will consider the curve L *in this plane* of instantaneous centers of rotation at all moments of time. L is called a fixed centroid, and we will call it a "rail." On the other hand, we will consider the curve C *on the plate* consisting of all points that are instantaneous centers of rotation at some moment of time. C is called a movable centroid, and we will call it a "wheel." These "nonserious" names suggest that we can obtain the original motion if we consider our "wheel" curve as rolling along the "rail" curve without sliding, and incorporate the remaining points in this rolling (Figure 14). From this we can deduce that

the arc length of the "wheel" equals the length of the corresponding arc of the "rail" guide along which the "wheel" rolls. Here we permit the "wheel" to intersect the "rail" as it rolls.

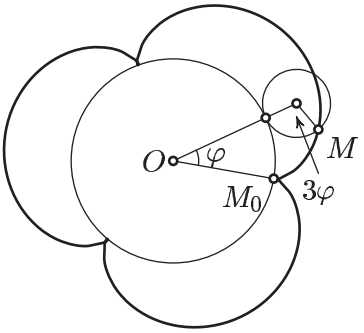
We often understand "roulette" to mean the trajectory described by the points in the plane as it moves like a solid plate satisfying condition (b) at all moments of time i.e., as it rolls. We have learned to take normals and tangents to all roulettes. Here it turns out that we do not even need to know how to take the tangent to a "wheel" and "rail" (it would have been necessary if we had used a composition of velocities). In our mechanical considerations we have left the 17th century behind; remarkably, however, the method of taking normals to a general roulette was discovered by Descartes, who defined them using oscillations (without knowing how general are the motions generated by oscillations).

Epicycloids

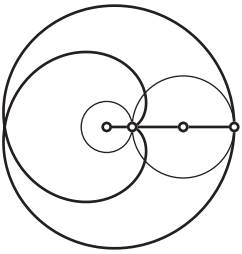
We now consider roulettes obtained when one circle rolls along another. Let a circle of radius r roll along the outside of a circle of radius R. The trajectories of the boundary points of the rolling circle (the "wheel") are called epicycloids. Their form depends on $k = \frac{R}{r}$ (see Figure 15). If k is an integer, then as the moving circle rolls once around the boundary of the fixed circle it makes k revolutions and the epicycloid will have k cusps and k arches. An epicycloid with $k = 1$ is called a cardioid, since it is reminiscent of a stylized heart. If $k = \frac{p}{q}$ is a fraction in lowest terms, then as the moving circle makes q revolutions it rolls p times around the fixed circle. If k is irrational, then there is no periodicity and the observed point never returns to its original position. One can prove that the infinite trajectory we obtain in this case forms a ring $R \le OA \le R + 2r$, approaching each of its points as closely as we wish but never falling on it.

It is easy to construct the tangents to these epicycloids using instantaneous centers of rotation, the points where the circles touch. Prove that the tangent to an epicycloid at a point passes through a point of the corresponding moving circle that is diametrically opposite the point of contact with the fixed circle.

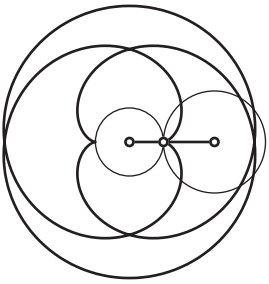
Remark. In constructing epicycloids and solving these problems we must remember the following. If A is the initial position of the point being observed (Figure 17) and at some moment of time the moving circle rolls on the fixed one at point B, then the epicycloid contains a point C on the boundary for which the arc lengths BA and BC are equal. Taking into



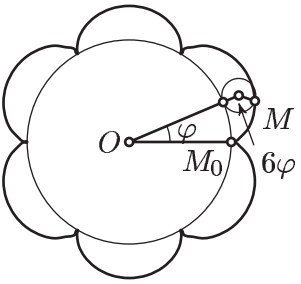
(a) $k = 3$



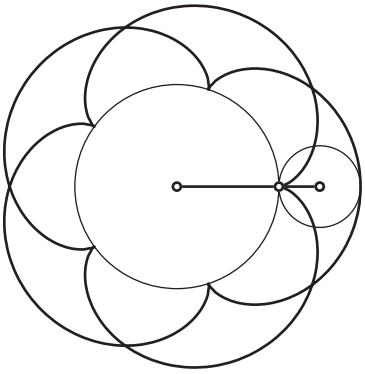
(b) $k = 6$



(c) $k = \frac{1}{2}$



(d) $k = \frac{5}{2}$



(e) $k = \frac{2}{3}$

Figure 15.

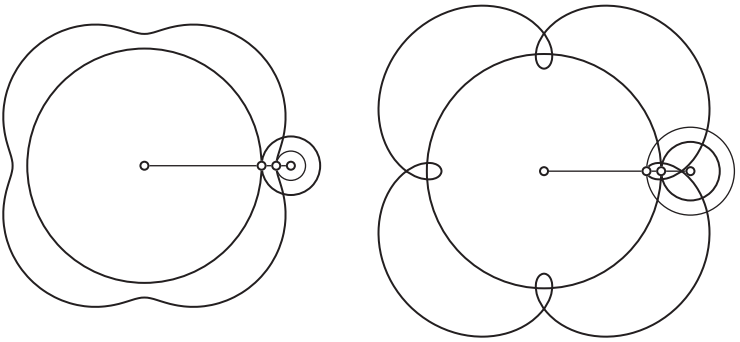


Figure 16.

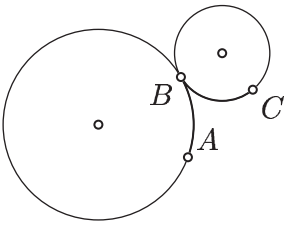


Figure 17.

account the difference in the radii, we obtain

$$\frac{\cup BC}{\cup BA} = \frac{R}{r} = k.$$

The trajectories of the interior (respectively, exterior) points of the moving circle as it rolls in this way are called shortened (respectively, extended) epicycloids. See Figure 16; we will limit ourselves to integer values of k.

Problem 5. Let A revolve uniformly around point O_1, which in turns revolves uniformly around point O. Let $OO_1 = r_2, O_1A = r_1$. Let both revolve clockwise and let v_1, v_2 be the magnitudes of their linear velocities. Show that A will move along some epicycloid (perhaps shortened or extended). What relations determine the nature of the curve?

Hypocycloids

Roulettes obtained as a circle of radius r rolls along the inside of a circle of radius $R > r$ are called *hypocycloids*. (Similarly, these may be shortened or extended.)

Analogously, we could also consider a hoop of radius R the inside of whose boundary rolls along a fixed circle of radius $r < R$. The corresponding roulettes are called pericycloids. But it turns out that they coincide with epicycloids (see the supplement at the end of this chapter).

Problem 6. Suppose the rotations described in Problem 5 take place in two opposing directions, one clockwise and the other counterclockwise. What will be the trajectory of the point A?

We have not had the goal of giving rigorous proofs of all the results we have obtained from kinematic considerations. In some cases this can be done easily: mechanical arguments are replaced by mathematical ones almost automatically (it is enough to replace velocities by derivatives). In other cases it is more complicated to find such "replacements" (for example, where we consider the motion of a plate or where forces are varying). However, purely mathematical approaches cannot replace a mechanical interpretation completely, which in many cases makes it possible to see a simple and beautiful answer.

III The Brachistochrone, or Yet Another Secret of the Cycloid

Galileo's Error

At the very beginning of the 17th century the young Galileo tried to verify experimentally his conjecture that the velocity of free fall is constant. When he brought his observations from the Tower of Pisa to his laboratory, he was very disturbed by the fact that the body fell "too quickly." In order to slow down this motion, Galileo decided to replace free fall by motion along an inclined plane, assuming that that would have constant velocity. Continuing these experiments, Galileo turned his attention to the fact that when a body slides down an inclined plane its speed at the end does not depend on the plane's angle of inclination but only on its height H, and is the same as the final speed of a body falling freely from that height (as you well know, in both cases $|\bar{v}| = 2gh$). From inclined planes, Galileo went on to considering a mass point moving under the force of gravity along a polygonal line. Comparing times for different polygonal lines that join two fixed points A and B, Galileo remarked that if we join A, B by a quarter of a circle (this can always be done—how?) and inscribe two polygonal lines M and L in it so that L is "inscribed" in M (see Figure 18), then a mass point falling from A to B falls faster along M than along L (try to

prove this). Increasing the number of line segments and passing to the limit, Galileo obtained that a mass point falls faster along the quarter circle than along any polygonal line inscribed in it. From this Galileo reached the inescapable conclusion that a quarter circle joining a pair of given points A, B (not lying on the same vertical line) is the path of fastest descent for a mass point moving under the force of gravity (later, the path of fastest descent would be called a brachistochrone). Later it would be clear that Galileo's assertion was not only unjustified but was, in fact, wrong.

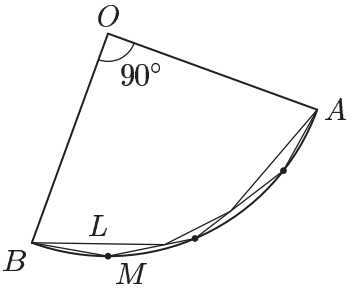


Figure 18.

Switzerland. The End of the 17th Century.

"Johann and Jacob Bernoulli, on the occasion of a conversation on mathematical topics during a walk in Basel, lighted on the question of what form a chain would take that was freely suspended and fastened at both ends. They soon and easily agreed in the view that the chain would assume that form of equilibrium at which its center of gravity lay in the lowest possible position. . . . The *physical* part of the problem is disposed of by this consideration. The determination of the curve that has the lowest center of gravity for a given length between the two points A, B, is simply a *mathematical* problem."[7]

Investigating the chain curve (the shape taken by a flexible, heavy, nonstretchable string suspended between two points), the Bernoulli brothers were interested in other problems where curves are sought that minimize some quantity or other. In 1696 Johann Bernoulli published a note entitled *A New Problem That Mathematicians Are Invited to Solve*. Incidentally, this "new" problem had already been studied by Galileo. This was the problem

[7]Ernst Mach, *The Science of Mechanics*, translated by Thomas J. McCormack, 6th ed., Open Court, La Salle, 1974, p. 85. This translation uses the English forms John and James Bernoulli.

of finding a brachistochrone, a curve joining a fixed pair of points along which a point mass descends most rapidly under the force of gravity. The brachistochrone problem, which was beyond even the great Galileo at the beginning of the century, turned out to be very timely at the end of the century. It was quickly solved by Johann Bernoulli himself and by his brother Jacob, by their teacher Leibniz, and even by Newton and L'Hôpital. We will discuss Johann Bernoulli's solution: it uses concepts from geometrical optics in a completely unexpected way.

"This solution of Johann Bernoulli's, achieved entirely without a method, the outcome of pure geometrical fancy and a skillful use of such knowledge as happened to be at his command, is one of the most remarkable and beautiful performances in the history of physical science. Johann Bernoulli was an aesthetic genius in this field. His brother Jacob's character was entirely different. Jacob was the superior of Johann in critical power, but in originality and imagination was surpassed by the latter. Jacob Bernoulli likewise solved this problem, though in less felicitous form. But, on the other hand, he did not fail to develop, with great thoroughness, a *general* method applicable to such problems. Thus, in these two brothers we find the two fundamental traits of high scientific talent separated from one another—traits, which in the very greatest natural inquirers, in Newton, for example, are combined."[8]

Fermat's Principle

As early as 140A.D., Claudius Ptolemy created a detailed table showing how the angle of refraction of a light ray passing from air to water depends on the angle of incidence, but it was only in 1621 that Willebrord Snell (1580–1626) conjectured the analytic relation connecting these angles:

$$\frac{\sin \alpha_{\text{incidence}}}{\sin \alpha_{\text{refraction}}} = k,$$

where k is the *coefficient of refraction*, a constant for each given pair of media.

In 1650 Fermat gave a remarkable interpretation of this law. It began with the fact, already known to Heron of Alexandria, that the equality of the angles of incidence and reflection can be derived from the assumption that reflected light takes the shortest path (Figure 19).

Fermat assumed that *the path along which light propagates between any two points is the path requiring the least time compared to all other paths between these points*. This assertion is now called "Fermat's principle." In particular,

[8] Ibid., pp. 522–523.

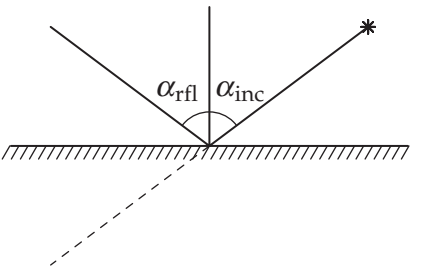


Figure 19.

from Fermat's principle it follows that if the speed of light in a homogeneous medium is constant, then the least time is achieved along a path of shortest length. This implies that the path of light in a homogeneous medium with no barrier is rectilinear, and also implies the law of reflection. If the density in the medium is variable, and the speed of light is different in different regions, then the path along which light propagates in shortest time cannot be rectilinear. We will consider what happens in the case of refraction. (Everything we discuss below is related to the planar case.)

Let the line l separate two media (in the plane), with the speed of light equal to c_1 in one and c_2 in the other; Let A_1, A_2 be points lying on different sides of l. We find a point B on l such that $\frac{\sin\alpha_1}{\sin\alpha_2} = \frac{c_1}{c_2}$, where α_1 is the angle of incidence and α_2 is the angle of refraction (Figure 20). The existence and uniqueness of B are easy to prove. Let C be any other point of l. Drop perpendiculars CE, CF onto A_1B, A_2B, respectively.

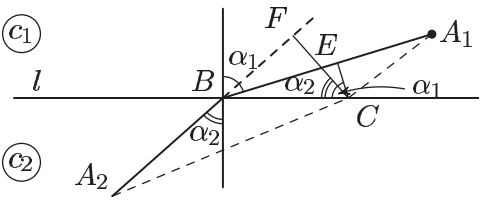


Figure 20.

Then $\angle ECB = \alpha_1$, $\angle FCB = \alpha_2$, and it takes as long to travel along BE with speed c_1 as to travel along BF with speed c_2. This means that it takes light as long to travel along the path A_1BA_2 as it does to travel along A_1E with speed c_1 and FA_2 with speed c_2. Since the segments A_1C and A_2C are longer than A_1E and FA_2, respectively, it takes light longer to travel along A_1CA_2 than along A_1BA_2. Thus the point C does not give the least time.

Thus, Fermat's principle implies Snell's law of refraction, and the coefficient of refraction of a light ray from one medium to another equals the ratio of the speeds of light in the two media.[9]

Fermat's principle also implies that in a complex layered optical medium consisting of horizontal "strips" in each of which the speed of light is constant, $c_1, c_2, \ldots$ (see Figure 21), light will propagate along a polygonal line in the plane, with vertices on the lines that divide the strips. Moreover, if α_i is the angle made with a vertical line by the segment in the region with speed of light c_i, then $\frac{\sin\alpha_j}{c_j}$ is constant over the entire polygonal line. Indeed, if $\frac{\sin\alpha_j}{c_j} \neq \frac{\sin\alpha_{j+1}}{c_{j+1}}$ for some j, then according to Fermat's principle light cannot propagate along such a polygonal line: we could move a vertex on the boundary shared by the corresponding strips without changing the others, so that in general, the time spent by the light would decrease.

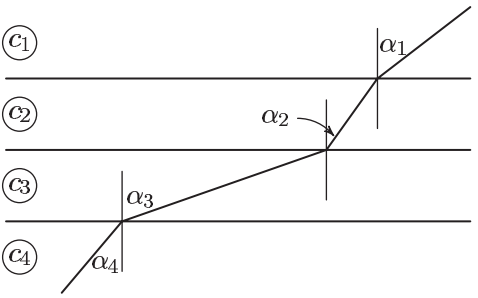


Figure 21.

Suppose the speed of light changes continuously in some inhomogeneous optical medium but is the same at all points on the horizontals, i.e., it has the same value $c(y)$ at all points with the same y-coordinate. Here $y = 0$ corresponds to the initial position of the point, i.e., the position from which the light ray emanates. Then in the limit we find that the path light takes between two points in this medium is a curve L for which

$$\frac{\sin\alpha(y)}{y} = \text{const}, \tag{3}$$

where $\alpha(y)$ denotes the angle that the tangent to L at a point with ordinate y makes with a vertical line.

In order to get to the brachistochrone problem, note that we obtained relation (3) from Fermat's principle, using only the fact that at a fixed point

[9]Fermat's principle was based on the wave theory of light constructed by Huygens in 1672–1673.

of our inhomogeneous medium the magnitude of the speed of light is fixed and does not depend on the direction in which light is being propagated (in our examples it was constant on the horizontals). But as we remarked above, for a body moving only under the force of gravity we have $|\bar{v}| = 2gh$, where y moves vertically, "losing" altitude, and we obtain that in this problem also the speed of light at each fixed point of the plane is constant and does not depend on which path is taken. Therefore, all consequences of Fermat's principle can be carried over to this case. Thus in order to fall from one given point to another in the least possible time, a point mass must move along a path L that joins these points (we assume that the points do not lie on the same vertical line) and for which

$$\frac{\sin\alpha(y)}{\sqrt{y}} = \text{const}, \tag{4}$$

where $\alpha(y)$ is the angle that the tangent to L at a point with ordinate y makes with a vertical line.

It remains only to search for a curve satisfying condition (4).

A Cycloid Again!

Seventeenth-century mathematicians were accustomed to the fact that the cycloid is the "magic wand" for answering many questions. And here it again decisively reaffirmed its "reputation"—the brachistochrone also turned out to be a cycloid!

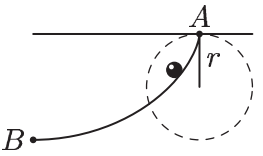


Figure 22.

Indeed, if $\alpha(y)$ denotes the angle made with the vertical by the tangent to a cycloid with parameter r at point with ordinate y, then $\sin\alpha(y) = \sqrt{\frac{y}{2r}}$ (see formula (1) on p. 99). Here the cycloid is obtained from a circle of radius r that rolls without sliding along the line $y = 0$. Moreover, as we have already noted, the cycloid is the unique curve satisfying this relation. Thus, the brachistochrone joining two given points A and B (not lying on the same vertical line) is part or all of an arch of an inverted cycloid (see Figure 22), where the "upper" point A is at a cusp of the cycloid. Since we

are only considering one (the first) arch of the cycloid, its parameter r is uniquely determined by B.

A brachistochrone can also be longer than half of a cycloidal arch. In this case, a mass point moving under gravity along a brachistochrone first moves downwards, reaches a vertex of an inverted cycloid, and then begins to move upwards. Nonetheless, such a motion turns out to take less time than if the point mass went from A to B along a straight line!

For comparison, note that although an inverted cycloid is both a tautochrone and a brachistochrone, in the first case we have to take an arch that ends at a vertex of a cycloid and in the second case one that begins at a cusp.

Some Problems

Let us return to optics. Now we know that if the speed of light in a planar inhomogeneous medium changes according to the law $c(x, y) = k\sqrt{H - y}$ (i.e., analogous to how the speed of a point mass changes as it moves under the force of gravity), then in such a medium light traveling between two points will propagate along an arc of an inverted cycloid with cusps along the line $y = H$.

Now try to solve some problems about the path light takes between two points in an optically inhomogeneous medium, given the law by which the speed of light changes in this medium.

Problem 7. The speed of light changes according to $c(x, y) = k(y-a)$. Prove that light will propagate between two points along arcs of semicircles with diameters on the line $y = a$ containing the initial point.

Problem 8. The speed of light changes according to

$$c(x, y) = \frac{k}{\sqrt{a - y}}.$$

Prove that in this case light will propagate between two points along parabolic arcs.

If in Problems 7 and 8 we interpret $c(x, y)$ as the speed of some mechanical motion, then the trajectories for the propagation of light that we obtained in solving these problems will be brachistochrones for the corresponding mechanical systems.

> The fundamental problem of mechanics consists of determining the position of a moving body at any moment of time.
>
> From a physics textbook

The Analogy Between Mechanics and Optics

Thus, in mechanics we are usually looking for the trajectory of a point mass when the forces acting on the point and the initial position and velocity vector (the initial conditions) are given. However, we can be interested not only in individual trajectories but in describing the entire set of trajectories when the forces change according to some given law. (An additional task of the initial conditions will then be to single out a specific trajectory from this set.) Then Galileo's classical result on the motion of a body that is thrown (horizontally or at some angle to the horizontal) states that in the case of the force of gravity the set of trajectories consists of parabolic arcs.

The use of optics in purely mechanical problems gave rise to the idea of trying to isolate the set of possible trajectories for a specific mechanical system using some sort of minimality condition analogous to Fermat's principle. Leibniz thought about such a condition, but its first formulation belongs to Pierre de Maupertuis (1698–1759). However, his construction was far too general and did not contain precise assertions. The first precise formulation belongs to Leonhard Euler, who had studied mathematics with Johann Bernoulli. It is related to the following special situation.

Let a point mass move in the plane under the action of some force such that its potential energy depends only on its position: $U = U(x, y)$. By the law of conservation of energy, the speed $|\bar{v}|$ of the point also depends only on (x, y):

$$|\bar{v}| = \sqrt{\frac{2}{m}(E - U(x, y))}.$$

We consider a planar inhomogeneous optical medium in which the speed of light changes according to the law $c(x, y) = \frac{k}{\bar{v}(x,y)}$. Euler's principle states that *the trajectories of light propagating in such a medium will coincide with the possible trajectories of the original mechanical system* (a point of mass m with potential energy $U(x, y)$). It is clear that Euler's principle can be formulated without mentioning the propagation of light.

In particular, Problem 8 and Euler's principle imply Galileo's assertion, stated above, about the trajectory of a mass point moving under the force of gravity.

Now let us clarify Euler's principle. For simplicity, we will limit ourselves to the case when $U(x, y)$ and thus also $|\bar{v}|$ depends only on y. Since the potential energy is constant on horizontal lines, the force will be directed vertically, the horizontal component of the acceleration will be zero, and the horizontal component of the velocity vector will be constant, i.e.,

$$|\bar{v}(y)| \sin \alpha(y) = \text{const}, \tag{5}$$

where $\alpha(y)$ is the angle between the velocity vector and the vertical when the ordinate on the trajectory is y. Equations (5) and (3) give us Euler's principle for this special case. In the general case we must take into account that the direction in which the force acts is perpendicular to the curves of constant potential energy and that, consequently, the components of the velocity vector that are tangent to these curves do not change.

In modern mechanics, principles that generalize Euler's principle (e.g., Hamilton's principle) play an exceptionally important role.

Epilogue

The heroic story of the cycloid was completed at the end of the 17th century. It arose so mysteriously in the solution of the most varied problems that no one doubted it played a quite exceptional role. The cycloid was revered for many years, but time passed and it became clear that it was not connected to the fundamental laws of nature in the same way as, say, conic sections. Problems reducing to the cycloid played a tremendous role in the establishment of mechanics and mathematical analysis, but when the greatest edifices of these fields were constructed it turned out that these problems were particular ones, far from the most important. An instructive historical illusion had taken place. However, being familiar with the instructive history of the cycloid, we can see many fundamental facts of the history of science.

Supplement

In this supplement we will explain, as promised, why the pericycloids (see p. 116) coincide with the epicycloids. Recall just what we have to prove.

Claim. *Let a hoop of radius R, hanging on a fixed circle of radius $r < R$, begin to roll without slipping along this circle. Then a point of the hoop describes the same trajectory as a point of a wheel of radius $R - r$, rolling outside the same circle of radius r* (*see Figure* 23).

Denote the wheel radius $R - r$ by ρ. Recall that curves described by points on the boundary of a wheel that is rolling as we have described are called epicycloids, and curves described by points of the hoop are pericycloids. We will prove that under the given relation among the radii, $R = r + \rho$, the pericycloids coincide with the epicycloids.

Let us focus on one point that is on the wheel and the hoop. Suppose that at the start, the points we observe on the wheel and hoop are at the same point A on the boundary of the fixed circle (Figure 24).

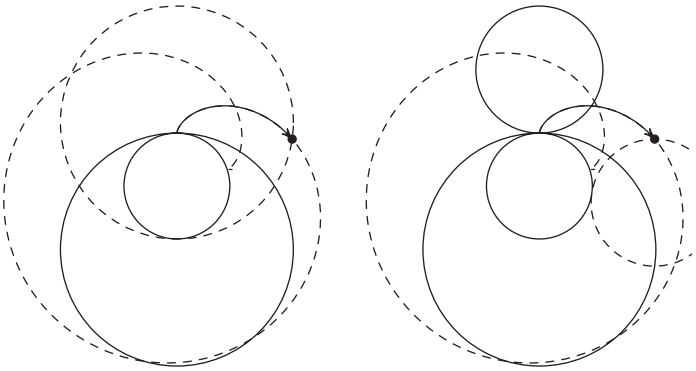


Figure 23.

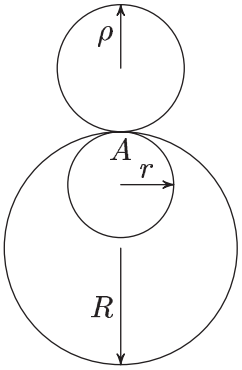


Figure 24.

For specificity, suppose the wheel and hoop roll around the circle counterclockwise. If at some moment the wheel is tangent to the fixed circle at a point B, then the point we are observing on its boundary (a point on a epicycloid) is at a point C for which the arcs AB and BC have the same length (the arc BC must take into account the direction in which the wheel rolls). See Figure 25(a).

Analogously, the position of the point C' that we observe on the hoop (a point on a pericycloid) at the moment when the hoop is tangent to the fixed circle at B' is found by setting the lengths of the arcs AB' and $B'C'$ equal, taking the direction of rolling into account. See Figure 25(b).

We will prove that for any point B on the boundary of the fixed circle we can select a point B' (also on the boundary of the fixed circle) so

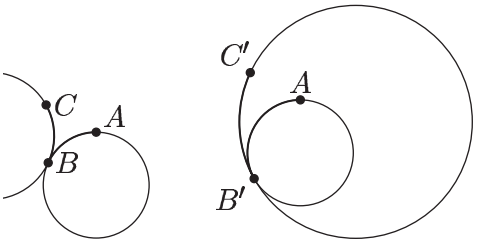


Figure 25.

that the corresponding points C (epicycloid) and C' (pericycloid) coincide (Figure 26(a)). From our proof it will also be clear how to choose B' from B.

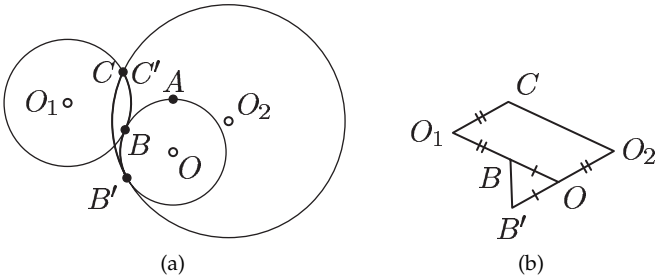


Figure 26.

We take B' so that the ratio of the lengths of the arcs AB and BB' will equal $\frac{\rho}{r}$. Then the radian measure of the arc BC will equal the radian measure of the arc BB', say φ radians. We have

$$\text{length}(AB) = \text{length}(BC) = \rho\varphi, \qquad \text{length}(BB') = r\varphi.$$

Therefore, $\text{length}(B'C') = \text{length}(AB') = r\varphi + \rho\varphi$, and the radian measure of arc $B'C'$ also equals φ. Let O be the center of the fixed circle, O_1 be the position of the center of the wheel at the moment when it is tangent to the fixed circle at B, and O_2 be the position of the center of the hoop when it is tangent to the fixed circle at B'. The points O, B, O_1 are colinear, as are O_2, O, B'.

Suppose $0 < \varphi < \pi$. We have (Figure 26(b)) $OB = OB' = r$, $O_2B' = R$, $OO_2 = R - r = \rho$, $O_B = O_1C = \rho$, $O_1O = r + \rho = R$, $\angle BOB' = \angle OO_1C = \varphi$. This means that the quadrilateral OO_1CO_2 is a parallelogram, and so $O_2C = R$, $\angle CO_2B' = \varphi$. Thus, the point C lies on the circle of radius R with center at O_2, and the radian measure of arc $B'C$ equals φ. This in turn

implies that C coincides with C'. Thus we have proved that, if arcs of a wheel and hoop with the same radian measure $\varphi < \pi$ roll around a fixed circle, then the resulting points on the epicycloid and pericycloid coincide.

It remains to be convinced that this assertion is also valid when $\varphi \geq \pi$. Let us see how Figure 26(a) changes when $\varphi = \pi$ and when $\pi < \varphi < 2\pi$. The difference in the circumferences of the hoop and wheel equals $2\pi r$, the circumference of the fixed circle. So at the moment when the wheel and hoop make complete revolutions, the points we are observing again fall on the boundary of the fixed circle at the same position, say, A_1. The case $2\pi < \varphi < 4\pi$ reduces to the case $\varphi < 2\pi$ if we take A_1 as the initial point instead of A. If we take A_1 as the initial point and also reverse the direction of rolling, then the case $\pi \leq \varphi \leq 2\pi$ reduces to the case $\varphi \leq \pi$.

Blaise Pascal

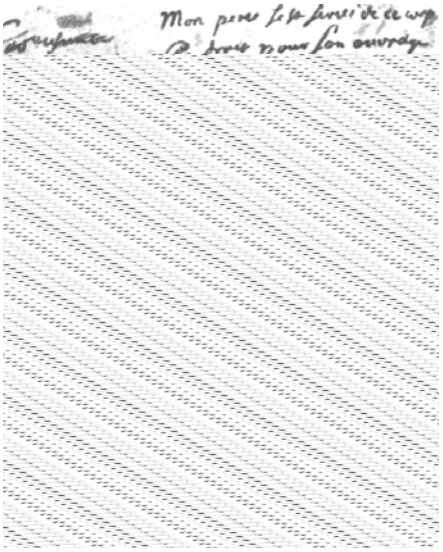
Pascal had his abyss, it followed him. Baudelaire, *The Abyss*[1]

Blaise Pascal was inherently multifaceted, a characteristic of the Renaissance that had almost become passé in the 17th century. The natural sciences (say, physics and mathematics) had not yet completely separated from the humanities, but studies in the humanities and the natural sciences were already no longer commonly combined.

Pascal entered the history of the natural sciences as a great physicist and mathematician, one of the creators of mathematical analysis, projective geometry, probability theory, computational methods, and hydrostatics. France counts him as one of its more remarkable writers: "Narrow minds are surprised by Pascal as the most perfect writer in the greatest century of the French language. . . . Each line coming out of his pen is revered as a precious stone" (Joseph Bertrand). Far from everyone agreed with Pascal's thoughts about man, his place in the universe, and the meaning of life but no one was indifferent to the lines for which the author paid with his life and which have surprisingly not aged. In 1805, Stendahl wrote, "When I read Pascal, it seems to me that I am reading myself." And 100 years later in 1910, Leo Tolstoy read "the wondrous Pascal," "a man of great mind and great heart," and "I could not but be moved to tears, reading him and being conscious of my complete unity with this man who died hundreds of years ago." It is instructive to compare how ideas in the natural sciences and the humanities have aged.

Let us recall one side of Pascal's legacy—his practical achievements. Some achieved the highest level of distinction, but today few know their

[1]Charles Baudelaire, *Les Fleurs du Mal* (*The Flowers of Evil*), translated by Richard Howard, David R. Godine, Boston. Translation copyright 1983 by Richard Howard. Reprinted with permission from the publisher.



Pascal in his youth.

creator's name. For the writer Turgenev, the standards for convenience and simplicity were "Columbus' egg"[2] and "Pascal's wheelbarrow." Learning that the great scientist had invented a most ordinary wheelbarrow, he wrote to the poet Nekrasov: "Incidentally, in one place I speak of Pascal's wheelbarrow—you know that Pascal invented this so obviously simple machine." And Pascal also originated the idea of the omnibus, a coach available to everyone ("for 5 sous"), with a fixed route, that was the first form of regular urban transport.

Pascal was one of the most notable people in the history of humanity; an immense literature is devoted to him. What aspects of his life and legacy have not been touched by "Pascalology"? There is peculiar testimony to his special popularity in France: Pascal's portrait has been reproduced on French currency (other French writers who have achieved this honor at various times include Corneille, Racine, Molière, Montesquieu, Voltaire, Hugo, and Saint-Exupéry).

Sticks and Coins

When we learn to draw graphs, in the kaleidoscope of anonymous curves we sometimes find ones that are named after people: the spiral of Archi-

[2]This refers to a 16th century anecdote about Columbus balancing an egg on its narrow end.—*Transl.*

17th century) that when Archimedes sought formulas (and had not proved them), he cut a body into infinitesimal layers (indivisibles) and then used mechanical arguments. We know from Galileo's notes that he thought a lot about the method of "indivisibles" but did not write the book he planned.

Soon after mathematicians of the 17th century started to work on the problem of measuring curvilinear areas and volumes, they were confined by the method of exhaustion. The first who chose to proceed along the slippery path of infinitesimals was Johannes Kepler (1571–1630). His *New Stereometry of Wine Barrels* appeared in 1616, where he studied a practical rule for measuring the volume of a barrel using a ruler inserted into the bung-hole. He did not carry out a proof the way Archimedes did, but worked boldly with infinitesimals and expressed his confidence that it was possible to construct a strict proof. Kepler wrote that he set forth Archimedes' principle "only so far as it is sufficient to satisfy a mind that loves geometry, and for strict proofs that are complete in all their parts one needs to look in Archimedes' own books, if one does not fear the thorny path of reading them." This position (we can carry out strict proofs but will not do that) became a long-standing and convenient defense against the need for strict proofs. Here are some examples. Fermat: "It would have been easy to give a proof in the spirit of Archimedes.... It suffices to give notice this time and for always, in order to avoid constant repetition." Pascal: "One of the methods differs from the other only in the way it is expressed." Barrow: "This proof could have been made longer by apagogic arguments (reasoning by contradiction—S.G.), but to what end?" But there were critics who tried to stop those who turned wildly to infinitesimals, invoking the name of Archimedes. An essay by Alexander Anderson (c.1582–1620), a student of Vieta, was directed at Kepler; it was called *A Vindication of Archimedes* (1616). A hundred years later Michel Rolle (1652–1719) stated that "precision no longer ruled in geometry since the time the new system of infinitesimals became involved in it."

Already in the statement of his second law, Kepler considered the area swept out by the line segment joining the Sun and a planet as the "sum" of these segments. Every mathematician who followed tried to devise a safer procedure for working with infinitesimals. Cavalieri (c.1598–1647) was close to Galileo and was honored by Galileo with the highest praise—he was called "Archimedes' rival." Cavalieri devoted two books to the method of indivisibles (1635, 1647). He started with the idea that the area of a figure is determined by the lengths of the segments in which the figure intersects a family of parallel lines (similarly for volume). Cavalieri was sure that his procedure had an advantage over Kepler's method: "Everyone who sees Kepler's tract on the motion of Mars can easily convince himself on the basis

of our studies how easy it was for him to fall into error. . . starting from the proposition that the area of an ellipse is equivalent to the set of all distances of the planet, rotating along an elliptical curve, from the Sun." Cavalieri thought that it was necessary to work carefully with nonparallel segments, but Kepler was not wrong! Only intuition can protect mathematicians from mistakes in working with infinitesimals.

Cavalieri applied his methods to calculating the area of the curvilinear trapezoid under $y = x^n$ (in contemporary terms, $\int_a^b x^n dx$). With great difficulty he gradually increased n, getting up to $n = 9$ between 1635 and 1647. But by this time Fermat (1601–1655) already knew how to calculate the area for all rational numbers $n \neq -1$ (he communicated this to Cavalieri in 1644, but his first results date to 1629). Mathematicians began to feel superior to the ancients. In 1644, Torricelli wrote, "Without a doubt, Cavalieri's geometry is surprising in its economy of means for discovering theorems. . . . This is really the royal road through the thicket of thorns. . . I feel sorry for ancient geometry, that it either did not know or did not want to acknowledge the study of indivisibles."

What is the situation in the case $n = -1$, which was omitted in Fermat's considerations? Here a surprising property turns out to hold: in finding the quadrature of a hyperbola, logarithms appear ($\int_1^x dy/y = \ln x$). This remarkable fact crystallized gradually, beginning with the work of Gregorius Saint-Vincent (1584–1667) in about 1647. Logarithms appeared with John Napier (1550–1617) at the very end of the 16th century because of kinematic considerations, which very much recalls Galileo's first mechanical constructions. However, for a long time they were treated as a purely computational method (tables!) and did not enter into theoretical research. As Torricelli wrote, Napier "only followed arithmetic practice." Roughly speaking, there was still no logarithmic or exponential function and these functions began to appear only after the middle of the 17th century, to a significant degree in connection with quadratures. In essence, the quadrature of a simple algebraic function turned out to be transcendental. The question of the quadrature of a circle and its parts was studied in detail, and here it turned out that the quadrature of an algebraic function ($\sqrt{1 - x^2}$) led to a trigonometric (circular) function. Incidentally, a sinusoid also came up in calculating the area under a cycloid (see the "companion of the cycloid," p. 100).

Gradually, problems about tangents to curves began to occur more often in the range of interests that mathematicians had. The ancients had only known how to take tangents to conic sections, and Archimedes also knew how to construct the tangent to his spiral. Insofar as this problem

is concerned, from the very start the seventeenth-century mathematicians were one up on the supporters of the ancients. Beginning in 1629 Descartes (1596–1650) and Fermat, competing with each other, worked out the general principles of constructing tangents, and Fermat connected them to maximum and minimum problems. In parallel, Torricelli and Roberval (1602–1675) proposed a synthetic method of constructing tangents, interpreting them as the directions of the velocity while moving along a curve, and cleverly representing motion along a curve as the composite of simpler motions. In the 1650s and 1660s, starting with the results of Descartes and Fermat, François Walther de Sluze (1622–1685), Johann Hudde (1633–1704), and Huygens found completely automatic rules for constructing the tangents to broad classes of algebraic curves. Characteristically, none of these authors hurried to publish his rule. In 1659, Hudde wrote to Frans van Schooten (1615–1660), "I ask you to keep secret everything that I write to you, and not to speak of it to anyone, or anything of the kind. It is necessary that my best discoveries either be known only to my most intimate friends, or that they become known to everyone." This is a characteristic illustration of the times. Information was basically spread through letters, books were rarely published, and the first journal (*Journal des Savants* in Paris) began to appear in 1665. Rapid publication was not yet understood as a natural means of preserving priority. It was considered completely acceptable to "hold back" a method in order to extract the maximum use of it oneself.

In 1668 Niklaus Kauffman (1620–1689), better known as Mercator, published a remarkable method for calculating logarithms in his book *Logarithmotechnia*:

$$\int_0^x \frac{dx}{1+x} = \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots,$$

where any precision is guaranteed by taking a sufficient number of terms (the series for $\ln 2$ was obtained earlier by William Brouncker (c.1620–1684)). It later turned out that this series had been known by Hudde (1656) and Newton (1665), but they had not hurried to publish it. Gradually, series became the most important method for calculations and also for theoretical studies. For example, James Gregory (1638–1675) had a very interesting plan for applying series to prove that π is transcendental and to prove that certain problems (calculating the arc lengths of an ellipse or hyperbola) do not lead to elementary functions.

We have given a very cursory description of the status of infinitesimal quantities in the first half of the century, and have not only omitted many famous pages in their history (results of Pascal and Fermat) and

many worthy names (Wallis, Fabri), but also blurred the picture by not discussing the many stages that led to establishing results. The authorship of these stages is very tentative and has often been incorrectly assigned to one mathematician or another: "the discovery came about as a result of almost imperceptible advances, and an argument about priority would be like the violin and trombone arguing for a month about the precise moment when a certain melody appears in a symphony" (Bourbaki).

By the beginning of the 1660s mathematicians had accumulated quite a number of facts. A group of problems was outlined that could be solved using infinitesimals. Two basic directions crystallized: calculating quadratures and constructing tangents. The situation with these problems was essentially different. At that time sufficiently general methods had appeared for the tangent problem, which was newer, but for the quadrature problem everything remained at the level of separate problems and synthetic methods. For example, Descartes was certain that general methods did not exist for these problems. Still, a remarkable connection between these problems was known. They turned out to be inverses of one another, and this was most naturally seen with the help of kinematic considerations: finding the (instantaneous) velocity along a path leads to constructing a tangent, and the path is found from the velocity by using a quadrature. This connection, which Galileo had already outlined, appeared in rather complete form in Isaac Barrow's (1630–1677) lectures given in 1669–1670, although it was still barely being exploited.

Activity in the theory of infinitesimals fell off noticeably at the end of the 1660s. Fermat and Descartes were no longer alive and Huygens had already done his major work. The remaining problems yielded to synthetic methods with difficulty, and there was no constellation of mathematicians of the first magnitude in the mathematical heavens, as there had been twenty years earlier. A sharp break was needed, which required a very talented person who would for a time dare to turn from moving ahead but would rethink everything from the very beginning, rid the theory of synthetic methods, and only go forward after simplifying and systematizing the methods for solving well-known problems. It was necessary to turn the theory of infinitesimals into the calculus—a collection of sufficiently simple, formal, but broadly acting recipes. One had to change the theory from an art into a craft. In this form it would not only go beyond the narrow circle of initiates, but it would also allow the strongest mathematicians to travel along part of the road without wasting energy and to concentrate their efforts on deeper questions. Characteristically, the giants who were already at work, above all Huygens, did not feel the need for this: they did their work the old way. This task had to be taken on by a mathematician

of the next generation.

"And God said: let there be Newton! And there was light," said Alexander Pope in a popular quatrain. Isaac Newton (1642–1727) conceived of the calculus during a two-year stay in Woolsthorpe where, after graduating from the University of Cambridge, he returned to his farm during 1665–1667 because of the plague and was cut off from the outside world. In those two years he obtained his most remarkable results in mechanics and mathematics. Before that he had attended Barrow's lectures and possibly learned from him the idea of systematically considering curves as functions of time: "possibly Doctor Barrow's lectures could have led me to consider forming figures using motion, although now I do not remember it." This is a very instructive statement! Newton constructed the calculus of fluxions. His independent variable was always time, and fluxions were velocities, derivatives with respect to time. He worked out detailed rules for calculating fluxions (our differentiation rules). He also studied the inverse problem—finding the fluent. This is the operation of integration, and Newton systematically explained which rules he could obtain by exploiting the fact that it is the inverse of differentiation (finding the fluxion, in Newton's terminology).

This gives many convenient methods, since everything looks simple when it involves fluxions (derivatives). This scheme—differentiation precedes integration—is how analysis is usually constructed even today. But Newton's main hobby was series. He thought highly of his formula for the binomial $(1+x)^k$ for any k (not necessarily a natural number). He perceived series as a universal method for solving problems in analysis and saw no limitations to this method.

In October, 1666, Newton put together a rough draft of his theory and in the summer of 1669 he sent a summary of his results to Barrow, and through him to John Collins (1625–1683) in London. In 1670–1671 Newton prepared a detailed essay on the method of fluxions but did not find a publisher, and his essays on analysis only began to appear in print after 1704. Some word of his work had spread among mathematicians, and some could acquaint themselves with the manuscript that Collins possessed. Newton did not hurry to publish but watched quietly as some of his results were rediscovered and published by others (e.g., his results on series were published by Mercator). Hardly anyone in his circle could appreciate the importance of the calculus, and sooner paid attention to concrete results. Newton himself thought highly of concrete results and promoted his series method rather than the calculus. Thus, in the 1670s "only Newton in Cambridge and J. Gregory, alone in Aberdeen, remained active, but they were soon joined by Leibniz with all the passion of a neophyte" (Bourbaki).

Leibniz and his Mathematical Journey

Throughout his life, Leibniz aimed at global problems and universal theories. His journey in mathematics was not a standard one, and it was partly because of this that he had a preference for method in a century when most valued concrete results. Leibniz had many plans during his life, some strikingly grandiose. New ideas supplant old ones, not infrequently carrying away an author lacking in realism. He barely finished writing any of the books he planned, and most were abandoned at the start (he finished the better part of only a few books on philosophy). But how hard it is to stay realistic when your ideas are far ahead of your century!

Already when he was 13 or 14, Leibniz dreamt of reconstructing logic, of creating an alphabet of human thought in which one could record all thought processes. The main idea of his life gradually ripened: the creation of "a universal characteristic," "a universal language," "Universal mathematics is, so to say, the logic of imagination;" it should be concerned with everything "that in the realm of imagination lends itself to precise definition." Language must be defended from recording incorrect thoughts: "chimerae, which are not understood even by the one who creates them, cannot be recorded by their signs." He dreamt of a machine that would prove theorems and wanted to turn thought into calculation, arithmetizing it so that it would be possible to replace discussion by computation and settle arguments with mathematical calculations. Three times Leibniz started to realize his ideas, which were grandiose and far ahead of his time, but each time he stopped after taking only the first steps. Only in the 20th century, when much of what Leibniz conceived turned out to be reality in the area of mathematical logic, did it become clear that his ideas were not so much Utopian as they were perspicacious.

Leibniz was interested in a variety of applications of mathematics, and he believed its possibilities were endless. He prepared to become a lawyer and at age 18 tried to construct jurisprudence as a mathematical theory with axioms and theorems. He thought of applying probabilistic ideas to legal procedures. At age 20 he refused a chair at the University of Nuremberg; he was not drawn to a peaceful academic career. Leibniz's plans were more ambitious: "In my soul I have long cherished something else" and "I do not consider it suitable for a young person to sit, pinned to his chair; my spirit burned with the desire to achieve greater scientific fame and to see the light." He received an invitation from the duke Johann Philip and moved to Mainz. Leibniz wanted to exploit his situation and, within the scope of a rather modest state, to create a perfect legal code. His plans gradually became more broad and at the same time less realistic. He thought of

reorganizing all of juridical knowledge, and began with three grandiose monographs. It is likely that when Bernard Fontenelle, the permanent secretary of the French Academy of Sciences, called him a great lawyer in his *Eulogy of Leibniz*, he had grounds to do so.

Leibniz had more than a few interesting ideas, but soon he had a series of completely different ones. The famous diplomat Johann Christian von Boineburg, living in Mainz, captivated Leibniz with grandiose plans to change the politics of Europe. Provincial Mainz was too small for their ideas. They undertook a proposal of the Elector (Kurfürst) of Brandenburg to find a justification for electing the German prince to the Polish throne. Leibniz composed a memorandum which was brilliant but did not keep him from losing in this matter; correct, practical diplomacy turned out to be more effective than a political pamphlet. The next project involved organizing a union of German states against France. It contained a number of sharp-witted steps but was not successful. Finally, the third grandiose project was to draw France into a war with Turkey in order to weaken its influence in Europe. To realize this project Leibniz went to Paris. The natural result was that Leibniz lost the support of the Elector, who was not very interested in an advisor who was trying to reconstruct European politics by going over his head.

It is possible that these circumstances, in which Leibniz found himself without something to do, switched this ebullient nature over to mathematics. In Leibniz's original plans, mathematics had played an auxiliary role. In 1666 he published *A Dissertation on the Combinatorial Art* in Leipzig, in which he said that he was not interested in discovering new arithmetical truths; mathematics had to help him work out a "logic of discovery." And in Mainz he found time for "mathematical leisure." In 1676 he worked on building an arithmetic machine and was interested in Pascal's machine. Leibniz sent several mathematical results to Paris. In the autumn of 1672 these were the subject of discussions with Huygens, who was working in Paris at the time. The issue was the summation of a numerical series $a_1 + a_2 + \cdots + a_n + \cdots$ by choosing a sequence $b_1, b_2, b_3, \ldots$ for which $a_n = b_n - b_{n+1}$. Then $a_1 + a_2 + \cdots + a_n = b_1 - b_{n+1}$. Leibniz considered a number of examples where this rule worked and fortunately it fit an example proposed by Huygens:

$$\frac{1}{1 \cdot 2} + \frac{1}{2 \cdot 3} + \cdots + \frac{1}{n(n+1)} + \cdots$$

(here $b_n = \frac{1}{n}$). Neither one knew that this method was not new and that this was a very special case. Nevertheless, Leibniz had a very high opinion of his accomplishments. Later he evaluated the situation sensibly: "When

I came to Paris in 1672, I was a self-taught mathematician but I had little experience, and I had no patience to go through a long chain of proofs... I wanted to swim by myself without a teacher.... In this great mathematical ignorance I paid attention only to history and law, and saw studying them as my goal. However, mathematics was for me a pleasant diversion."

In 1673, Leibniz visited London as part of a diplomatic mission from Mainz. Contacts with English mathematicians were sobering for him. He recognized that his basic results were not new, and that modern mathematics was far ahead of him. Leibniz had just one way to catch up to modern mathematics—to start from the beginning. Twenty-seven is not the most appropriate age for starting a young person's science, but this did not put Leibniz off. He had all the grounds for later calling himself the "most studious of mortals" (letter to Jacob Bernoulli, 1703). In the autumn of 1673 Leibniz began years of study in mathematics skillfully guided by Huygens, who divined that this self-confident "overage child" had an authentic gift. "...Huygens, who, I suppose, considered me to be more capable than I really was, gave me a copy of his just published *Pendulum*. For me this was the beginning or the occasion for deeper mathematical work." Thus, everything began with the great book *Pendulum Clocks*. Then followed Saint-Vincent, Descartes, Sluze, John Wallis (1616–1703), and most of all Pascal. Leibniz saw that Pascal essentially applied a very general method to a particular problem and, surprised that "Pascal's eyes were closed," tried to articulate this method and apply it to other problems. Thus his so-called method of the "characteristic triangle" appeared in which an infinitesimal triangle is replaced by a finite one, which was real progress compared to the method of indivisibles. It would not have been bad for Leibniz to read the more classic texts, but he was in a hurry. Indeed, he could have made his way "to geometry really by the back door." Results appeared that astonished Huygens, for example, the series

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots .$$

Later it turned out that Gregory had known this result. Huygens expected that one could use this series to obtain the quadrature of the circle (Gregory, by contrast, expected it was a way to prove that π is transcendental).

Leibniz did not just occupy himself with analysis. He tried to find a formula for the solution of a general algebraic equation (really general—particular problems were of little interest to him), analyzing Cardano's formula in the complex domain (Huygens was surprised by the relation $\sqrt{1+\sqrt{-3}} + \sqrt{1-\sqrt{-3}} = \sqrt{6}$), and worked on compasses for finding the roots of any equation (similar to using ordinary compasses to find square

roots).

All the same, the major results involved infinitesimals. Leibniz wrote that already in 1673 he had "completed some one hundred pages" but still "did not consider this work ready for publication. But I am bored working on trivialities when an Ocean has opened up before me."

He had obtained many theorems during the first year of his "apprenticeship," but most of them could be found in the work of Gregory or Barrow. However, general methods allowed him to obtain everything more simply and uniformly. Leibniz's path had been chosen: he was developing the calculus of infinitesimals.

The nature of his talent and his previous scientific experience were the best preparation for achieving this goal. He thought in a clear way about the class of functions that should be considered in analysis (the very word "function" was first used by Leibniz in 1673). He decisively rejected the idea of limiting this to algebraic functions (the geometric curves of Descartes) and thought it was necessary also to consider transcendental functions (this is Leibniz's term; Descartes spoke of mechanical curves in these cases). From the start, as he developed the calculus he worked out its notation, which in the end took the form that has come down to our times.

Leibniz, as no one before him, understood the importance of scientific notation, and not only in mathematics. The calculus of infinitesimals gave him an excellent way to realize this idea. Good notation not only simplifies the use of the calculus but is essential for mastering it. In 1678 Leibniz wrote to Ehrenfried von Tschirnhaus (1651–1708), "One must be concerned that the symbols are opportune for discoveries. This is achieved in the greatest measure when the symbols briefly express and somehow represent the deepest nature of the thing, and then the work of thinking is surprisingly shortened." Throughout, Leibniz looked for the chance to introduce convenient notation. It is worth recalling that the method of solving a system of linear equations using determinants goes back to him, and in connection with this he wrote to L'Hôpital (1693), "Part of the secret of analysis consists of the art of employing well the symbols that are used, and as a small example you see, sir, that Vieta and Descartes still have not learned all its secrets." We should stress that the notation in Newton's calculus was not well developed. He himself wrote that he "doth not place his Method in Forms of Symbols, nor confine himself to any particular Sort of Symbols for Fluents and Fluxions."[1] Significantly, Huygens did not appreciate the usefulness of analytic notation. With his gifts he was in a position not to

[1] "An account of a book entitled Commercium epistolicum Collinii et aliorum, de analysi promota," *Philos. Trans. Roy. Soc. London*, **342** (1714–1715), pp. 173–224. This was published by the Royal Society and is believed to have been written by Newton himself.—*Transl.*

need it. Leibniz tried to explain its advantage: "I completely imagine that you have available a method equivalent to mine for the calculus of differences. What I call dx or dy you may denote by other letters. However, this is roughly the same as if you had always wanted to substitute letters in place of roots or powers.... Judge for yourself how difficult that would have been...." What Huygens was able to manage without was absolutely necessary for converting analysis into an everyday practical method. It is likely that notation was the decisive reason why we use Leibniz's variant of analysis today.

As early as 1674, Leibniz was sure that "all learning about sums and quadratures can be reduced to analysis, something which no one hoped for up to now." By the end of 1675 Leibniz had developed his first approximation to the calculus, and he had reason to be convinced of its effectiveness. An important moment was his solution of a problem of Florimonde de Beaune (1601–1652), on which Descartes had worked but could not complete: "Already last year I had posed a question to myself which can be taken to be among the most difficult in all of geometry, since the methods that have been disseminated so far yield almost nothing here. Today I found its solution and am carrying out its analysis" (November 11, 1675). He was speaking about finding a curve with constant subtangent (the interval between the projection of a point A on the x-axis and the intersection of the tangent to A with the x-axis). The difficulty is that the solution is related to the logarithmic function. By the middle of 1676, differential and integral calculus had been joined once and for all. He was amazed that "thanks to this calculus everything appears before the eyes and in the mind in exquisite compactness and clarity."

Leibniz, as well as Newton, tried to create a powerful method without worrying at that stage about a sufficiently rigorous basis for the calculus. "Newton and Leibniz, turning their backs on the past, decided temporarily to search for the justification of the new methods not in strict proofs but in the abundance of results and in their mutual agreement" (Bourbaki). While still in the learning stage it seemed to Leibniz that Gregory was too keen on "proofs in the old manner." Leibniz regarded concrete results above all as a possible illustration of his method. This may tell us that he never could do computations easily and always envied calculators "of iron or copper." Later (in a 1696 letter to L'Hôpital) he connected this to the fact that he was busy with many different things at the time: "My mind, occupied with many subjects, is not able to concentrate to the necessary extent and because of this I stumble every minute, and when I exert my attention I have the unpleasant sensation of some sort of heat." In 1699, "calculations become more pleasant when I divide them up with someone

and I am not occupying myself with calculations for a long time, if no one is helping me."

In 1675 in Paris, Leibniz had a suitable partner in his countryman Tschirnhaus. Their abilities were often complementary, and this made their collaboration especially fruitful. Tschirnhaus mostly studied algebraic equations but was also interested in quadratures. Leibniz was hurt that his friend could not appreciate the usefulness of the calculus: "...some quadratures that you obtained are verbose but elegant and are beautiful in their own right, and I consider them only as consequences of the general calculus. I write this, my friend, since I see with regret that you often waste much time only because you do not want to pay enough attention to some of my observations" (1678).

It seems that Leibniz had heard that Newton possessed some sort of powerful techniques and decided to discuss his new method with him. Through the intervention of Henry Oldenburg, the secretary of the Royal Society, there was an exchange of letters in 1676. Leibniz wrote about the problems that he knew how to solve, asked to find out about Newton's methods, and promised to tell about his own method. Leibniz had written earlier to Oldenburg that creating the method was the only thing to which he attached significance. Newton was not surprised by Leibniz's results. He immediately noted that De Beaune's problem was reduced to the quadrature of a hyperbola (by logarithms), and regarding the series for π, he remarked that it had taken 1000 years to compute the first 20 digits. Newton spoke very sparingly of the method. It is only clear that the center of gravity in his opinion lay in power series. Newton claimed that he could solve any differential equation using them. The main part of the information was enciphered in two anagrams in which the first letters of the words were written in a code (5accdae10effh...) whose meaning Newton revealed much later. This was an ancient method of claiming priority. Perhaps the focus on power series kept Leibniz from realizing that Newton had the calculus.

Leibniz did not agree that series solved all the problems. "We still do not, as far as I know, have available a general inverse method for tangents." He saw a different picture. It was necessary to reduce the solution of differential equations to known quadratures. It was important to examine whether the elementary functions and the quadratures of the hyperbola and circle (logarithmic and trigonometric functions) were enough. Gregory used impressive arguments to show that these quadratures were not enough to calculate the arc length of an ellipse or hyperbola. Then one had to "establish some other higher fundamental figures" (elsewhere, "higher transcendences in geometry") which would be sufficient to solve differen-

tial equations. Newton was also not caught unaware by this statement: he wrote that for any α and β the integral $\int x^{\alpha}(1+x)^{\beta}dx$ reduces to known quadratures. The correspondence was broken off at Newton's initiative and in 1677 Oldenburg, through whom it had been conducted, died.

Mathematics and the "Conquest of the Minds" of Sovereigns

Leibniz's life changed decisively. Only rough drafts and sketches of articles remained from his Paris period. He perceived a plan for preparing a comprehensive book, *The Mathematics of the Infinite*, but changes in his life diverted him from mathematics.

We do not know if Leibniz's political ambitions bubbled up again, or if he did not find it possible to devote himself to a life of science (perhaps his Protestantism interfered with obtaining a position at the Academy of Sciences in Paris). One way or the other, starting in late 1676 he was in the service of the duke Johann Friedrich in Hannover. He went to Hannover by a circuitous route: he visited London, where he was seen with many mathematicians but did not meet with Newton, and he met with Spinoza in Holland.

Thus Leibniz was only able to obtain a position with a second-rate ruler, and at first he was only the duke's librarian. It was not the most enviable position for a 30-year old scholar-politician who had not given up his ambitious plans. But Leibniz was full of enthusiasm and dreamt about creating the best library in the world, and was not restrained by the fact that its size realistically did not give him the means to do that. He was permitted to practice law but preferred not to be a full-time lawyer, for which he had no taste, as opposed to working on global legal problems. Leibniz was allowed to take part in diplomatic activities in a very limited way. He was entrusted with preparing a text promoting the right of the duke to take part in peace negotiations between France and Germany. Johann Friedrich was a Catholic monarch in a Protestant state, and Leibniz wanted to use this situation to realize his hidden aim of uniting the Catholic and Protestant religions.

In 1678 a new duke, Ernst August, was on the throne, and things began to get worse. But Leibniz was full of projects across an unusually large spectrum: improving a furnace, producing nails and hammers, improving carriage wheels, fishing rods, and ships' paddle wheels, foundries, fire fighting, reorganizing the archives, composing *The Legal Code of Ernst August*, and so on. Almost none of the projects found support. The most far-fetched was a plan for improving water engines in the mines in the Harz Mountains. This project was interrupted in 1685, since it was thought to be

hopeless. In each project Leibniz was clever and resourceful but he often lacked realism. Success came when reasoned thought was combined with talented practicality. That is how it was with the steam engine: "...Leibniz, scattering brilliant ideas around him, as always, without worrying about whether the discovery of these ideas would be attributed to him or to someone else. As we now know from the correspondence of Denis Papin (1647–c.1712), published by Gerland, Leibniz gave him the basic idea: to use a cylinder and piston" (Friedrich Engels). As a curiosity, we recall that Leibniz proposed to the priest Francesco Grimaldi, who was going to China, that he acquaint the enlightened emperor with the binary number system and use it to convert him to Christianity by proving the unity of God.

For a long time, a constant struggle for influence in the court kept Leibniz from mathematics. His return was stimulated by two circumstances. Since 1682, the journal *Acta Eruditorum* had begun to appear in Leipzig with Leibniz's support, and he proposed to publish his results there. In 1683–1684, the journal published Tschirnhaus's articles on quadratures, in which Leibniz discovered the results of his recent discussions with the author and without the required references. Once Leibniz had unsuccessfully tried to convince Tschirnhaus of the effectiveness of the calculus, and now the latter had himself published results in that direction. It is very likely that Tschirnhaus did not remember that the source of his statements was Leibniz. It happens that misunderstood ideas can hide deeply within oneself, and then after some time arise as if they were one's own.

In May of 1684, Leibniz published an article sharply criticizing Tschirnhaus (without claims of priority and without giving his full surname) and in October his famous article appeared, which we spoke about at the beginning. In seven pages he formulated the basic rules of differential calculus, discussed the connection with maximum and minimum problems and points of inflection, and gave some examples (a proof of the law of refraction and De Beaune's problem). He gave an optimistic assessment: "That which a person versed in this calculus can obtain directly in three lines, other most learned men were obliged to seek by following complicated roundabout ways." At this time Leibniz was not really engaged in analysis. He only published some of his mathematical "inventory." Leibniz published a few additional articles. Among them was an 1686 article "on the deeply hidden geometry and analysis of indivisibles and infinities." It contained the first appearance in print of the integral, which was still called the sum but was denoted by $\int$. The term "integral" was introduced by Johann Bernoulli. The article clearly stated the inverse relationship of the operations of differentiation and integration, and underscored the need to

consider transcendental functions in analysis. It included short historical notes. Newton was called a "most deeply talented geometer" and noted that the publication of his methods promoted "a scientific advancement of no small importance." This marked the close of the second period in Leibniz's mathematical life.

Matters at the court took a turn for the worse. In 1685 the Harz project finally failed. The duke aimed at becoming the ninth Elector (a prince who took part in electing the emperor). Leibniz was assigned a clearly limited but valuable role. He was supposed to research the history of the house of Welf, to which the duke belonged. This was needed to strengthen the duke's claims. The inquisitive Leibniz became completely engaged in this modest work of historiography. In particular, it gave him the opportunity to get away from Hannover. In 1687 he went on a three-year trip to work in the archives of Germany and Italy. He had not left Hannover for ten years, and his contacts with scholars were extremely limited. He had tried to replace travel with an active correspondence. While still in Mainz he had about 50 correspondents, this grew to 70 in Hannover, and to 200 by the start of the new century. But all these letters could not replace personal contacts. Summing up his journey, Leibniz wrote, "The trip has in part served to free me from my usual duties and to heal my spirit, and I obtained satisfaction from conversations which I was in the habit of having with many erudite people who were skilled in the sciences." Moreover, it was confining for Leibniz to be in the service of the Hannoverian duke. For his plans it was important to "conquer the mind of the great sovereign." He obtained an audience with Emperor Leopold in Vienna (" I survived the day that I had been wishing for for twenty years."). Among the proposals that were met with favor was a project to organize an Academy of Sciences in Vienna. But soon the emperor, busy with the war with France, stopped short of establishing the Academy.

After 1690 Leibniz was in Hannover again. He thought it would take two or three years to complete *A History of the Welfs*. But this estimate, as always, was too optimistic. His conception was too substantial and grew along with the work. Leibniz was unable to limit the task and the book hung over him like a heavy weight to the end of his days.

In Hannover, a letter was awaiting Leibniz that had been sent to him in 1687 by Jacob Bernoulli (1654–1705). Bernoulli had read Leibniz's articles and his spirit was imbued with the new calculus. While he waited for Leibniz's reply he began to work actively in analysis, also involving his younger brother Johann (1667–1748). Leibniz found the understanding for which he had waited many years. He had not even dreamt of better disciples. Leibniz had gotten his scientific school (which Newton lacked).

In contacts with the Bernoulli brothers, Leibniz began to develop analysis systematically. They published articles in *Acta Eruditorum*, exchanged letters, and discussed problems. Later the Marquis de L'Hôpital (1661–1704), a student of Johann, joined the triumvirate. In 1692 Johann Bernoulli prepared lectures on differential calculus but did not publish them, and in 1696 L'Hôpital published the first course on differential calculus, *L'Analyse des Infiniment Petits pour l'Intelligence des Lignes Courbes*, or *The Analysis of Infinitely Small Quantities for the Study of Curves*. We will not dwell on the results that Leibniz and his colleagues obtained during these years, but will discuss how his view of analysis changed as a consequence of them.

At the end of the century, it seemed to Leibniz that everything had been accomplished in mathematics: "From now on I consider pure mathematics only as an exercise for developing thinking skills. For practical purposes, almost everything in it has been discovered with the help of the new methods." In September, 1692 he communicated his plans to Huygens: "I want us to bring the analysis of numbers and curves to completion during this century, at least in the main, in order to deliver the human race from this trouble so that henceforth all the acumen of human reason will be turned toward physics." But later, as we see from a letter to L'Hôpital, he was no longer so optimistic: "One should not be surprised that the analysis of infinitely small quantities is taking only its first steps and that we cannot manage the situation even with quadratures, and with the inverse problem of tangents, and even in the least measure with solving differential equations. . . ." He clearly saw that natural problems did not reduce to known quadratures and he did not see a way to systematize "higher transcendences." This was a problem for the next two centuries.

Leibniz's scientific authority grew. One mark of this was his election to the French Academy of Sciences in 1699, as soon as they were permitted to elect non-Catholics. But it became even harder for him to combine his service with science. He broke out of the limitations of Hannover. In the 1690s he was also in the service of two German monarchs. In 1700–1711 he added service to the Elector of Brandenburg, Friedrich III, who became the king of Prussia. Here, Leibniz's project was to organize a scientific society, but intrigues forced him to abandon Berlin before its opening ceremonies. The idea of organizing an imperial academy in Vienna was renewed and this was strongly promised in 1713, but then Charles VI decided to give up this very expensive plaything. The geography of Leibniz's interests widened: "I do not belong to those who are nourished by passion for their homeland or in some other country, my thoughts are aimed at the good of the whole human race; for I consider the heavens as my native land and all good-thinking people as my fellow citizens." Leibniz wrote this to

the Russian Tsar Peter I in January, 1712. They had become acquainted in 1711 at the wedding of the Tsarevich Aleksei and had met several times afterward.

Peter took Leibniz into service as a confidential legal advisor to help with regulating Russian legislation. They discussed the question of organizing an academy in St. Petersburg. The range of questions discussed with the Tsar was immense: finding a route from the polar seas to the Pacific Ocean, Christian missions to China, uniting Orthodoxy with Catholicism and Protestantism (convoking an ecumenical council), and creating a broad anti-French coalition. It seems that they found a common language. This activity of Leibniz's did not find favor with the Hannoverian duke. Although he was made a confidential advisor, Leibniz did not get his wish to "rise" to the rank of vice-chancellor. The new duke (since 1698) Georg Ludwig insistently expressed his wish finally to see the "invisible book," the long-awaited *History of the Welfs*. He essentially removed Leibniz from all activities and tried to restrict his external contacts. This firmly strengthened Leibniz's reputation as a "monarch hunter," about which Arnold Eckhard, who helped him in his historical research, said unkindly during Leibniz's final illness, "If a ruler or a dozen nobles promise him a salary, then he will be able to get up." And the gravely ill scholar tried with his last strength to complete the unfinished *History*.

There was no question of systematic scientific investigations. In 1695 he wrote, "There are no words to describe how unfocused I am. I search for old things in the archives and plan unpublished manuscripts, with the help of which I hope to shed light on the history of the Braunschweig house. I receive and send more than a few letters. I have so much that is new in mathematics, so many thoughts in philosophy, so many other new things to say in literature that I cannot let die, that I often do not know what to do first, and I feel how right Ovid was when he cried, 'Plenty has made me poor'.... It is already more than twenty years since the French and English saw my calculating machine.... Now with the help of workers I have gathered together the machine is ready, allowing multiplication of up to 12 digits.... But first of all I wanted to finish my *Dynamics* in which, I claim, I have finally found the true laws of material nature.... My friends, who know about the higher geometry I have constructed, insist on publishing my *Science of the Infinite*, containing the basis of my new analysis. To this I have to add the new *The Nature of Position* on which I am working, and more significantly more general things about the art of discovery. But all these works, excluding the historical ones, are done under cover. After all, you know that at the courts they seek and expect something very different! Therefore, from time to time I have to investigate questions of in-

ternational law and the law of imperial princes, especially of my master. . . . Meanwhile I often have to discuss religious disagreements. . . . And all the same I try to put in order my reflections on law." In 1697: ". . .If you weigh all that, . . . then wish me to have assistants, young people or friends, astute and diligent, who would want to support me. For I can give a great deal, but I cannot complete everything that I see, and I would willingly communicate it to others if it would bring them fame, if only it would serve some common cause, the good of mankind, and the very glory of God." In a letter to Johann Bernoulli from 1697: "daily reflections on themes not only in mathematics but also physics, and the deepest philosophy, history, and law, reflections which I record in the briefest way so that I do not lose them. . . . Add to this my ideas on constructing natural law. . . ; but first of all I am busy with a new analysis for reasoning of a higher sort. . . . I leave it to you to decide whether I have enough time for fundamental studies in geometry."

He often thought about mathematics while in his carriage (we know that is where he thought of the rule for differentiating an integral with respect to a parameter, in 1697). His ideas were overflowing; he expanded his thinking to include the creation of a "geometry of position." "I have not yet decided to publish my projects on the nature of position for if I do not make them convincing, giving some essential examples, then it will be thought of as a fantasy. Nevertheless, I foresee that the work cannot fail" (letter to L'Hôpital, 1694). It seems that nothing was published and Euler later tried to unravel the great idea. When differential geometry and then topology were created in the 19th century, it was thought that this was the realization of Leibniz's project.

The last years of Leibniz's life were darkened by a controversy with Newton over priority. The argument gradually grew into an accusation of plagiarism against Leibniz. It was implied that Leibniz may have become acquainted with Newton's manuscripts in London, but today the independence of Leibniz's discovery has been established. There was no sufficiently detailed text in London, during his first visit Leibniz was not prepared to grasp Newton's theory, and there was no one who understood the calculus well enough to communicate it to Leibniz. By his second visit to London, Leibniz already possessed his own calculus. At first the controversy went on without the participation of Newton and Leibniz. It is surprising that Newton, who always stayed away from such arguments and cared little about preserving his priority, energetically joined the dispute this time. It is likely that Leibniz hurt him very much by not once acknowledging him as the creator of a new calculus (his publications had appeared by then). A real persecution was organized against Leibniz in England. A special

commission was constituted and a collection of materials was prepared. In 1714 Leibniz tried to write his own "History and Origin of the Differential Calculus," but he could not withstand the pressure in England.

Things became still more complicated in 1714 when the duke became King George I of England. Leibniz counted on going to London and becoming the royal historiographer, but in an insulting way he was even kept from traveling to the coronation (he was made to complete the "History"). He had played his role and the fact was that the king did not want to have in his entourage someone who had plunged into opposing Newton, who was celebrated throughout England. Leibniz died in 1716. He was buried modestly under a tombstone with the short inscription "The Remains of Leibniz."

Leibniz once wrote to Peter I, "Although I must often work in political and legal fields, and noble princes sometimes use my counsel in these matters, all the same I preferred science and the arts since they constantly promote the glory of God and the welfare of all mankind... science and crafts are the real treasure of mankind, for through them art overcomes nature and civilized peoples are distinguished from barbarians. Therefore, I have loved science since my youth, have studied it and had the happiness... of making varied and very important discoveries that have been praised in print by impartial and famous people. I did not just find a powerful monarch who was rather interested in it."

Evidently, over the years Leibniz's priorities shifted: for a long time he had given preference to politics over science, but life cruelly taught him how thankless the position of a scientist at court is.

Leonhard Euler

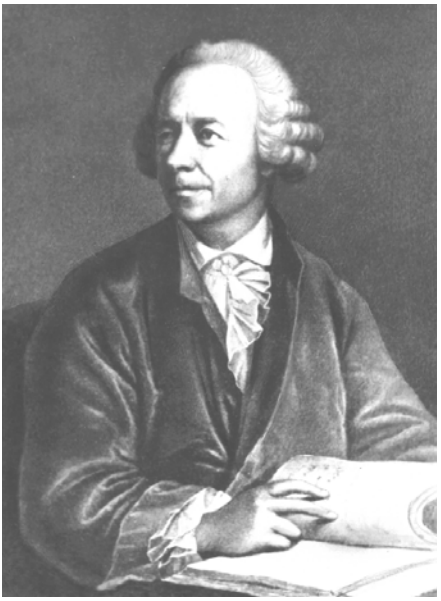
Thus, Euler ceased to calculate and to live. Condorcet

In early 1783, Princess Ekaterina Romanovna Dashkova was named director of the Petersburg Academy of Sciences. Twenty years earlier she had been the closest associate of Catherine II ("Catherine the Great") when the latter ascended to the Russian throne. Known for her ingenuity, the princess had thought of the perfect way to convince the academicians of her devotion to science. She persuaded the elderly Euler, who for a long time had not been on good terms with the academic establishment and had not attended Academy conferences, to accompany her on her first visit there. The blind Euler appeared with his son and grandson. Dashkova later recalled, "I said to them that I asked Euler to take me to the session since, regardless of my own ignorance, I consider that by such a ceremonial act I was testifying to my respect for science and enlightenment."

Several months later the proceedings of the Academy recorded, "At the conference session of 11 September 1783 Academician N. I. Fuss assumed the duties of the Secretary,[1] who was absent because of the passing of his famous father Mr. Leonhard Euler, who died of an apoplectic stroke on 7 September at 11 o'clock in the evening, at the age of 76 years 5 months and 3 days, completing his long and brilliant journey and making his immortal name known throughout Europe." The first sign of this unfortunate event appeared at the beginning of September in the form of mild dizziness, when Euler calculated the speed at which a balloon rises. On the day he died he discussed the results of computing the orbit of Uranus, which had recently been discovered by William Herschel, with the astronomer Andrei Leksel.[2]

[1] Johann-Albrecht, Euler's eldest son.

[2] Also known as Anders Lexell.—*Transl.*



Leonhard Euler.

Euler's exceptional personality and his unprecedented role in the history of the Academy led to the search for an out-of-the-ordinary way to honor his memory. On October 23rd, Academician Nicolaus (Nikolai Ivanovich) Fuss, Euler's student and the husband of his granddaughter, delivered his eulogy. The academicians decided to commission, at their own expense, a bust of the "immortal Euler, deserving of admiration equally for his genius and for his merits," and their "illustrious leader" (Dashkova) "added to this a magnificent column which will serve as the base for this bust." They discussed in great detail the publication of the deceased's works, including the quality of the printing paper. At first the bust was placed in the library, and later it was located opposite the president's chair in the hall where the Academy met. A picture remained in the library, entitled *Silhouettes of the Group of Academicians of the Mathematics Class, Installing the Bust of the Late L. Euler*.

Word of the honors paid to the scholar spread far beyond the boundaries of Russia. The Marquis de Condorcet,[3] Permanent Secretary of the

[3]In less than 10 years he would take part in the French Revolution and his name would be crossed off the rolls of the Petersburg Academy for "reprehensible behavior... against a sovereign."

French Academy of Sciences, said in his eulogy, "Thus a people, which at the beginning of this century we took for barbarians, gives an example to civilized Europe—how to honor great persons for their lives and how to honor them in death. And other nations should blush that they could not anticipate Russia or even imitate it, and not only in this matter." Although from far-off Paris the situation in Petersburg seemed happier than it really was, the attitude towards science during the 60 years of existence of the Petersburg Academy was unrecognizable.

The Academy's First Years

As early as the last years of the 17th century, Peter I ("Peter the Great") had thought about organizing an Academy in Russia. Beginning in 1711, he discussed his plans with Leibniz three times and even took him into the Russian service. Leibniz, a great fantasizer who dreamt of spreading academies throughout the world, responded enthusiastically to Peter's idea when he met the sovereign for the first time. Leibniz did not think that the absence of science in Russia was an obstacle to the creation of an Academy and even found some advantages in this. However, few in Russia shared this optimism. V. N. Tatishchev, one of Peter's most educated associates, told him that "...there is no one to teach, for without lower schools an academy will be a great expense and be useless." Peter replied, "I can reap great stacks, but there are no mills." Therefore he decided to build a "water mill" first, although "there is no water nearby and there is water rather far away," not in the hopes of "making a canal" but in the hopes that the mill "will force my successors to bring the water." There were many problems along the way but in 1724 the Senate decided to create an Academy of Sciences. At the time the word "science" did not even exist in Russian and the Academy was called the "Academie des Sciences."[4]

Peter died in 1725 and so did not see the opening of the Academy. It fell to his successors "to take part in building the mill." His wife, who succeeded him as Catherine I, carried out her husband's plan (not without some hesitation), although she did not share his interest in science. As one contemporary wrote, "the scientists' eulogies were not understood by Her Majesty." The fate of the Academy always hung by a thread. It was seen as a German phenomenon and so the Russian party, and in particular Alexander Menshikov, was inclined against it. The public did not understand the function of the Academy well and the academicians showed how strong

[4]Its Latin name was "Academia Scientiarum Imperialis Petropolitanae."—*Transl.*

they were. The Petersburg diary, published in *The Saint Petersburg Gazette*, kept a record of the public lectures given by academicians on the occasion of the coronation of Peter II (1727), when the academicians Joseph Nicolas Delisle and Daniel Bernoulli discussed the revolution of the earth around the sun, the academician Gotlib Baier delivered a "funeral ode in Latin verses," and "at the same time fountains running with white and red wine were turned on for the people who were strolling all night in the Tsaritsa's meadow."

The academicians tried to mend their relations with the Russian public and opened the doors of the Academy to visitors twice a week. Sometimes one could see something surprising there. On February 24, 1729 "Professor Leutman managed to change a picture of the state coat of arms (using a prism) into a portrait of the reigning emperor." Some academicians established themselves by successfully organizing "amusing fires" and illuminations, composing celebratory odes, and casting horoscopes. More lofty matters were not highly valued, except perhaps for making maps recommended by some navigators. The by-laws of 1747 asserted, "It can only be to the benefit and glory of the state for there to be such persons who know about the course of heavenly bodies, time, navigation, and the geography of the whole world and of their state." Peter II died in 1730. His successor Anna Ioannovna visited the Academy only once, and the Academy disappeared from the Petersburg diary for a long time.

The academicians had been assembled into a "scientific society" under Peter I. It gradually became clear that a first-class group could not be brought together: eminent scholars thought of a voyage to Russia as questionable and even risky. Leibniz was no longer alive, and his closest colleague Christian Wolff declined to accept the post of president. The first president was Laurentius Blumentrost, a court physician. Instead of inviting eminent scholars, they tried to invite their children in the hopes that scientific ability was inherited and that a famous name would adorn the rolls of the Academy. Thus an invitation to the famous Johann Bernoulli (1667–1748) was addressed to his son. In an extended correspondence it was not clear for a long time whether the invitation was for the oldest son Nicolaus (1695–1726) or for the middle son Daniel (1700–1782). In the end, both went: Nicolaus, a former law professor in Rome, became professor of mathematics with a salary of 1000 rubles per year and Daniel became professor of physiology at 800 rubles. Their father's parting words were, "...it is better to withstand a severe climate in a land of ice where the muse is present than to die of hunger in a land with a moderate climate that detests and offends the muse." Could he have known then what would happen in a year and what his oldest son would become!

Euler in Petersburg

In 1726 the Bernoulli brothers were enviously seen off by Leonhard Euler, their father's student: "I had an indescribable wish to go with them. . . . But that could not come true so soon, and meanwhile the young Bernoullis who had been invited vigorously promised that while they were in Petersburg they would plead for a suitable position for me."

Leonhard Euler was born on April 15, 1707[5] in Basel, Switzerland. His father, Paul Euler, was a village pastor. In his youth he successfully studied mathematics under Jacob Bernoulli (1654–1705), Johann's older brother. He received his first lessons from his father and took his later classes at the gymnasium in Basel while simultaneously attending mathematics lectures at the university, where Johann Bernoulli taught. Euler was soon studying primary sources independently, and on Saturdays Johann Bernoulli met with the talented student to discuss points that were not clear. Leonhard was friendly with his sons, especially with Daniel.

In 1723 Leonhard received the degree of Master of Arts, and at his examination gave a speech in Latin comparing the philosophies of Descartes and Newton. Paul Euler thought his son should follow him in his career, and Leonhard obediently studied theology. Both father and son understood that a career in science had no prospects. Even though science was not especially prestigious (at the time the Swiss were fond of saying, "let the Germans study it, the Swiss have more important work to do"), there were more candidates for professorial positions than there were vacancies.

In 1727, Euler tried to get a position as physics professor at Basel but was doomed in advance to failure. At the same time he successfully took part in a competition at the French Academy of Sciences on the best way to position the masts on a ship. It is noteworthy that "in mountainous Switzerland, which Euler had so far never left, he of course had never had the chance to see a ship except in pictures. . ." (Aleksei Krylov).[6] This was Euler's first but not last contact with naval science.

Daniel Bernoulli kept the promise he made when leaving for Petersburg. Even before Euler tried for the position in Basel he knew he could get a position in Petersburg as an adjunct in physiology at a salary of 200 rubles. Bernoulli rushed him, recommending that he come "this winter." It did not bother Euler that he would have to work in medicine. In those days medicine was not considered to be so remote from mathematics. We do not have to go far for examples: his teacher Johann Bernoulli alternated

[5] April 4th on the Julian calendar that was still in use in Russia.—*Transl.*

[6] Aleksei Krylov (1863–1945) was a leading Russian scientist who applied mathematics and mechanics to problems of shipbuilding and was a member of the Academy of Sciences.

between mathematics and practical medicine (and teaching Greek). Euler began studying anatomy and physiology, and later surprised those around him with his medical knowledge. He could not leave as quickly as Bernoulli wanted, but in the spring of 1727 Euler received "a draft in the amount of 130 rubles for passage" and left for Russia. He arrived in Petersburg on the day Catherine I died.

As did Daniel Bernoulli, within the realm of physiology Euler preferred to study hydrodynamic problems of blood circulation. To a significant extent these problems stimulated the creation of the field of hydrodynamics. In his first years in Petersburg, Euler hardly thought that his life would be so firmly connected to the Academy. The very existence of the Academy then seemed extremely problematical. Fuss later wrote, "Euler was the pride and glory of our Academy over a span of fifty years. Before his eyes it began its existence, died several times, and was resurrected." Euler felt very uncomfortable when the Academy's downfall seemed to him to be a reality. At one of the most difficult moments a mass exodus of academicians from Russia began after Peter II's death in 1730, and a desperate Euler discussed entering the naval service. But this was not needed. On the contrary, a vacancy opened up that allowed Euler to assume the position of professor (academician) with a chair in physics, although at the comparatively low salary of 400 rubles. Within two years Daniel Bernoulli abandoned Russia and Euler took his chair in mathematics, although his salary of 600 rubles was only half of what Bernoulli had received for the same position.

After this, Euler became an outstanding figure at the Academy. Most of the academicians were not too zealous about their responsibilities, which were not clearly defined. Euler did not neglect his duties in any way: he constantly gave reports at academic conferences, sometimes over two or three sessions in a row, gave public lectures, wrote a textbook on arithmetic for the Academy's gymnasium and wrote popular science articles for the *Notices* of the *Saint Petersburg Gazette*, was on commissions to study pumps for fighting fires, weights, a "sawing machine," and magnets, and took various examinations. Euler went into many technical projects in detail. Jumping ahead, we can recall his work on hydraulic turbines and his findings about projects for bridges across the river Neva, among them a single-arch wooden bridge by the inventor Ivan Kulibin, who worked in the mechanical workshop at the Academy.[7] Euler was always interested in Kulibin. There remains one moment in their relationship that is unclear. For 40 years Kulibin studied perpetual motion (a "self-propelled machine") and claimed that Euler did not rule out the possibility of creating such a

[7]Kulibin (1735–1818) had no education but suggested many innovations, some of which were realized.

machine (". . .perhaps in his time some lucky person will make such a machine and reveal it"). On the other hand, there is opposing evidence. We have to say that the Petersburg academicians were constantly considering projects on perpetual motion. But recall that in 1775 the Academy in Paris refused to consider such projects.

Beginning in 1733, Euler took part in "examining" maps and cartographic activities gradually became one of his main academic duties. The question arose of making a general map of Russia based on existing provincial maps, and Euler proposed his project saying, "I am certain that through my work and that of Professor Geinzius[8] Russian geography will be brought to a much better state than geography in the German lands." Sharp differences with Academician Delisle led Euler in 1740 to decide to curtail his cartography work. His health probably also played a role in this decision. On August 21st he wrote to Academician Goldbach, "Geography is killing me. You know that I paid for it with an eye and I now find myself in similar danger. When they sent me part of a map to look at this morning, I immediately felt a new attack since this work, which always requires looking over a large expanse all at once, tires my vision much more than simple reading or writing." Euler lost his right eye in 1735, when he completed a government assignment in three days for which the academicians needed several months. It is not completely clear whether this assignment involved cartography (we can understand Euler in this way) or astronomical calculations (as Condorcet wrote).

In 1740 there was another case where Euler avoided an assignment that was given to him (no other examples are known): he redirected constructing a horoscope for "Ivan Tsarevich," the soon-to-be emperor Ioann Antonovich, to the court astronomer. Incidentally, Pushkin told a different version of this story: "When Ioann Antonovich was born, Empress Anna Ioannovna sent Euler an order to construct a horoscope for the newborn. Euler at first refused, but was compelled to obey. He made a horoscope together with another academician. They constructed it according to the rules of astrology, as conscientious Germans, but did not believe it. The conclusion they reached frightened the two mathematicians, and they sent the empress another horoscope that predicted every happiness for the newborn. But Euler kept the first one and when Ioann Antonovich's unhappy fate came about he sent it to Count K. G. Razumovsky."

One should be constantly amazed that these many obligations still left Euler time for his main work, the study of mathematics. In these very years he took shape as a great scientist. Critically reexamining the work

[8]The astronomer Gotfrid Geinzius (1709–1769).—*Transl.*

of Leibniz and Newton in mathematical analysis and that of Fermat in number theory, he found his own path in science. Almost all his books and articles were published later, but the most important part of Euler's scientific destiny was decided in his first ten years in Petersburg. Only his fantastic capacity for work and astonishing purposefulness allowed Euler to combine the unobtrusive world of mathematical studies with day-to-day academic worries. He later wrote that a young scientist's specialty must "be his main subject and he [should] not... be distracted from it by any other work." According to Euler, that was possible for him in Petersburg: "For such a desirous situation that made his name well known Doctor Gmelin[9] is not the only one who is obliged to everyone, but also I myself and all others who had the pleasure of spending some time at the Russian Imperial Academy. We should recognize how obliged we are for the favorable circumstances in which we have just found ourselves. Concerning me personally, if this excellent situation had been absent I would have had to study mainly other sciences, from which according to all signs I only would have become dull. His Royal Majesty[10] recently asked me where I learned what I know. I answered truthfully that I owe all to my stay at the Petersburg Academy of Sciences."

In 1733, Euler married Katharina Gsell, the daughter of a Swiss academic painter who had been brought from Holland by Peter I. Of their 13 children, three sons and two daughters survived. For the pious son of a village pastor, his family was a fortress in which he could protect himself from the stormy tempers of the northern capital. A measured family life and small pleasures were necessary for Euler to work peacefully. No scientific work could make him neglect his family duties. For example, he was never indifferent to financial problems. He is quoted as saying, "Where they pay more, that is where I must go."

The year 1740 was probably the most difficult year in Euler's life. On the one hand, all the signs were positive: his academic salary reached its maximum, 1200 rubles (as much as Daniel Bernoulli had received). He understood a lot about Russian society and particularly about the subtleties of academic relationships. Field Marshal Burkhard Christoph von Münnich (Minikh) rendered him "the honor of his particular favor." Unlike most of the academicians, Euler was even on good terms with Johann Schumacher, the all-powerful head of the Academy. The most important thing for Euler was to be able to work quietly, and throughout his life he generally avoided conflict. Recall his good relations over many years with Johann Bernoulli, who constantly argued not only with other scholars but also with

[9] Johann Gmelin, a naturalist who arrived at the Academy a year after Euler.—*Transl.*
[10] Friedrich II ("Frederick the Great").

his brother Jacob and son Daniel.

On the other hand, the great scholar, nearing only the end of his thirty-third year, had managed to undermine his health through constant overwork. In 1740 he fell into a deep depression, which was connected not only to his health but also to constant strain because of the instability of political life in Russia. Euler's endurance had been enough to survive the Biron decade[11] but the new regency that was in place after the death of Anna Ioannovna frightened him. He recalled that "something dangerous was expected" and "after the death of the worthy Empress Anna and the regency that followed, things began to go bad." At that time the possibility arose of going to Friedrich II in Berlin, and Euler submitted a request to leave: "...I find it necessary, for the sake of my weak health and for other circumstances, to seek a more pleasant climate and to accept the summons made to me by his Prussian Royal Highness. Because of this I ask the Imperial Academy of Sciences to grant me the favor of releasing me and to supply me and my household with the necessary passport for travel...." He promised to maintain contact with the Academy and "as my health improves, to return to Russia from the German lands." Meanwhile, Euler wrote to Prussia that he had "firmly decided to live under the glorious rule" of Friedrich. On May 29, 1741 Euler was released from service, and later his request was met "to be made an honorary member of the Academy with a pension set at two hundred rubles a year." It was customary to convert departing academicians into honorary members with a commitment to render assistance to the Academy. From Euler they received a promise "through the usual correspondence and other mathematical pieces to serve more than when he was in active academic service."

In the Service of a "Royal Philosopher"

Thus, Euler came to Berlin. Friedrich II was not in the city. War had distracted him from his patronage of the sciences: in his own words, he was always at war with the "three whores," Maria-Theresa, Elizabeth, and the Marquise de Pompadour. Year after year, the opening of the Berlin Academy of Sciences was postponed (it opened in 1744). Meanwhile, the king sent his new geometer an affectionate letter from his camp at Reichenbach. He was attentive to Euler and invited him to a court ball. According to Condorcet, the Queen Mother was surprised by the scholar's terse answers to her questions: "But why do you wish not to speak with me at all?" The reply followed: "Madam, forgive me but I am unaccustomed to it. I

[11] Ernst Biron was Anna Ioannovna's favorite and a powerful figure at the Russian court.—*Transl.*

came from a country where he who speaks is hung." Euler was gradually drawn into the life of Berlin. His assignment here was no less than in Petersburg. He recommended books on ballistics to the king and published three volumes on this theme himself. He investigated correcting the level of the canal between the Havel and Oder rivers, studied the condition of the saltworks at Schönebeck, took part in organizing state lotteries and a reform of widows' insurance, and gave his opinion on a range of projects. They quickly understood that he could do a diverse range of things and refused him nothing. After organizing the Berlin Academy in 1744, Euler became the director of its Mathematics section.

But relations with the king took a form that was not the best. It is indicative that Euler's salary was half that of Pierre de Maupertuis (1698–1759), the President of the Academy. Euler rarely received praise from the monarch. Here is one of the few examples. Euler worked a lot on concrete problems in optics and in 1759 he made eyeglasses for Friedrich that were just right for him. Here is how the king complimented him: "...I cannot but praise your efforts to extract utility for people from the scientific studies that occupy your time. My work does not allow me to devote the required attention to your work at this time, but I will do it at the first opportunity." Euler tried to interest the king in differential equations, but without success. In 1747 Euler "inopportunely" published a tract against free-thinking, and his position was out of fashion at the Prussian court. At that moment the scientist felt uncomfortable: "I remark that the propensity for fiction begins here to prevail over mathematics, so that I fear that my person will soon be superfluous here." Euler thought about going to London.

In Berlin it was thought that scholarly duties included serving as decoration for the reception rooms, to make discussions more pleasant. The French scholars Maupertuis and the Marquis d'Argens (1704–1771) knew how to do this brilliantly, and Euler did not. D'Argens wrote to Friedrich about one of his colleagues: "The difference between his discussion style and Euler's manner is like the difference between Horace's essays and the works of the most learned and most pedantic Wolf." In 1746 Friedrich's brother August-Wilhelm met Euler and he gave the king his impressions: "...M. Maupertuis introduced me to the mathematician Euler. I found that in him the truth that all things are imperfect is confirmed. Thanks to diligence he has developed in himself logical thinking and acquired a name for it, but his appearance and his clumsy way of expressing himself obscure all his excellent qualities and interfere with gaining satisfaction from them." Friedrich answered, "Dearest brother! I had already thought that a conversation with M. Euler would not give you particular satisfaction. His epigrams consist of calculating new curves, some sort of conic sections,

or astronomical measurements. Among the scholars there are strong calculators, commentators, translators, and compilers who are useful in the Republic of Science but do not shine elsewhere. They are used like Doric columns in architecture. They belong to the lower story to support the whole building and the Corinthian columns are their decoration." What eloquent testimony to the views of an enlightened monarch about science and scientists!

Euler divided his time between science and home, but he did not belong to the category of scientists who are uninterested in external events and who avoid contact with people. His scientific knowledge was encyclopedic. He knew a lot about botany, chemistry, anatomy, and medicine, knew ancient and Eastern languages well, and spoke Russian. After his death it was recalled that he knew well "the best writers of the ancient world," "ancient mathematical literature," and "the history of all ages and peoples." Fuss wrote in his memoirs that Euler knew the *Anaeid* by heart and remembered which verses were first and last on each page of his copy. It is possible that this is not what was appreciated by the Prussian court, and also posthumous evaluations are always kind.

For some time Euler had become the butt of jokes made up by the king: "A certain geometer, who had lost an eye from his calculations, thought of composing a minuet using a plus b. If it had been performed for Apollo, the geometer would have risked being skinned like Marsyas." This may contain an allusion to a tract Euler wrote on the mathematical theory of music. The king found out that Euler did not stop his calculations while at the theater and the scholar became the hero of a new epigram. Incidentally, Euler did not appreciate theater but greatly enjoyed puppet shows.

Euler, who was solidly winning a reputation as one of the strongest and perhaps the strongest mathematician in Europe, was doomed to remain a second-rate person in Friedrich's circle. He once substituted as president of the Academy and after Maupertuis left he thought about taking that position. But the king intended the presidency for Jean Le Rond d'Alembert (1717–1783), a remarkable mathematician who was ten years younger than Euler. D'Alembert declined it, but this did not decide the question in favor of Euler. The "French peril" was one of the reasons for Euler's thinking about leaving Berlin.

At the same time in Russia, the situation of the Academy changed for the better with the coronation of Elizabeth. After a long absence from the Petersburg diary, this notice appeared in 1742: "The calm of the capital was varied by a few shows by the scholars meeting at the Academy of Sciences. Beginning on February 17 in Library Hall, twice a week from 10 to 12 o'clock, lectures in physics by Georg Kraft were open to the public,

and the number attending these discussions, which are coming into fashion, was significant. Classes in drawing from nature are also open there." The 18-year old Kirill Razumovsky, brother of the empress's favorite, was named President of the Academy. Before this the future president traveled to various university cities for two years and acquired diplomas. Euler's contacts with the Academy had not been interrupted. None of the honorary academicians took his duties so conscientiously. During his 25 years in Berlin Euler published 109 articles in the publications of the Petersburg Academy, and published 127 in Berlin during the same time. He performed various services for the Russian Academy: he was concerned with stocking the library, choosing themes for competitions for academic prizes, seeking candidates for vacancies in academic positions, and acquiring "magic lanterns" and fireworks for festivals at court (this was one of the most important jobs of the Academy). The intensity of Euler's correspondence with the Russian academicians is striking, especially with Schumacher, who ran the Academy.

At the start of the 1750s Euler established a *pension* in his house for his students. He combined working with them with teaching his eldest son Johann-Albrecht, and besides the income from the students was not unimportant for the tight family budget. Some of the first to come were students of the Academy's university, Semyon Kotel'nikov and Stepan Rumovsky, who were future academicians (a third student, Safronov, was sent home within a year because "he was so drunk that he could hardly be kept on"). Euler was constantly concerned about financial problems. He tried to have his family want for nothing. In 1753 he acquired an estate in Charlottenburg with a beautiful house, a garden, a large amount of arable land, 6 horses, and 10 cows. His father died in Switzerland and his mother came to live with her son. Euler went to meet her in Frankfurt-am-Main. His biographers have always wondered why he did not use this natural opportunity to visit his native Basel. Were the reasons sentimental or financial?

The Seven Years' War[12] made life more difficult. The currency was worth half of what it had been but salaries did not increase. The advancing Russian army destroyed the Charlottenburg estate. But Field Marshal Pyotr Saltykov, recognizing the owner's name, ordered immediate compensation for the damage. Later, Elizabeth herself added the enormous sum of 4000 rubles. These details testify to the special nature of Euler's relationship with Russia. He tried to maintain contact with Russia even during the war years. These were not only scientific contacts. For instance, in 1762 he requested that 3 centners of "Russian butter," a centner of "good

[12] In which Prussia and Russia were on different sides.—*Transl.*

white honey," "several poods of Vologda candles," etc., be sent to him via Stettin.[13]

After the end of the war in 1763, Euler thought more decisively about returning to Russia. As early as 1746 and 1750 he had received invitations through Razumovsky, but politely put off the decision for an indeterminate length of time. Euler hardly went anywhere in 1763 and d'Alembert took upon himself the unexpected role of intermediary in negotiations with the king. He was evidently able to convince both parties, because in August he wrote to Euler: "Finally, I am happy that I have held onto a person such as yourself for the king and the Academy." Within a week, he wrote in another letter: "I have completely persuaded His Majesty that in you the Academy will bear an irretrievable loss which will inflict a blow on the the king's reputation. I daresay that before my departure I will be entrusted with bringing your interests to his attention." Euler was not allowed to leave, but within two years a scandal broke out. Euler provoked the king's anger by interceding for the Academy treasurer during an inspection.

Negotiations over his departure were begun again with new zeal and Catherine II, who had succeeded to the Russian throne, very much wanted to obtain Euler for St. Petersburg. Euler communicated his conditions: a salary of 3000 rubles (the same as the president, while salaries for academicians did not usually exceed 1200 rubles), an academic position in physics for his son Johann-Albrecht, suitable positions for his other sons (an artilleryman and a doctor), an apartment where he would be free from quartering soldiers, and, finally, establishing for himself the position of vice-president with the corresponding rank. Euler was not able to become president of the Berlin Academy and wanted at least in part to realize his ambitious plans in Petersburg (he did not lay claim to the presidency, believing that in Russia this had to be held by a nobleman). The academician Goldbach, Euler's friend (more about him below), worked in the Ministry of Foreign Affairs with the high rank of Privy Councilor (the civilian equivalent of a general). Apparently Euler wanted to become a general in his declining years. On January 6, 1766 Catherine wrote to the Chancellor, Count Mikhail Vorontsov: "M. Euler's letter to you gave me great satisfaction because I recognized in it his desire once again to enter my service. Of course, I find the desired title of Vice-President of the Academy of Sciences completely merited, but this will require taking certain measures. First of all I will establish the title. I say will establish, since it has not existed up to now. As things are currently there is no money for a salary of 3000 rubles, but for a man with such qualities as M. Euler I will supplement his

[13] 1 centner $\approx$ 100 kg and 1 pood $\approx$ 16.4 kg.—*Transl.*

academic salary from state funds so that together they will make up the required 3000 rubles. He will have a state apartment without the slightest shadow of a soldier. Although there is no free chair in physics at the Academy with a salary of 1000 rubles for his eldest son I will endow it, and I will also authorize a free practice for the second (the physician) and will give [him] a position, if he will desire to enter service. The third son (the artilleryman) will be accommodated without any difficulty.... I am sure that my Academy will be restored from the ashes by such an important acquisition, and I congratulate myself in advance that a great man has returned to Russia." Recognizing Euler's desire to take part in rebuilding the Academy, the empress promised "to make no changes in the Academy before his arrival, to the end of better discussing improvements with him...." With great diplomatic mastery, they refused Euler the rank: he could only have the rank of Collegiate Councilor, the civilian equivalent of a colonel, which was unworthy of the great scholar: "I would have given him the rank when he wanted it if there had not been the risk that this rank would compare him with a number of persons who do not stand up to M. Euler. In view of his fame, a rank that would render him the respect he is due would be better." Euler probably understood quickly that the generous empress knew how to explain clearly the limits of what was permitted, agreed to all the conditions, and resolved "to finish my days in the service of this incomparable sovereign."

It turned out that Friedrich was not inclined to part with his geometer. In particular, he used the possibility that he would keep the scientist's son in the army. All the same, Euler obtained permission to leave. Later, the king used Euler as a target for his wit for the last time: "Herr Euler, who madly loves the Great and Little Bears, went north so as better to observe them. The boat, weighted down with his *XX* and his *KK*, was shipwrecked. All was lost, which is a pity because because there was enough there to fill six volumes of articles dotted with numbers from beginning to end. In all probability, Europe is missing the pleasant amusement it would have had in reading them" (from a letter to d'Alembert). Soon Friedrich consoled himself by obtaining the young Joseph-Louis Lagrange (1736–1813) in place of Euler, instructively justifying the expediency of his move to Berlin: "The greatest geometer in Europe must live near its greatest king."

In Russia Again

Euler arrived in Petersburg on July 17, 1766. He had been away for exactly 25 years and was nearing 60. From the start Euler took seriously Catherine's

proposal to participate in reorganizing the Academy. He brought a detailed design with him and did not lean towards autonomy for the Academy but rather towards binding the activities of the Academy tightly to those of government institutions. But it gradually became clear that the empress was not inclined to trust Euler with the leadership of the Academy. Euler learned once more the lesson that enlightened monarchs want their scholars to know their place. As the most senior academician, the dean, he had no small influence on academic life, but no one remembered the promised position of vice-president. And as the head of the Academy Catherine installed (following Elizabeth's tradition) the younger brother of her favorite, Count Vladimir Orlov. But a small discrepancy arose: Razumovsky still occupied the presidency, and as commander of the Ismailov regiment he had supported Catherine during a revolt. He could not be insulted, so they established the post of Director of the Academy for Orlov. The new director did not relate badly to Euler: he worried about his health and obtained medicine for him but could play tricks on the old man, pretending to be a poor Swiss supplicant "to check [Euler's] vision." There was a conflict not long before Orlov's departure in 1774, after which Euler stopped coming to meetings at the Academy. However he continued to be interested in its activities and the academicians frequently held sessions in Euler's apartment.

Euler brought a stack of manuscripts with him to St. Petersburg which he had not been able to publish in Berlin, since publishing had almost ceased during the war. But he brought even more ideas in his head that were almost ripe but were still unrealized. And life told Euler that he should hurry. Soon after his arrival he lost vision in his second eye but did not stop working and dictated his essays to his son, who had not the slightest idea about mathematics. The imperial oculist Baron Michael de Wenzel removed a cataract from one eye but predicted that overwork would inevitably lead to a return of Euler's blindness. That soon happened, but Euler preferred losing his vision to passivity. He tried to direct the work of others, such as his son and Academicians Kraft, Fuss, and Lexell, but most of all he dictated what he knew and wanted to impart to others. During half a decade he dictated more than 400 articles and 10 large books. Deafness became added to blindness. His wife died in 1766 and Euler married her sister (that was the simplest way to maintain order at home). His house and a large part of his property caught fire but nothing could make Euler interrupt his work. During the summer of 1777 Johann Bernoulli III (1747–1807), Daniel's nephew, visited Euler. Here are Johann's impressions: "His health was rather good, for which he was obliged to follow a measured and correct way of life. As for his vision, which is for the most part lost and

was at one time completely gone, he now uses it better than many imagine! Although he cannot recognize anyone's face, read black on white, or write with a pen on paper, he writes his mathematical calculations very clearly and orderly with chalk on a black table and at the usual size. Then they are written down in a large book by one or another of his assistants, Fuss or Golovin[14] (more often by the former). And articles are prepared from these materials under his direction. Thus in the period of five years during which M. Fuss lived in Euler's house, 120 or 130 articles were brought to completion."

Euler maintained his ability to work until his final days. His second Petersburg period lasted 17 years. In 1783 the son of the village pastor finished his days as the greatest mathematician in Europe. Euler was buried in the Smolensk cemetery. The inscription on his monument reads: "Here rests the mortal remains of the wise, just, illustrious Leonhard Euler." After 50 years it was discovered that the monument had disappeared and it was found only by chance, during the burial of his daughter-in-law, where "the stone had sunk little by little into the earth from its own weight and because of the powdery turf." They felt uncomfortable at the Academy and resolved to establish a new monument "worthy of the famous geometer." Later Euler's remains were reinterred in the necropolis of the Alexander Nevsky Monastery, where his monument can be seen today.

A Great Legacy

Euler's scientific legacy is astounding in its completely unprecedented volume. During his life, 530 books and articles saw the light of day. In his last years his academic output did not resemble the scientific productivity of a blind scholar, and he jokingly promised Count Orlov that his works would fill the Academy's *Commentaries* for 20 years after his death. This estimate turned out to be "optimistic": it took the Academy 47 years to publish Euler's works. The number of items reached 771, but Gustav Enestrom's bibliography, composed in 1910, contained 886 titles broken down by these rubrics: philosophy, mathematics, mechanics, astronomy, physics, geography, and agriculture. Since 1910, the Swiss Society of Naturalists has been publishing a collection of Euler's writings, widely distributed internationally through subscription. It had been estimated that Euler's *Opera Omnia* would occupy 75 large volumes. So far 76 have been published, 30 of them in mathematics, and the project is nearly complete. His scientific

[14]Mikhail Golovin.

correspondence was expected to fill 8 additional volumes, and this work is ongoing.

This scale reflects not only the amazing speed at which Euler worked but also his habit of systematically writing scientific texts, including ones that were prepared relatively hastily. The great range of topics reflects not only the breadth of his interests and his ability to go quickly to the far corners of science, but also the many academic obligations he had in Petersburg as well as Berlin. Some publications are like brief remarks. Euler easily entered into scientific contacts, gave advice on various questions, and willingly thought about random isolated problems sent to him by his correspondents. It might look as if he spread himself too thin and was omnivorous, but that would only be at first glance. Random questions and problems served as fertile soil for well-planned thought. Euler could pause in his thinking at opportune times if he saw no realistic possibility of moving ahead. He could organize things so that the many daily affairs of life did not seriously divert him from the fundamental direction of his work.

How paradoxical it is to say, without much exaggeration, that Euler worked almost exclusively on mathematics all his life. In other areas of science (e.g., mechanics or astronomy) his success was mostly related to applying mathematical methods. His philosophical stance throughout his whole life was that discoveries in the natural sciences should be obtained via a theoretical (to a significant extent mathematical) treatment of a few general, unquestioned principles. In his Swiss dissertation, the nineteen-year old Euler wrote: "I did not think it necessary to confirm this new theory by experiment because it was completely derived from the most reliable and irrefutable principles of mechanics and thus there simply can be no doubt as to whether it works in practice." Euler even tried to derive Newton's laws from more general principles, and in celestial mechanics he tried not to obtain empirical formulas from observational data but rather to derive them directly from the law of universal gravitation. He always tried to go from theory to application. Although Euler was, in fact, involved with experiments all his life, this was not his major direction. Sergei Vavilov[15] wrote, "...Euler's genius was in essence mathematical... he did not have a good feeling for experiments (although he did do experiments)...." Elsewhere he wrote, "Euler's mathematical genius did not have the physical intuition of Newton and Huygens that allowed them to guess the solution without a precise mathematical statement of the problem or methods for solving it."

[15](1891–1951), a Russian physicist who was president of the Academy of Sciences.

Arithmetic

Turning to Euler's mathematical legacy, it is natural for us to begin with his arithmetical works. Euler's first publications are from 1732, his fifth year in St. Petersburg. Euler had two great predecessors in arithmetic, Diophantus and Fermat. If we bypass the prehistory associated with the name of Diophantus (3rd century), then Pierre Fermat (1601–1655) was the first to discover that arithmetic contains not only surprising facts about specific numbers but also general assertions—theorems. Fermat left the statements of many such theorems in the margins of Diophantus's *Arithmetica* (opportunely published in 1621) and in letters and notes. Fermat was one of the best mathematicians of his time, was at the very center of the heroic epic of the creation of analysis and analytic geometry, and corresponded with the leading mathematicians. Significantly, he was unable to generate serious interest in arithmetic problems among any of his most earnest correspondents. He found interested collaborators only among mathematicians of lesser calibre such as Frenicle de Bessy (1605–1675). For reasons that are hard to unravel, some scientific theories fascinate everyone (e.g., analysis in the 17th century) while others are worked on by individual scientists who try in vain to attract their colleagues' attention. Think of projective geometry, which was created by Girard Desargues (1591–1661) and Blaise Pascal (1623–1662) (who were far from unknown), and which was forgotten for a century and a half and rediscovered by Gaspard Monge (1746–1818) and his students. In the 1670s Fermat's notes were partially collected and published, but it is hard to imagine the fate of Fermat's arithmetic if it had not been for Euler.

In 1849, Pafnuty Lvovich Chebyshev (1821–1879) wrote: "All the research making up the general theory of numbers begins with Euler. In this research Fermat preceded Euler.... But the research of this geometer had no direct influence on the development of the science: his propositions remained without proof and without application. In this situation Fermat's discoveries served only as a challenge to geometers to investigate the theory of numbers. But despite the general interest of this research, no one accepted the challenge until Euler. And this is understandable: this work did not require new applications of methods that were already known or new advancements with methods that had been used before. It required the creation of new methods, the discovery of new starting points, in a word the foundation of a new science. This was done by Euler."

Soon after Euler arrived in St. Petersburg in 1727 he evidently heard about Fermat's work from Christian Goldbach (1690–1764), and he maintained his interest in number theory all his life. Euler's prominent col-

leagues had something to do with this interest, at least unknowingly. Daniel Bernoulli (1700–1782), who himself did not mind studying arithmetic problems, wrote in 1778 to Euler's student Fuss about his teacher's arithmetic works: "...do you not find that the prime numbers are given, if you will, too much honor, so that so much energy is wasted on them, and does this refined taste of our century not repel you?" Euler discussed arithmetic problems first of all with Goldbach, a very original mathematician who still was not among Euler's strongest contemporaries such as Jean Le Rond d'Alembert (1717–1783) or Alexis Clairaut (1713–1765).

The situation changed only towards the end of Euler's life, when thanks to his work attitudes towards number theory began to change and he was able to discuss these problems with Lagrange in the letters of 1772–1773.

As early as 1729, Euler learned from Goldbach about Fermat's claim that the numbers $F_n = 2^{2^n} + 1$ are prime for all n. In 1732 he discovered that this was false, and in particular that F_5 is divisible by 641. Euler's observation was not the result of sorting through all the possibilities: finding the divisors of F_5 directly was unrealistic even for such a computational virtuoso as Euler. He discovered first that the divisors of F_n have the very special form $k \cdot 2^{n+2} + 1$ if they exist, and then it was not hard to find that $641 = 5 \cdot 2^7 + 1$. It is surprising that Euler's first run at proving Fermat's confident assertions led him to the only erroneous one. Happily, this did not shake his belief and interest in Fermat's arithmetic.

A second class of prime numbers in Euler's field of vision were the Mersenne primes $M_p = 2^p - 1$, where p is prime. The divisors of M_p must simultaneously have the forms $2pk - 1$ and $8l \pm 1$. Using this, Euler proved that $M_{31} = 2{,}147{,}483{,}647$ is prime. Since that time new Fermat primes (primes of the form F_n) have not been found, but the world record for Mersenne primes has steadily increased. The record in 1983 was $p = 86{,}243$ and today computers find Mersenne primes with unbelievably many digits.[16]

Regarding Mersenne primes, Euler filled a gap that had been left open since Euclid. Euclid knew that if M_p is prime, then $\frac{M_p(M_p+1)}{2}$ is a perfect number, i.e., a number that is the sum of its proper divisors. Euler proved that every even perfect number has this form (it is still unknown if there exist odd perfect numbers). Euler was interested in whether there exist polynomials $P(n)$ that take prime values for all natural numbers n. He answered this negatively but noted that the values of the polynomial $41 - n + n^2$ are prime for all $n \leq 40$.

[16] In December 2005 it was found that $2^{30{,}402{,}457} - 1$, which has 9,152,052 digits, is prime. This is the 43rd Mersenne prime that has been found. More details are at http://www.mersenne.org/prime.htm.—*Transl.*

Euler supplied a proof of "Fermat's Little Theorem," which states that $a^{p-1}-1$ is divisible by p, where p is prime and a is an integer not divisible by p. Not stopping there, he found and proved its generalization to nonprime divisors: if a and m are relatively prime, then $a^{\varphi(m)}-1$ is divisible by m. Here $\varphi(m)$, called the Euler phi function, is the number of natural numbers relatively prime to m and less than it. When p is prime we have $\varphi(p)=p-1$. In discovering that this function, which is defined on the natural numbers, has remarkable properties he opened an important chapter in number theory, namely, the theory of arithmetic functions. Euler proceeded very logically. He noticed that for certain values of a the number a^k-1 is divisible by p for $k<p-1$ and for others it is not. In the latter situation a is called a primitive root modulo p. An experiment convinced Euler that primitive roots exist for all primes p, but he was unable to prove this. Later on, Legendre and Gauss found proofs. Euler knew how to prove difficult theorems but he knew his capabilities and appreciated them soberly. He never concentrated his thinking on one difficult problem for years but instead advanced along a broad front against the secrets of mathematics.

Another assertion that was stated by Fermat without proof attracted Euler's attention. This concerned expressing squares n^2 in the form $kp-1$, where p is prime. For $p=3$ such squares do not exist (why?) and for $p=5$ we have $2^2=5-1$. Fermat claimed that for each prime p of the form $4l+1$ there exists a square of the form $kp-1$ and for $p=4l-1$ such squares do not exist. In 1747, after several unsuccessful attempts, Euler proved Fermat's claim and then continued in a natural direction: for which p can $kp+2$ be a square and, more broadly, for which p can $kp+a$ be a square, where a is a fixed number? For $a=2$ the conjecture is that squares of such form exist for $p=8l\pm1$ and not otherwise. The general conjecture is that squares of the form $kp+a$, where p is prime, exist (we say that a is a quadratic residue modulo p) or do not exist (a is a quadratic nonresidue) simultaneously for all prime p in the arithmetic progression $b+4ak$, $k=1,2,3,\ldots$. This statement was later called the "law of quadratic reciprocity." Euler could prove it only for $a=3$, apart from $a=-1$. Lagrange and Legendre later considered different values of a until the 19-year old Gauss found a complete proof of Euler's conjecture (this is presented in the chapter on Gauss).

The next circle of questions inherited from Fermat was the solution of equations in integers. Fermat's most famous claim, "Fermat's Last Theorem," is that the equation $x^n+y^n=z^n$ has no solution in positive integers for natural numbers $n>2$ (for $n=2$ such solutions exist and are called Pythagorean triples). In 1738 Euler found a proof of Fermat's Last Theorem for $n=3,4$ but he gave up trying to prove it for larger values of n, despite Fermat's unmotivated statement that there exists a proof for arbitrary n.

Fermat's Last Theorem was proved by Andrew Wiles in 1995.

Once Fermat proposed to Frenicle and Pierre Brulart de Saint-Martin to construct a right triangle with integer sides for which the sum of the legs and the hypotenuse are both squares, i.e., to solve this system of equations in integers: $x+y=u^2$, $x^2+y^2=v^4$. Fermat suspected that he had posed an "impossible" problem. Euler studied the system, which is remarkable in that its smallest solution consists of the 13-digit numbers 1,061,652,293,520 and 4,565,486,027,761.

Euler considered the equation $x^2-Dy^2=1$, $D\neq a^2$, which he called the Pell[17] equation. He discovered a connection between its smallest solution and the expansion of $\sqrt{D}$ as an infinite continued fraction. Many examples convinced Euler that he had found a periodic continued fraction, but a proof of this was found only later by Lagrange.

Fermat claimed that every prime of the form $4k+1$ can be written as the sum of two squares, and that this can be done in only one way (it is easy to show that primes of the form $4k+3$ are not sums of squares). Euler established that the converse is true: if a representation of a number N as the sum of squares exists and is unique, then N is a prime number. He showed that this property can sometimes be used to prove that a number N is prime. For example, 1,000,009 is a composite number since in addition to its representation as 1000^2+3^2 it can be written as 235^2+972^2. Furthermore, Euler showed that the forms x^2+2y^2 and x^2+3y^2 have an analogous property. Primes of the form $8m+1$, $8m+3$ can be written uniquely in the form x^2+2y^2, and numbers having more than one representation of this form are composite. Analogously, only prime numbers admit a representation as x^2+3y^2 and these have the form $6m+1$. After this Euler moved on to the general problem: is it true that N admits a unique representation of the form x^2+Dy^2, where D is fixed, if and only if N is prime? This claim turned out to be true for all $D\leq 10$ but for $D=11$ there is a composite number with a unique representation. The situation intrigued Euler. He called a number D convenient if only prime numbers can be written as x^2+Dy^2. Euler obtained a criterion for verifying that numbers are convenient, and out of curiosity wrote down convenient numbers one after the other: after 10 comes 12, 13, 15, 16, 18, 21, 22, 24, ... The convenient numbers gradually become rarer. There are 62 in the first thousand numbers but Euler persisted in continuing his calculations, probably hoping to notice some kind of pattern. He discovered only three more convenient numbers, namely, 1320, 1365, 1848, although without losing patience he checked all numbers up to 10,000 and then some. Euler had grounds for hypothesizing that the

[17] Named for John Pell (1611–1685).—*Transl.*

set of all convenient numbers is limited to the 65 that he had found. Gauss carried out Euler's study in a more reasonable way but did not find any new convenient numbers. It has now been proved that the set of convenient numbers is finite, but it is unknown whether there are any beyond 1848. This work is very characteristic of Euler's creative method, performing huge computational experiments to verify a conjecture and with the aim of seeing new patterns. Among the great mathematicians perhaps only Gauss completely mastered this inductive method.

With this we complete our survey of that part of Euler's work in arithmetic where he followed Fermat. He included Fermat's claims in a picture of the multiplicative theory of numbers, i.e., concepts related to divisibility, that was well thought out, and he correctly foresaw practically all of its basic theorems and problems. The proofs of some of the key assertions were left for Euler's successors. We can see the features of Euler's scientific style in some examples. Before him lay several excellent problems on which one could have concentrated for years if not for a lifetime, but no specific problem took priority with Euler over recreating the whole picture, over the irrepressible desire to move forward. He constantly returned to unsolved problems and knew how to ration the time he devoted to one problem or another. The difficulty of the problems that arose and the awareness that he needed to give up on obtaining a strict proof led Euler to formulate a method for establishing mathematical truths apart from proofs. An experiment was first and foremost not only for thinking through a problem or conjecture. A thorough numerical experiment on a large scale was, in Euler's internal value system, sometimes equivalent to establishing the truth. He spoke of "discovered but not proven truths" and tried to have this kind of argument accepted in mathematics. Obtaining a strict proof remained an important goal for Euler, but at some stage he would knowingly stop searching for one and instead turn to careful heuristic considerations.

Analytic Number Theory

Number theory is obliged to Euler for an idea that soon changed its face completely, namely the application of mathematical analysis to arithmetic. It was hard to imagine this possibility. At first it surprised Euler himself: "Although we are considering here the nature of integers, to which the Infinitesimal Calculus does not seem applicable, nevertheless I came to my conclusion with the help of differentiation and other devices."

For different values of s, Euler considered the sum of the infinite series

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots + \frac{1}{n^s} + \cdots . \tag{1}$$

This was later called the Riemann zeta function and plays an exceptional role in arithmetic. Using an argument that was not rigorous, Euler proved that this infinite sum coincides with an infinite product based on prime numbers,

$$\zeta(s) = \left(1 - \frac{1}{2^s}\right)^{-1} \left(1 - \frac{1}{3^s}\right)^{-1} \left(1 - \frac{1}{5^s}\right)^{-1} \cdots \left(1 - \frac{1}{p^s}\right)^{-1} \cdots . \quad (2)$$

His reasoning went as follows: for $s > 0$ the factor $(1-p^{-s})^{-1}$ can be thought of as the sum of the infinite geometric progression $1+p^{-s}+p^{-2s}+p^{-3s}+\cdots$. Multiplying these infinite sums together for all prime p and considering only the products of terms that equal 1 for all but a finite number of p, we arrive at the infinite sum (1). We must add a lot here to make this argument rigorous, beginning with the idea of the sum of an infinite number of terms and the product of an infinite number of factors. Euler did not include this. He felt that these considerations led to an exceptionally serious arithmetic result, but he himself could only produce a new proof of the theorem that the set of primes is infinite, a result that goes back as far as Euclid. Jacob Bernoulli knew that the sum of the n terms $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}$ goes to infinity as $n \to \infty$, i.e., that $\zeta(s)$ goes to infinity as $s \to 1$, which cannot be obtained from the product (2) if the number of different p is finite. You can say that the mountain gave birth to a mouse, but Euler's intuition did not mislead him. This became clear when Dirichlet proved that the number of primes in an arithmetic progression whose first term and difference are relatively prime is infinite, a generalization of Euclid's theorem, starting with Euler's proof described above (Euclid's proof does not extend to arithmetic progressions beyond $1, 2, 3, \ldots$).

Euler opened the door to yet another mystery in the world of prime numbers. His analytical intuition, which was far ahead of his technical capability, suggested that when x is large $\sum_{p<x} \frac{1}{p}$ is close to $\ln \sum_{n<x} \frac{1}{n}$, which is the first step in finding the distribution of the primes within the natural numbers. Euler felt that the function $\zeta(s)$ can be extended even to those values of s for which it cannot be defined as the sum of a series. Moreover, he noted a connection between the values of ζ at the points s and $1-s$ that would later be stated by Riemann in the form of a remarkable functional equation. Euler studied the values of $\zeta(s)$ at the integers. We will describe below how he figured out the case where the argument is even and how he expected to apply the symmetry between s and $1-s$ to study ζ at the odd points. But he failed, understanding that at the negative even points the continuation of ζ equals zero. We note that only recently a little has been learned about the arithmetic nature of the values of the zeta

function at the odd points: in 1979 it was proved[18] that $\zeta(3)$ is irrational and in the summer of 2000 it was announced[19] that the set of values of the zeta function at the odd integers contains an infinite number of distinct irrationals.

Series and Infinite Products

Infinite sums and infinite products were Euler's beloved objects in analysis. Newton frequently used infinite sums (series) and particularly power series $a_0 + a_1x + \cdots + a_nx^n + \cdots$, for example, in studying the binomial $(1 + x)^\alpha$ for noninteger α. Without really stressing this, Newton had in mind series for which the sums of n successive terms converge, as with decreasing geometric progressions. Although Euler well understood that a series might not be summable, he dared to work with series without worrying about convergence: he multiplied and divided series formally, differentiated them term-by-term, etc. This foreshadowed modern work in algebra with formal series. Not limiting himself to formal operations, Euler wanted to assign numerical values to divergent series. More than once his successors condemned him for what were, in fact, dubious claims such as $1 - 3 + 5 - 7 + \cdots = 0$ and $\cdots + \frac{1}{n^3} + \frac{1}{n^2} + \frac{1}{n} + 1 + n + n^2 + n^3 + \cdots = 0$. On the other hand, Euler took the partial sums of the harmonic series $1+\frac{1}{2}+\frac{1}{3}+\cdots+\frac{1}{n}$ and noted that if we subtract $\ln n$, then the difference will approach the constant 0.577216 …, which is now called Euler's constant. This is an important example of how the nature of divergence arose. Euler felt that divergent series were unavoidable in mathematics, and although he did not have the necessary apparatus his astounding intuition protected him against drawing erroneous conclusions from nonrigorous reasoning. At the same time his imitators, who did not have such powerful defenses, committed more than a few mistakes and absurdities.

Euler looked at infinite series as polynomials of infinite degree and by analogy formulated rules for expanding them into infinite products of linear factors. If the sum of a series $1 + a_1x + a_2x^2 + \cdots$ equals zero at the points $\alpha_1, \alpha_2, \ldots, \alpha_n, \ldots$, then it coincides with the infinite product $(1 - \frac{x}{\alpha_1}) \cdots (1 - \frac{x}{\alpha_n}) \cdots$. Euler gave no basis for this claim and did not state it rigorously but went straight to examples. He started with the infinite series

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots = x\left(1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots\right).$$

[18] By Roger Apéry.—*Transl.*
[19] By Tanguy Rivoal.—*Transl.*

Its sum is zero for $\alpha_{\pm k} = \pm\pi k$, which leads to

$$1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \cdots = \left(1 - \frac{x^2}{\pi^2}\right)\left(1 - \frac{x^2}{4\pi^2}\right)\left(1 - \frac{x^2}{9\pi^2}\right)\cdots .$$

By formally multiplying out the expressions in brackets, collecting the coefficients of x^2, and comparing the result with the coefficient in the series on the left, we obtain

$$1 + \frac{1}{4} + \frac{1}{9} + \cdots + \frac{1}{n^2} + \cdots = \frac{\pi^2}{6}.$$

This is the value of the zeta function at $s = 2$. Jacob Bernoulli had already investigated this series but could not find its sum. Euler had considered it for a long time. First he found its sum to seven decimal places, 1.6449340, and later calculated it to eight more places. While understanding that his computations were, strictly speaking, invalid Euler first of all found $\frac{\pi^2}{6}$ to seven places and compared it to the answer he knew. The results agreed! This happened in 1735. Comparing coefficients for more terms in the series and product, Euler easily found that $\zeta(4) = \frac{\pi^4}{90}$ and $\zeta(6) = \frac{\pi^6}{945}$. He understood that $\zeta(2n) = c_n\pi^{2n}$ and was interested in the nature of the coefficients c_{2n}. Euler obtained recurrent relations that were enough to compute them but he was not satisfied with that.

At almost the same, Euler was concerned with another numerical sequence that arose in a completely different problem. He wanted to apply integrals to estimate the sum of a large number of terms $S(n) = f(1) + f(2) + \cdots + f(n)$. He obtained the following formula, now known as the Euler–Maclaurin formula:

$$S(n) = \int_0^n f(x)dx + \frac{f(n)}{2} + \frac{f'(n)}{12} - \frac{f'''(n)}{720} + \frac{f^{(5)}(n)}{30240} + \cdots ,$$

and so on for successive derivatives. These were mysterious coefficients that Euler knew how to compute but for which he did not know a simple pattern. How surprised he was when he discovered that the coefficients were equal to $\frac{(-1)^{n-1}c_n}{2^{2n-1}}$. Nature only bestows such surprising coincidences on the greatest mathematicians! After all, there is no direct connection between these problems. But then Euler recalled a remarkable numerical sequence B_n that Jacob Bernoulli found in computing the sums of the kth powers of the first n natural numbers. The numbers B_n are now called the Bernoulli numbers, and it turned out that $B_n = \frac{(-1)^{n-1}(2n!)c_n}{2^{2n-1}}$. Moreover, in the expansion of $\frac{z}{e^z-1}$ in powers of z, $\frac{B_n}{n!}$ equals the coefficient of z^n. The

Bernoulli numbers were known before Euler, but he was the first to see that they can arise in a hidden way in the most different problems.

Euler constantly worried that his calculation of $\zeta(2n)$ was unfounded. He thought of another argument that strengthened the conclusions he drew from his numerical experiments. Among the examples he considered was one based on the expansion of $1-\sin x$ in a series and in an infinite product. He arrived at the relation $\frac{\pi}{4} = 1-\frac{1}{3}+\frac{1}{5}-\frac{1}{7}+\cdots$, which Leibniz had already derived rigorously directly from the geometric definition of π. Euler placed a very high value on the fact that these results agreed with each other: "This is a great confirmation for our method, which some might see as insufficiently reliable. Therefore we should not at all doubt the other results to which the same method leads." Euler insisted that unproven claims should be trusted seriously if they passed experimental trials and were confirmed indirectly. He thought that in the state of affairs that was then current, mathematics would lose a lot by holding narrowly to the Euclidean rules for establishing truth. By the way, he did not stop searching for a rigorous foundation and within 10 years found an essentially simpler basis for the expansion of $\sin x$ (based on the connection between trigonometric and exponential functions in the complex domain).

Euler continued to work with infinite products. He calculated the series corresponding to the infinite product $s(x) = (1-x)(1-x^2)(1-x^3)\cdots$ and noted that many powers are missing from it:

$$s(x) = 1 - x - x^2 + x^5 + x^7 - x^{12} - x^{15} + x^{22} + x^{26} - x^{35} - x^{46} + \cdots .$$

The signs of the nonzero terms change every other term. For Euler, it was easy to unravel the pattern for the sequence of exponents of nonzero terms. He considered the sequence of differences $1, 3, 2, 5, 3, 7, 4, \ldots$ and divided it into the sequence of natural numbers and the sequence of odd integers. He then obtained this representation for the original sequence of exponents: in the kth pair of exponents the powers are $m = \frac{1}{2}(3k^2 \pm k)$, and the sign of x^m is $(-1)^k$. However, Euler did not succeed in proving that the infinite product and series are equal, even in a formal way: "I searched in vain for a strict proof of the equality between this series and the infinite product $(1-x)(1-x^2)(1-x^3)\cdots$, and I proposed this question to several of my friends whose abilities in this regard I know and all agreed with me that this transformation of a product into a series is valid, but no one could unearth any sort of key to the proof. Thus this known but unproved truth. . . ." Incidentally, the numbers of the form $\frac{3k^2-k}{2}$ were known by the Greek mathematicians, at least by Nichomachus in the first century. These are the so-called pentagonal numbers.

Euler arrived at the problem we are discussing by starting with another one. Let $a_m(b_m)$ be the number of representations of a natural number m as the sum of an even (odd) number of different terms. By analyzing how the term x^m arises in multiplying by $(1-x)$, $(1-x^2)$, ..., it is not hard to see that the coefficient of x^m is precisely equal to $a_m - b_m$. This means that the claim Euler was trying to prove is equivalent to saying that $a_m = b_m$ for all m other than $\frac{3k^2 \pm k}{2}$ and that for these numbers $|a_m - b_m| = 1$ (we can even specify the sign). It was this claim that interested Euler, and considering infinite products and series was only a means of proving it.

Euler associated yet another remarkable arithmetic statement with the series $s(x)$. This was a claim about $\sigma(n)$, the sum of the divisors of n. Working with $\frac{s'(x)}{x}$, Euler obtained

$$\sigma(n) = \sigma(n-1) + \sigma(n-2) - \sigma(n-5) - \sigma(n-7) + \cdots .$$

Euler called this "an extraordinary law of numbers in relation to the sum of their divisors." Seeing no way to prove it directly, he verified the law for $n \leq 20$ and then for $n = 101$ (a prime number) and $n = 301$, and wrote: "These examples that I have just developed will no doubt remove any scruple which one could still have about the truth of my formula. But one could be all the more surprised by this nice property, not seeing any connection between the composition of my formula and the nature of the divisors, the sum of which the proposition centers upon."[20]

Additive Number Theory

Problems about the number of representations of a natural number as a sum of terms of some nature (as Euler said, problems about "partitions of numbers") were at the center of his attention for a long time. His first push in this direction may have come from problems contained in a letter of Phillipe Naudé (1684–1745), whose name is unknown today.[21] Euler applied the machinery of infinite products to these problems. Here are some examples. Euler claimed that

$$(1+x)(1+x^2)(1+x^3)\cdots = \frac{1}{(1-x)(1-x^3)(1-x^5)\cdots}.$$

The argument consists of the fact that if we multiply the left-hand side sequentially by $(1-x)$, $(1-x^3)$, $(1-x^5)$, ..., then all nonzero degrees will

[20] Taken from the English translation by Todd Doucet of paper E175 in the Euler archive at http://math.dartmouth.edu/~euler.

[21] It is notable that Euler started not only with great sources, as in the case of Fermat, but sometimes with completely random problems.

gradually vanish and this implies the identity (this argument can be made rigorous by using the theory of limits). After removing the parentheses on the left we obtain the series $1 + a_1x + a_2x^2 + \cdots$, where a_k is the number of representations of k as a sum of distinct natural numbers. The right-hand side can be written as follows, using the sum of an infinite geometric progression:

$$(1 + x + x^2 + x^3 + \cdots)(1 + x^3 + x^6 + x^9 + \cdots)(1 + x^5 + x^{10} + \cdots)\cdots,$$

and equals $1 + b_1x + b_2x^2 + \cdots$, where b_k is the number of representations of k as the sum of odd terms, some of which can be the same (why?). Euler concluded that the numbers of representations coincide, i.e., that $a_k = b_k$. Try to prove it directly and you will see that it is not obvious how to approach this problem.

The following argument starts with the identity

$$(1 + x)(1 + x^2)(1 + x^4)(1 + x^8)\cdots = 1 + x + x^2 + x^3 + \cdots.$$

In order to convince yourself that this is plausible, you can multiply both sides by $(1-x)$ and observe how the nonzero powers of x successively drop out on both sides. It follows immediately that each nonnegative integer can be represented in one and only one way as the sum of different powers of 2 (the numbers of such representations are the coefficients in the power series obtained after transforming the product on the left).

Euler's method was later called the method of generating functions. A function $a(n)$ of a natural number n is placed in correspondence with a function that is the sum of an infinite series, $A(x) = a(0) + a(1)x + a(2)x^2 + \cdots$. Euler's idea, confirmed by many examples, was that the arithmetic properties of the sequence $a(n)$ appear distinctively in the properties of the function $A(x)$. Characteristically, a purely arithmetical proof of Euler's results on partitions that were proved by him analytically was found only in the second half of the 19th century. Euler's method was later used to prove a series of remarkable results. For example, Carl Jacobi (1804–1851) not only reproved Lagrange's theorem on representing a natural number as the sum of four squares but also found the number of such representations.

The problems about partitions differed from the arithmetic of Diophantus and Fermat not only in their methods but also in their statements. They initiated the additive theory of numbers (as opposed to the multiplicative theory). Goldbach's famous problems, which were posed in a letter to Euler, belong to additive number theory. Among them is the widely known conjecture that every odd (even) number can be represented as the sum of three (two) primes. For sufficiently large odd numbers this was proved

in 1937 by Ivan Vinogradov (1891–1983). Euler, faithful to his principles, studied these problems thoroughly. He verified the conjecture that each odd number n is the sum of a prime and two times a square for $n < 2500$; to this day, it has not been proved. He stated several new conjectures. For example, Euler's conjecture that every prime of the form $8k+3$ is the sum of two times a prime of the form $4l+1$ and an odd square remains unproved. We recall still another of Euler's arithmetic conjectures whose origin is difficult to reconstruct, namely that $3^{\sqrt{2}}$ is transcendental. A generalization of this claim constituted one of Hilbert's problems[22] and was solved in 1934 by Aleksandr Gelfond (1906–1968). This is yet one more example of astonishing foresight!

Analysis

We have already spoken of Euler's work in analysis in connection with series and infinite products. Differential and integral calculus were created during the 17th century, in their final form, in the works of Newton and Leibniz. Euler was the "scientific grandson" of Leibniz, via Johann Bernoulli. By the end of the 17th century, the question had arisen of creating a handbook for infinitesimal calculus. This goal was pursued in *Analyse des infiniment petits pour l'intelligence des lignes courbes* (*Analysis of the Infinitely Small for Understanding Curved Lines*), published in 1696 by the Marquis de L'Hôpital, a student of Johann Bernoulli. In his thinking through of analysis, Euler devoted a considerable part of his life to creating a complete sequence of books on the subject. In 1748 *Introductio in analysin infinitorum* (*Introduction to Infinitesimal Analysis*) appeared in two volumes. The second volume is about analytic geometry. The first volume is a remarkable textbook that students can read with interest even today and contains all of "ordinary" analysis that, in Euler's opinion, needs to precede infinitesimal analysis. There is much elementary material here and many elementary problems. Here is one: "After the Flood the human race propagated itself starting from six people. Suppose that after 200 years the number of people had grown to 1,000,000. It is required to find by what portion the number of people had to increase each year." But there is also a detailed study of elementary functions, including their expansion in series and continuation into the complex domain. Here also are the calculation of $\zeta(2n)$ and the theory of partitions of the natural numbers. *Differential Calculus* appeared in 1755, three volumes of *Integral Calculus* came out in 1768–1770, and an additional volume was published after Euler's death.

[22] In a famous speech in 1900, David Hilbert (1862–1943) posed 23 problems for the coming century.—*Transl.*

We are able to talk only a little about Euler's results in analysis. Above all he made a principal contribution to the evolution of the concept of function. At that time mathematicians clearly understood that functions were the fundamental objects of analysis and knew a large number of concrete functions, but they were only approaching an understanding of the general concept. From the point of view of a mathematician working on applications, a function was always given by some analytic expression. On the other hand, in constructing differential and integral calculus it was often inconvenient to work with explicit expressions. Here a geometric view of functions is more effective. Euler, whose field of vision included both applications and the general theory, developed the two viewpoints of functions in parallel. He was the first who dared to identify general functions with arbitrary (continuous) curves having unique points of intersection with vertical lines. As Riemann wrote, "Euler was the first to introduce these (arbitrary—S.G.) functions into Analysis and, relying on a geometric view, applied infinitesimal calculus to them."

Euler did not only develop analysis for arbitrary functions, but he also pointed out a real situation where arbitrary functions arise in applications. In 1748, investigating a formula for how the form of an oscillating string changes with time, Euler emphasized that at the initial moment of time the form of the string can be arbitrary. At the same time d'Alembert, who had found this formula a year earlier, was having a lot of trouble over the belief that initial form had to be given by an analytic expression (in particular, he came to the conclusion that the problem of the oscillations of a string bent along a parabolic arc was insoluble). In 1761 Lagrange underlined his debt to Euler in using general functions: "...they are necessary for a large number of important questions in dynamics and hydrodynamics.... M. Euler was, I maintain, the first who introduced this new type of function into analysis in his solution of the problem on vibrating strings...." Since Euler's time, the terminology has changed in an essential way: his general ("discontinuous" or "mechanical") functions are continuous from our viewpoint, and after Lagrange functions that are continuous in his sense came to be called analytic. Euler was certain that general functions do not admit an analytic representation. He firmly opposed Daniel Bernoulli, who thought (in connection with the string problem) that general functions are superpositions of harmonics. After 70 years Joseph Fourier (1768–1830) confirmed that Bernoulli was correct if infinite superpositions are considered, i.e., Fourier series.

As remarkable as Euler's results were in the area of formulating the general concept of function, they do not compare in any way to his colossal work in selecting and studying special classes of "good" functions that are

needed for applications. In studying special functions he definitely went beyond the bounds of elementary functions. We have already discussed the zeta function, introduced as early as 1730. Continuing the work of Wallis, Euler sought a function $\Gamma(x)$ that takes the values $n!$ at the integer points and then a function $B(x, y)$ that equals $\frac{(n+m)!}{n!m!}$ (the number of combinations) at the integers. This is how the famous Euler integrals (the gamma and beta functions) came about.

Eighteenth-century mathematicians knew that the elementary functions were not sufficient and recalled Leibniz's dream of investigating the higher transcendental functions, but a sober evaluation showed that there were no standard ways to study this problem. Various examples of functions had been given by different mathematicians, but we now see clearly that this was a problem for the 19th century. At the same time we also see that Euler, following some mysterious instinct, guessed all the special functions that are the subject of higher analysis, almost without exception. We have mentioned the Euler integrals and the zeta function. To these we can add the Bessel functions, some form of the theta functions, and Gauss' hypergeometric function (which received this name much later!), for which various values of the parameters give most of the special functions arising in mathematical physics. Finally, Euler took the most important steps in the theory of elliptic integrals, including an addition theorem. Legendre, Gauss, Abel, and Jacobi started off from these results. We have become accustomed to the the idea that if a new natural class of functions appears, then we need to look for it in Euler's work. In recent years, the dilogarithm $Li_2(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2}$ has mysteriously appeared in the most diverse problems in number theory, algebra, topology, and geometry. It turns out that Euler knew about the remarkable properties of this function, and about addition theorems in particular.

The most important technical method that Euler did not yet know was the continuation of special functions into the complex domain. But he had already taken the first steps in the construction of complex analysis. He had considered the Cauchy–Riemann equations, as did d'Alembert (true, in connection with problems in hydromechanics), which give analytic functions of a complex variable. He made use of complex substitutions in computing real integrals and in the last years of his life computed real integrals using integrals of complex functions, coming close to Cauchy's theory of contour integration in the complex plane. Euler understood the inevitability of the "complex" world.

Euler's most notable result in complex analysis was his discovery of the connection between exponential and trigonometric functions in the complex domain, which cannot be seen if we stay within the limits of the

real numbers. Lagrange called Euler's formula $e^{ix} = \cos x + i \sin x$ "one of the most beautiful discoveries in analysis made in this century." Even today, the formula makes a strong impression. It can be obtained naturally through series or functional equations, and one rarely recalls how it appeared in the mathematics of the 18th century. Surprisingly, the logic of its discovery was straightforward enough. At the beginning of the century Johann Bernoulli, Euler's teacher, while studying the problem of integrating rational functions, turned his attention to the relation $\frac{1}{1+x^2} = \frac{1}{2i}(\frac{1}{x-i} - \frac{1}{x+i})$. If we integrate this formally, then we obtain an arctangent on the left and a logarithm on the right, but of an imaginary argument. After some easy transformations we obtain the formula

$$x = \frac{1}{2i} \ln \frac{1 - i \tan x}{1 + i \tan x}, \tag{3}$$

which is trivially equivalent to Euler's formula. Although Johann Bernoulli did not actually write down (3), he tried unsuccessfully to make sense out of this calculation with imaginary quantities. It was against this background that the famous discussion arose in 1712–1713 between Bernoulli and his teacher Leibniz on logarithms of negative numbers (what is $\ln(-1)$?), and in 1714 "Euler's formula" was briefly mentioned without the necessary substantiation by Roger Cotes (1682–1716), an associate of Newton who died young. Euler, being well informed about his teacher's problems and proceeding from computations, derived (3) in 1728. In 1739 he developed the theory of logarithms in the complex domain, so that all the formulas became correct and all the contradictions disappeared ($\ln(-1) = (2k+1)\pi i$, where k is any integer).

One cannot separate the search for special functions from the delineation of important classes of differential equations. No one had doubted that it was impossible to integrate an arbitrary differential equation explicitly. Euler actively participated in delineating those equations that arise in physics. He considered a series of equations in connection with problems in hydromechanics, the oscillation of strings and membranes, and the propagation of sound: here are Laplace's equation, some forms of the wave equation, and so on. An analytical view of physics was characteristic of Euler. He sought to reduce physics problems to the solution of one differential equation or another. In mechanics, he was the first to move from Newton's geometric language to the language of analysis.

In reviewing Euler's activities in analysis, we stress that he put aside his preference for analytic methods for solving both general mathematics problems and applied ones. But analysis was never Euler's only goal. Recall that he, as opposed to d'Alembert, persistently sought a purely al-

gebraic proof of the fundamental theorem of algebra (the existence of a complex root for every algebraic equation). He did not succeed in finding an algebraic proof and Georg Frobenius (1849–1917) noted with regret that Euler's remarkable algebraic work did not yield what it should have, and much of it was later incorrectly attributed to Gauss.

Geometry

Euler's studies in geometry have a more fragmented character. The second volume of *Introduction to Analysis* is the first textbook of analytic geometry. A lot of analytic geometry comes from Euler. He was the first to consider affine transformations (he introduced that terminology) and to study the group of rotations, and he connected the results he found with the motion of a rigid body. Euler thought about being able to apply analysis to geometry and made the first steps in differential geometry. As one of these he considered geometric problems related to cartography, beginning with the question of in what sense a planar representation on a map is similar to the corresponding picture on a sphere (the Earth's surface). The connection he discovered with complex numbers was unexpected by many.

Even in elementary geometry, Euler discovered facts that no one had noticed earlier. For example, the orthocenter of a triangle (the intersection of the altitudes), the center of the circumscribing circle, and the center of gravity lie on one line, called the Euler line. It seems that the theorem that the three altitudes meet at one point, omitted by Euclid, was not stated clearly by anyone before Euler.

Euler's theorem on polyhedra is probably the best known of his geometric assertions: $V + F = E + 2$, where V is the number of vertices, F is the number of faces, and E is the number of edges. It is interesting that Euler found this relation through examples but could not at first prove it in general, verifying the theorem instead for any pyramid or prism, for certain composite polyhedra, and for regular polyhedra. Euler trusted in mathematical experiments in geometry as well: "Thus, since the truth of this claim has been verified in all these cases there is no doubt at all that it holds for all bodies, so that this proposition is sufficiently established as valid." Only later did he find a general proof.

Euler did not call on his colleagues to compete in solving problems as Fermat had done, but instead readily exchanged solved and unsolved problems with them. This applies to his results on the traditional themes of mathematics competitions such as magic squares, amicable numbers, etc. Up to the present, popular books have preserved some simply stated

problems that Euler thought of or that he was the first to solve. One is the problem of the Knight's Tour on a chessboard, where no square can be visited twice. Another well-known problem is to prove that it is impossible to make the rounds of the seven bridges of Königsberg without going on the same bridge twice. It is clear from the example of this problem that Euler was intrigued by problems that can be solved in a nonstandard way, since this nonstandard way can have far-reaching consequences. In March 1736, Euler wrote to "Sir, the glorious and distinguished Marinoni"[23]: "A problem was posed to me about an island in the city of Königsberg, surrounded by a river spanned by seven bridges, and I was asked whether someone could traverse the separate bridges in a connected walk in such a way that each bridge is crossed only once. I was informed that hitherto no one had demonstrated the possibility of doing this, or shown that it is impossible. This question is so banal, but seemed to me worthy of attention in that geometry, nor algebra, nor even the art of counting was sufficient to solve it. In view of this, it occurred to me to wonder whether it belonged to the geometry of position (*geometriam Situs*), which Leibniz had once so much longed for."

Leibniz had, in fact, left several enigmatic remarks about a mysterious geometry, "which reveals position to us as algebra does quantities" (letter to Huygens, 1679). Euler tried unsuccessfully to clarify the details of this "geometry of position." He worked out a method for solving this problem that is essentially the beginning of topology. He felt that the problem was only the the echo of much deeper problems: "If we could give other, more serious problems here this method could be more useful and it could not be ignored." Within a month, in a letter to Ehler[24] in Danzig, he discussed a generalization of the bridges problem and said, "Thus you see, most noble Sir, how this type of solution bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle. Because of this, I do not know why even questions which bear so little relationship to mathematics are solved more quickly by mathematicians than by others. In the meantime, most noble Sir, you have assigned this question to the geometry of position, but I am ignorant as to what this new discipline involves, and as to which types of problem Leibniz and Wolff expected to

[23]This letter was written to Giovanni Marinoni (1676–1755), an Italian mathematician and astronomer in Vienna. The English translation of this excerpt is taken from Brian Hopkins and Robin Wilson, "The truth about Königsberg," *College Math. J.*, **35** (2004), pp. 198–207.

[24]Carl Leonhard Gottlieb Ehler, the mayor of Danzig. The translation of this excerpt is also taken from Hopkins and Wilson.

see expressed in this way." Thus Euler, following Leibniz, saw ahead to a new area of geometry, the geometry of shapes and without measurement, whose lines would become clearer in a century and a half.

Mechanics

The field of mechanics was within Euler's field of vision from the start. As early as 1736 he published *Mechanica sive motus scientia analytice exposita* (*Mechanics, or the Science of Motion Presented Analytically*). This was his first book and it appeared when he was 29. Euler carefully studied Newton's *Principia*, in which mechanics was presented in geometric language. He discovered that from the viewpoint of applications to concrete problems it was more effective to change to analytical language, using coordinates. In the end, a problem in mechanics is transformed into the purely mathematical problem of solving differential equations. Lagrange continued this direction in mechanics and said in the preface to his *Méchanique Analytique*, "There are no figures in this work, only algebraic operations." Euler clearly was aware that reducing a mechanics problem to a mathematics problem does not solve it: "...Although the principles of mechanics on which all laws of motion are based are evidently sufficiently well known and sufficiently applicable to general phenomena to use them to place changes in motion under analytical formulas, very often analysis is not sufficient to solve the equations.... We really do not see that the principles of mechanics lead us everyday to differential equations whose solutions can be found only when analysis is developed to a point that is still very far away."

Newton's mechanics did not go beyond the limits of the motion of point masses and then Descartes considered the motion of planar plates, but only Euler studied the detailed motion of a solid body of finite size. He did this in a book that saw the light 29 years after the appearance of his *Mechanics*.

Newton's mechanics begins with his three laws as axioms. Euler thought in essence that they needed more motivation and that they should follow from some sort of more primary laws of the universe. An attempt in this direction was a dubious endeavor in 1736. Aleksei Krylov (1863–1945) wrote that Euler only "diluted" Newton's laws and found the roots of Euler's desire in his religious practice. When Euler was in Berlin, a new way to work out a more natural basis for mechanics unexpectedly opened up to him. In 1744, Maupertuis proposed that all the laws of motion and equilibrium in nature could be derived from the fact that all motion takes place so that some quantity or action takes a minimum value. Maupertuis started with optics (Fermat's principle) and moved to mechanics but then interpreted his law as broadly as possible and, confusingly, gave his law of least action a theological meaning by claiming that least action is a

consequence of the "wisest use of the power of the Creator." Maupertuis did not go further into simple applications of mechanics but was inclined towards global problems, which soon involved him in a hot argument that cost him dearly. D'Alembert wrote, "This argument about action, if we are permitted to say, somewhat resembles certain religious arguments in the bitterness in which it is conducted and in the number of people who take part without understanding anything about it."

Euler was on Maupertuis' side from the very beginning. The theological interpretation was not foreign to him: "Indeed, since the edifice of the entire world is perfect and is erected by a wise Creator, nothing happens in the world in which the idea of some maximum or minimum does not appear." But first of all Euler sought a precise statement of the principle that would allow him to change the laws of mechanics. He found such a formulation in the case of central forces but did not give a proof. As Maupertuis himself wrote about Euler, "This great geometer not only established a more fundamental principle than I did but his gaze, more enveloping and more penetrating than mine, led him to the discovery of consequences that I did not draw out."

Maupertuis' assertions were so general that in the discussions (more precisely, the quarrels) people took part who were distant from physics, among them Voltaire, who had old accounts to settle with Maupertuis and who burst forth in the pamphlet *Diatribe du Docteur Akakia et du natif de Saint-Malo* (*The Diatribe of Doctor Akakia and the Native of Saint-Malo*). In the end, Maupertuis' spirit collapsed and Euler, his ardent defender, also caught it from Voltaire. He can be recognized in the scientist who tries to gain glory for himself among the European mathematicians by "what happens on the page of a maximum calculation." Voltaire is talking about a scholar who computes for no less than 60 pages instead of thinking and taking no more than 10 lines, who computes for three days and three nights without spending a quarter of an hour thinking about the right way. This is how Voltaire interpreted the brilliant calculator.

Euler was not infrequently reproached and here they reproached him for overrating Maupertuis' mistaken pronouncements, almost ostentatiously stressing the second-rate nature of his work. They even hinted that the practical Euler had tried to play up to Maupertuis, the all-powerful (before these debates) president of the Berlin Academy of Sciences. But it is thought that this feeling about Maupertuis' work was integral to Euler: he could appreciate pioneering work and understood how to imagine the ideas it contained even in incomplete form. Maupertuis expressed something that Euler could naturally have said. Euler consistently sought a more reliable basis for mechanics than Newton's laws, to which he was not ready to give

primacy. But he was not destined to guess that the necessary principle could be gotten from his beloved calculus of variations.

Astronomy

Euler's activities in astronomy were a continuation of his work in mechanics. His field of interest was celestial mechanics. He was able to realize his astounding computational abilities here. As the French astronomer Dominique Arago (1786–1853) wrote, Euler "calculates the way man breathes." Euler was one of the first to whom calculations were accessible that went beyond observational results. The old celestial mechanics only extrapolated from observations, while the new started off from the law of universal gravitation. The first steps in this direction were taken by Newton himself, who gave a theoretical definition of the acceleration of the moon's motion and explained various anomalies (called "inequalities") in this motion. As always, Euler clearly understood the essential problems of celestial mechanics. First of all it was necessary to try to explain the inequalities in the motions of the large planets Jupiter and Saturn on the basis of their mutual attraction, superimposed on the attraction of the sun. Euler made substantial progress towards the goal towards which he was aiming, namely to explain the so-called "great inequalities" that appeared in the systematic acceleration of Jupiter and deceleration of Saturn. However Euler was unable to reach this result by calculations that agreed well with observations, although he was on the right track. (Laplace later succeeded in doing this.)

The theory of the moon's motion was at the center of Euler's attention. The most burning issue was the problem of explaining the periodic motion of the orbit's perigee, with a period of 9 years. A calculation of the perturbation persistently gave a period of 18 years, until Alexis Clairaut (1713–1765) showed in 1749 that computing the perturbation terms of the next order gave the correct period. Euler recognized that Clairaut, concentrating his energies on solving this problem, had surpassed him: "...in this question M. Clairaut has, if you will, no stronger opponent than I, ... although I was in fact M. Clairaut's predecessor in this question I did not have enough patience to undertake such extensive calculations." Although Euler's theory did not succeed as much as Clairaut's did, its consequences were exceptionally important. On the basis of this theory, Tobias Mayer (1723–1762) constructed tables of the moon's motion with a level of accuracy not seen before. These tables made it possible to measure longitude on board a ship and this method was competitive with the method of using the chronometer that had been invented by John Harrison in 1735. Recognizing Mayer's merit in solving this long-standing practical problem (see the

chapter on Huygens), the British Parliament posthumously awarded him a prize of 3000 pounds in 1765. At the same time Euler was awarded 300 pounds "for theorems with which the recently deceased Professor Mayer of Göttingen constructed his Lunar Tables, allowing for the achievement of great progress in the matter of finding longitude at sea."

Euler worked a lot on calculating the elliptical (unperturbed) orbits of comets. In particular, he worked on Anders Lexell's (1740–1784) famous comet of 1769, which came unusually close to the Earth (on May 10, 1983 the comet came about as close to the Earth for the first time in 200 years).

Although Euler did not succeed in constructing a theory of motion of the planets starting only from Newton's laws and completely agreeing with experiments, he believed in the immutability of the law of universal gravitation. Sometime after his unsuccessful attempt to explain the inequalities in the motion of the moon, Euler, like his contemporaries, thought about "refining" Newton's law. However, further development of the theory of lunar motion showed, in Euler's words, that "the closer it agrees with Newton's law the better it represents the observed phenomena." Euler did not doubt that it was also valid for all of celestial mechanics. Euler's position towards solving the three-body problem is instructive: "I should first of all note that we would have gained nothing by using some convenient work on integrating these equations. On the one hand, I strongly doubt that at some time a method was found for this; on the other hand, even if one had been lucky enough to derive their integrals, then these integrals would have been quite complex and would have been almost no help for use in astronomy. Towards this end they could just as well as been replaced by successive approximations. But if we are talking about approximate expressions, then it would be so easy to obtain them immediately from differential equations."

"Letters to a Princess"

The interrelation between great scientists and monarchs is an interesting subject in the history of science. We have already had a chance to talk about it. It is not only that contacts with the world of power were needed to guarantee the existence of scientists and their work. They frequently entertained the hope that their knowledge would help educate the perfect monarch (recall Leibniz and the Elector of Hannover, the future king of England, and Descartes and Queen Christina of Sweden). Euler hardly had such plans for the princess of Anhalt–Dessau, the eldest daughter of the Margrave of Brandenburg–Schwerin, the niece of Friedrich II. It was probably pleasant for Euler to work with the inquisitive and bright princess, all the more so because her relationship with the scientist was different from

that of most of the king's relatives. The princess gradually had less time for studying and Euler decided to complete the gaps in his lessons with letters: "The hope of having the honour to communicate, in person, to your Highness, my lessons in Geometry, becoming more and more distant, which is a very sensible mortification to me, I feel myself impelled to supply personal instruction by writing, as far as the nature of the subjects will permit."[25] The possibility of systematically laying out his global view of the universe, life, and religion fascinated Euler. Little by little the letters to the princess were oriented towards eventual publication. The three volumes of *Letters to a German Princess* appeared in 1768–1774.

The letters are encyclopedic and create the impression that Euler was trying tell about everything that he was able to think through. We can get an idea of the range of questions discussed by listing how the first volume began: the idea of attraction, the speed of sound and music, light, vision, and the structure of the eye, the law of universal gravitation, the ebb and flow of the tides at sea, Wolff's monadology, "on the relation of the soul to the body," "on natural phenomena," "on the state of the soul after death," "on idealists, egoists, and materialists," "on the perfection of language," "on syllogism," "on morals and physical suffering," "on the destination of man," "the conversion of sinners," "on the wonders of the human voice," etc.

Most scientists are not accustomed to philosophical writings, although many have remarked on the value of popular presentations of scientific knowledge. Condorcet kindly wrote, "...this work presents something rather valuable in the clarity with which the chief and most important things in the fields of astronomy, optics, and the theory of sound are set forth. As for Euler's thoughts regarding philosophy, they are more witty than deep." Euler made use of the pages of *Letters* in his struggle against free-thinking in science, against materialism. He ridiculed "the narrow chemists, anatomists, physicists who leave everything to their experiments. No matter how much is said to them about the properties and essence of the soul, they agree only with what strikes their external senses." All this, together with Euler's musings on religion, provoked harsh responses from Lagrange and d'Alembert. On December 2, 1768 Lagrange wrote to d'Alembert, "...there is one essay that he should not have published for the sake of his honor: this is *Letters to a German Princess*." And on July 15, 1769 he wrote that *Letters* may have amused d'Alembert in its campaign against free-thinkers. In reply d'Alembert compared *Letters* to Newton's commentaries on the Apocalypse and wrote, "Our friend is a great analyst

[25]From the English translation by David Brewster, *Letters of Euler on Different Subjects in Natural Philosophy Addressed to a German Princess*, 3rd ed., W. and C. Tait, Edinburgh, 1823.

but a rather poor philosopher." In a letter of August 7th he wrote, "You had every reason to say that his honor is dear and that he should not have printed this work. It is simply unlikely that such a great genius as he is in geometry and analysis could in metaphysics be lower than the youngest schoolboy so as to say something so trivial and absurd, and this is indeed a suitable opportunity to exclaim: the gods do not give everyone the same gifts."

But the public loved *Letters*! The fact that there were four editions in Russian just in the 18th century testifies to this (the original was published in French). This contrasts with how slowly Euler's scientific works sold (in a letter to Miller, the conference-secretary of the Academy in Berlin, Euler wrote that of the 500 copies of *Differential Calculus* only 100 had been sold and that only 12 subscribers had been found, with difficulty, for *Theory of Motion of a Rigid Body*). And in our time, Vladimir Vernadsky[26] wrote of *Letters to a Princess*, "stopping in delight before the breadth and the considered blending into a whole that bubbles up from this product of his leisure time, no less characteristic of the 18th century as any creation of the art or music of that time."

The art of popularization that appears in the best pages of *Letters to a Princess* was one of the most outstanding manifestations of Euler's pedagogical mastery. Another was how well thought out the ideas he introduced were and how modern his notation was (the notation for trigonometric functions came from Euler, he was the first to consider their values outside the interval $[0, 2\pi]$, etc.). He devoted a lot of energy to educating his students, who always lived in his home. His writings were oriented not only towards communicating his results but towards demonstrating his art: "He preferred to teach his students to the small satisfaction he would have gotten by amazing them. He thought he would not have done enough for science if he had not added to the discoveries by which he had enriched science by candidly putting forth the ideas that led him to these discoveries" (Condorcet). This is the source of his readiness to publish unproven results together with the motivation for their being plausible and even with imprecise but instructive calculations. Here is how he answered a critic who discovered gaps in his work on the refraction of light (dioptrics): "You are mistaken, my dear sir, if you think this work is therefore useless. On the contrary, it is very valuable because it contains computations that are independent of the object itself and of its course and contains an application, and these can serve as a model; in short, these are calculations of a new kind, and that is rather not useless."

[26] (1863–1945), a member of the Academy of Sciences who worked in geochemistry and the philosophy of science.

Concluding Remarks

We have not been able to touch on many of the directions in Euler's activities: optics, cartography, ballistics, the theory of ships, etc. We want to stress again that in Euler's rich legacy mathematics occupies a special place, and in his mathematical works he was first of all an analyst. The great mathematicians of the 19th century learned from Euler's works. "Read Euler—he is the teacher of us all," wrote Laplace. In Gauss' words, "the study of Euler's works remains the best school in various fields of mathematics, and nothing else can substitute for it." No one has ever seriously called into question Euler's reputation as a great mathematician. However in later assessments it was said that Euler did not carry through many difficult problems to their definitive solutions. If we do not judge his activities as a whole but only by the major results that he completed, then he must yield his place to other great scholars. Although he did much in celestial mechanics, he did not leave behind results similar to Clairaut's explanation of the rapid motion of the perigee of the lunar orbit or his calculation of the perturbed orbit of Halley's comet and the prediction of its next return. In arithmetic, Legendre and Gauss found difficult proofs of the existence of primitive roots and the law of quadratic reciprocity, which were stated by Euler.

In 1842, Jacobi noted an important aspect of Euler's mathematical legacy in a letter to Pavel Nikolaevich Fuss[27] (1797–1855): "Recently, I thoroughly

[27] Also known as Paul Heinrich Fuss, Euler's great grandson, the son of Nicolaus, and permanent secretary of the Russian Academy of Sciences.—*Transl.*

studied Euler's integral calculus once more and was again surprised at how fresh this seventy-year old book remains while it is completely impossible to read d'Alembert's book, which was written at the same time. It seems to me that the reason lies in its examples, because these examples have not only the auxiliary value of illustrations but they constitute the whole content of the general propositions." Euler was persistently compared to d'Alembert throughout his life and Jacobi continued this after his death. In May, 1841 he wrote to Fuss: "It is surprising that now it is impossible to read even a line left by d'Alembert while at the same time one still reads Euler's best works with delight, and they died in one and the same year. It seems that d'Alembert exhausted his elegance on *belles-lettres*." Jacobi's and Friedrich II's tastes did not agree, but Jacobi was definitely wrong about d'Alembert.

Euler was appreciated most of all by those who studied his works and did not judge his legacy by the peaks, by those who studied with him and made use of his prophetic ideas.

In conclusion, we present a curiosity which, by the way, is more characteristic of the peculiarities of academic "democracy" in Russia than of Euler's merit. In the last year of the 19th century the Petersburg academicians were thinking ahead about celebrating the forthcoming 200th anniversary of the birth of the great scientist, to occur in 1907. At a general session on February 6, 1899 the Academy discussed a proposal by the Physical and Mathematical Sciences Section to erect a monument to Euler in St. Petersburg funded by international subscriptions. A mathematician, Academician Nikolai Sonin (1849–1915) came out decisively in opposition to the proposal. He said that Euler's works had aged, that he had been greatly surpassed by Lagrange and Gauss, that "the traces of Euler's work had practically been covered up." In general, monuments were to be erected for great scholars and Euler was perhaps outstanding, so that his bust, which had been placed in the conference hall shortly after his death, was quite sufficient for him. Also, he did not understand why the monument needed to be placed in Petersburg and not in Basel, where Euler was born, or in Berlin, where he worked almost as long as in Petersburg. The question was put to a vote. The vote was a tie which meant, according to the regulations of the Academy, that a monument to Euler was rejected. Democracy won!

Today there is the Euler International Mathematical Institute in St. Petersburg, but there is still no monument.

Joseph Louis Lagrange

> . . .my pursuits are reduced to doing geometry peacefully and quietly. And because I am not hurried and I work more for my pleasure than for work, I do as the grandees who build something: I do, I undo, and I redo several times until I am reasonably happy with my work, which nevertheless comes very rarely. Lagrange[1]

A Letter from Turin

In August 1755 the great Euler (1707–1783) received a letter from Turin, from the 19 year-old Lagrange who had written to him before. Euler no doubt had already formed the opinion that his correspondent was a talented and mature mathematician despite his youth. But all the same, the contents of the letter astonished the scientist.

Since the end of the 17th century mathematicians had increasingly turned their attention to problems that we now call variational but were then usually called isoperimetric. It all began with a problem posed by Johann Bernoulli (1664–1748) on the brachistochrone, the curve of most rapid descent between two points. By the way, problems about curves having one maximum-minimum property or another had arisen earlier: a circle is the curve of given length that encloses the greatest area (this is called the isoperimetric property and gives its name to this class of problems), a straight line is the shortest distance between two points, etc. The number of such problems grew and mathematicians solved them with satisfaction, matching their "secret key" to each problem.

[1] From a letter of Lagrange to d'Alembert, January 1, 1781; in *Oeuvres de Lagrange*, Vol. 13, J.-A. Serret and G. Darboux, eds., Paris, Gauthier-Villars, 1867–1892, p. 360.

Joseph Louis Lagrange.

However the fashion at the time when differential and integral calculus were blossoming required a search for a *general* method that would develop a calculus for the solution of isoperimetric problems. The remarkable mathematicians who studied these problems sensed their common features intuitively. Jacob Bernoulli (1654–1705) did a lot in this area. All the same, the picture remained rather mixed and there was much to work out in constructing a general method.

Euler was exactly 19 years old when his teacher Johann Bernoulli gave him the problem of a brachistochrone in a resistant medium. Then he added the problem of the shortest ("geodesic") curve on a surface. Variational problems were always in Euler's field of vision, and by 1732 he had crystallized a general method for solving them. It took another 12 years to perfect the method, and in 1744 his summary memoir appeared on the solution of "isoperimetric problems in the broadest sense." The method was illustrated by the solution of more than 60 of the most diverse problems.

Today we understand clearly the difficulty in solving variational problems: in some sense they were ahead of their time for 18th century analysis. At that time analysts primarily studied functions of one variable, and to a lesser extent functions of several variables. However, the curves that figure into variational problems are not characterized by a finite set of parameters. To all intents and purposes these problems deal with functions of an infinite number of variables, and this was the patrimony of 20th century

analysis (functional analysis).

Euler's principal observation was that curves that are solutions to isoperimetric problems are solutions to certain differential equations. In drawing conclusions about these equations Euler in fact saw the basic issue. He moved very carefully in order to stay within the bounds of ordinary analysis: he replaced curves by polygonal lines (after all, they depend on a finite number of parameters that characterize the vertices) and watched how the quantities in the problem changed when only one vertex changed. The desired differential equation is obtained but the path to it is rather thorny. Jean-Baptiste-Joseph Delambre (1749–1822) (do not confuse him with d'Alembert!), Lagrange's true friend and biographer, wrote that this method "does not have the simplicity that is desired in a question of pure analysis."

These words probably reflect Lagrange's opinion. With the decisiveness of youth he dared to pursue to its completion a scheme that he worked out for functions, where he considered the dominant linear part df of the increment of a function $f(x)$ corresponding to an increment dx of the argument x, and looked for some x for which $df(x) = 0$. He considered functions of curves, i.e., functionals (of a special form) $I(l)$, not being afraid that this is in fact a function of an infinite number of variables. For a fixed curve l he considered an arbitrarily small "perturbation" δl, defined the dominant part of the corresponding increment of the functional, δI, and for determining the curves for which $\delta I = 0$ obtained a differential equation that Euler had reached in a roundabout way and that is now called the Euler–Lagrange equation. We note that Lagrange providently introduced the new notation δ, which is similar to but different from the notation for the differential d. This new notation was introduced successfully and greatly helped advance the work.

A little information was enough for Euler to appreciate the advantage of Lagrange's improvements. A lively correspondence started, and the high evaluation of Lagrange's work by the great scientist inspired the beginning mathematician. The letters discussed increasingly complex formulations of problems: after all, the power of the new method should be demonstrated by the solutions of new problems that were inaccessible by the old techniques. One of Lagrange's letters generated Euler's own interest in extremal problems. As early as 1756 he submitted two reports to the Berlin Academy related to Lagrange's method. In the same year Lagrange was elected as a foreign member of the academy at Euler's instigation, a rare honor for a young scholar who had still not published his work (at the time less meaning was attached to this election than would be today).

Euler did not rush to publish his new results, letting his young colleague

take his time to prepare his work for publication. He explained his position in a letter on October 10, 1759: "Your analytical solution of the isoperimetric problem contains, as far as I see, everything one could want in this area and I am extremely happy that this theory, which after my first attempts I was hardly alone in studying, has been brought by you to the height of perfection. The importance of the question motivated me to the point where with the help of your illumination I deduced the analytic solution myself. However, I decided to conceal this until you publish your results, since in no way do I want to take away from you any of the glory that you deserve." What a remarkable example of scientific ethics!

Euler's letter added to Lagrange's determination to publish what he had done, and his memoir *Essai d'une nouvelle methode pour determiner les maxima et les minima des formules integrales indefinies* (*An Attempt at a New Method of Determining the Maxima and Minima of Indefinite Integral Formulae*) appeared in 1761–1762, in Volume II of the new journal of the Turin Mathematical Society, *Mélanges de Turin*. In 1764 Euler also published his results, prefacing his work with these words: "Since I worked long and fruitlessly on the solution to this question, I was surprised to see that this problem is solved in *Mélanges de Turin* so easily and happily. This excellent discovery delighted me all the more since it differs significantly from the methods I gave and it significantly surpasses them in simplicity." It is somewhat surprising that Euler did not mention the earlier correspondence. Euler proposed to call the new method the "calculus of variations" by analogy with differential calculus (δI is called a variation).

This was how Lagrange's scientific debut took place. In one sense it was unique. There are other examples where great mathematicians obtained their first powerful results at the same age as Lagrange. But this usually involved solutions to concrete problems. Interest in perfecting a method such as this comes with the years. We see that already in his first work Lagrange displayed what would always distinguish him in the future: clarifying the situation completely, perfecting the method, and searching for first principles is more valued than solving concrete problems.

Giuseppe Luigi

We have discussed Lagrange's first great work but it is still worth saying a few words about the earlier events in his life. Joseph Louis Lagrange was born on January 25, 1736 in Turin, Italy. At birth he was named Giuseppe Luigi. His great-grandfather had come from France and entered the service of the Duke of Savoy, and his grandfather and father continued to serve as treasurers for factories and construction. When the future mathematician

was born the family was ruined. He later said, "If I had been rich, I probably would have not achieved my position in mathematics. And in what other activity would I have achieved such success?" Incidentally, in the family's original plan Joseph Louis was destined for a career as a lawyer, and at the age of 14 he entered the university in Turin. However he soon transferred to the Royal Artillery School, in connection with his increasing interest in mathematics. At 19 he was a professor at this school (even earlier, according to some information).

His first attempts to discover something new in mathematics led Lagrange to find something that was already known. Contact with the exceptionally original Italian mathematician Count Giulio Fagnano (1682–1766) helped the young man understand that a serious study of modern mathematics must precede any independent work. And we have seen that Lagrange's first results were not the fortunate discovery of a young dilettante but the end results of strenuous work by someone who was becoming a professional. The ability to understand thoroughly and critically and to refine a previous attempt marked Lagrange's scientific work from the very start.

A circle of young mathematicians and physicists built up around Lagrange, and this was later transformed into the Turin Academy of Sciences. The journal *Miscellanea Philosophico-Mathematica Societatis Privatae Taurinensis*, or *Mélanges de Turin* for short and in French, appeared beginning in 1759. We have already said that Volume II contained Lagrange's memoir on the calculus of variations, while Volume I contained two of his works, including *A Study on the Nature and Propagation of Sound*. Some comments here on the problem of the vibrating string will be very instructive. In 1747–1748 this problem was studied by the three strongest mathematicians of the time, Jean Le Rond d'Alembert (1717–1783), Euler, and Daniel Bernoulli (1700–1782). There were essential differences in their interpretations. D'Alembert, who was the first to solve the string equation, thought that the initial position should be described by a function having a single analytic expression (it was still not clear what this meant). Euler insisted that the function could be completely arbitrary (even discontinuous, as we have said) and this was the first time in analysis that functions in general appeared, given by graphs rather than analytic expressions. Finally, Bernoulli considered harmonic oscillations with different frequencies and claimed that an arbitrary oscillation could be expanded as an infinite superposition of harmonic oscillations, which neither d'Alembert nor Euler believed.

Lagrange thought of a clever method that considered a string of constant density to be the limit of weightless strings with a finite number of

uniformly distributed identical weights. The question of oscillations of strings with weights was thought to be elementary. In passing to the limit, Lagrange confirmed Euler's opinion. Later, repeating this argument in *Mécanique Analytique* (*Analytical Mechanics*), he recalled, "By this very method in the first volume of *Mélanges de Turin* I proved the validity of Euler's construction, which had not been sufficiently grounded." Soon Lagrange had still another chance to convince himself how right Euler was by insisting on the need to use general (nonanalytic) functions in analysis. In the study of how air moves in tubes of constant cross-section, curves arise that become straight lines at some point (Euler called them "mixed" functions). Passing to the limit in the same way convinced Lagrange that Bernoulli was right. He was close to a proof that it is possible to expand an arbitrary function into harmonics (as a Fourier series), but a precise proof needed to wait for another forty years.

We have already seen how kindly Euler received Lagrange's first work. Lagrange's work on the string also attracted the attention of another of his great contemporaries, d'Alembert: "Until we meet again, sir, You deserve, if I am not mistaken, to play a great role in science and I applaud the start of Your success." As Delambre said, "among these famous geometers a twenty-three year-old young man suddenly appeared not only as their equal but as an arbiter among them, in order to cut short a difficult struggle, to point out to each where he was right and where he was wrong, to correct these errors, and to give the true solution, which had not been achieved although it had been foreseen." This observation exactly communicates the style of Lagrange's article, and the letters to him from Euler and d'Alembert indeed reflect their readiness to accept Lagrange as an arbiter.

The Foundations of Statics

Lagrange was the soul of the Turin circle. The articles published in *Mélanges de Turin* by his friends carry the distinct mark of Lagrange's strong influence. This applies especially to an article by François Daviet de Foncenex (1734–1799) who was, evidently, only an accessory to Lagrange's systematic consideration of the foundation of mechanics. Later, his famous *Analytical Mechanics* began with the subject of this article and strikingly demonstrates how thorough Lagrange was as he undertook this work.

This concerns a comparison of the two most important principles of statics: that of the lever and that of the composition of forces applied to the same point. At the foundation of the theory of the lever, Archimedes placed an axiom saying that a lever with equal arms and weights is in equilibrium and that the load at the fulcrum is doubled in this case. Many authors

tried to make Archimedes' argument more precise and to supplement it but they, in Lagrange's words, "while impairing simplicity... have gained almost nothing from the point of view of precision." Lagrange noted that it is natural to consider the first part of the axiom as being obvious from symmetry: "we cannot perceive a basis by which one weight outbalances the other." However, he did not see any logical basis for the fact that the loading on the fulcrum must necessarily be equal to the sum of the weights: "evidently all mechanists considered this assumption as the result of everyday observation, which teaches us that the weight of a body depends only on its mass and in no way on its form." Lagrange proposed to derive the second half of Archimedes' axiom from the first. He considered a homogeneous triangular plate ABC, where the base AB of the isosceles triangle is horizontal. The vertices A, B are loaded with equal weights P and the vertex C with weight $2P$. The plate rests on the median line MN parallel to AB (Figure 1). It will be in equilibrium, which follows from the first part of Archimedes' axiom applied to the two levers AC, CB with fulcra M, N. But then the lever CF with fulcrum E will also be in equilibrium (here F is the midpoint of AB and E is the intersection of MN and CF). This means that the loading on F must equal the weight $2P$ at C (strictly speaking, we are applying here the converse of the first part of Archimedes' axiom, which is easy to deduce), and this is exactly the loading on the fulcrum of the lever AB. Lagrange correctly noted that considering the equilibrium of a planar plate with respect to a pivot was a method he got from Huygens.

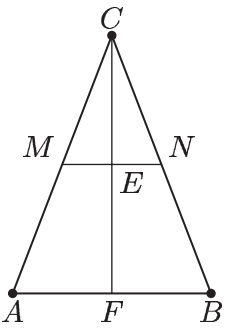


Figure 1.

Furthermore, Lagrange considered the principle of composition of forces applied to the same point, which is easily validated using the composition of motions. The essential difference in these principles is that in one case the forces are applied to different points and in the other to the same one. Nevertheless many statements in statics can be derived from either princi-

ple. This leads to wanting to refrain from taking the principle of the lever as an axiom at all, but Lagrange was alert to the fact that all known deductions of Archimedes' axiom from the law of composition of forces were rather artificial: "...although, strictly speaking, both the principles of the lever and of the composition of forces always lead to the same results, it is interesting to note that the simplest case for one of these principles becomes the most complex for the other."

Lagrange's intuition allowed him to discover a delicate point correctly, although he could not explain it completely. It is connected to the interaction between mechanics and geometry. The point is that the law of composition of forces applied at the same point does not depend on the parallel postulate, while in Lobachevsky space the loading on the fulcrum of a lever is greater than the sum of the weights of the points. In deriving the second half of Archimedes' axiom the claim is used that the altitude of the isosceles triangle in Figure 1 meets the median line at its midpoint, which rests on the parallel postulate and is not valid in Lobachevsky's geometry. Evidently Lagrange did not yet know this, although it is known that he thought about the problem of the fifth postulate.

The Principle of Least Action

In Volume II of *Mélanges de Turin* the memoir on the calculus of variation was followed by Lagrange's *An Application of the Method Proposed in the Preceding Memoir for Solving Various Problems in Dynamics*. Here Lagrange followed in Euler's footsteps. In 1744 Pierre de Maupertuis (1698–1759) had formulated a very general and hazy principle, according to which everything in nature, including mechanical motions, occurs in such a way that some quantity—action—achieves its minimum value. For the case of the motion of a point in a central field, Euler turned this vague statement into a completely precise one, defining action in this case as the integral of the velocity along a path, $\int v ds$. Lagrange generalized Euler's principle to the case of an arbitrary system of related points that interact with one another in an arbitrary way. Defining action in this generalized situation, Lagrange used the calculus of variations that he worked out to solve a diverse range of problems in dynamics, including hydrodynamics. He had no doubt that using this principle he could construct the whole edifice of mechanics. In *Analytical Mechanics*, he wrote, "This is the principle to which I improperly gave the name of Least Action and which I view not as a metaphysical principle but as a simple and general result of the laws of mechanics. In volume two of the *Mémoires de Turin*, the use which I made of this principle to solve several difficult problems of dynamics can be

found. This principle, combined with the one of Forces Vives and further developed following the rules of the calculus of variations, gives directly all the necessary equations for the solution of each problem."[2]

As Joseph Fourier (1768–1830) wrote, "He reduced all the laws of equilibrium and motion to a single principle and, what is no less surprising, he subjected them to a single method of calculation that he himself invented."

First Work in Astronomy

We have seen that Lagrange's activities first developed in the areas of those questions and problems that were traditional for mathematicians in the 18th century, as found in the sphere of interests of his older contemporaries Euler and d'Alembert. The logic of the time inevitably had to lead to his needing to try his powers in celestial mechanics. There was no more burning problem than that of the agreement of observations of the movement of heavenly bodies with the law of universal gravitation. It was necessary to clarify, on the one hand, whether the undoubted deviations from Kepler's laws (then called "inequalities") were explainable in the context of this law and, on the other hand, how various additional regularities in celestial mechanics came about. For example, why do we see only one side of the moon? The Paris Academy of Sciences took the explanation of this phenomenon as the theme of its prize competition for 1764.

We must say that themes for academic prizes in Paris were chosen with the greatest taste, and winning such a prize was very prestigious for a mathematician, especially a young one. Lagrange won first prize and gained an enthusiastic response from d'Alembert: "I read with great satisfaction the fruit of Your excellent works on librations. They deserve the prize you were given."

Really, the laws of the moon's motion were derived very precisely from the observations of Giovanni Cassini (1625–1712): the moon's axis of rotation is fixed relative to the surface, the period of rotation and the period of revolution around the earth coincide, the axis of rotation is at a constant angle to the plane of the ecliptic (the earth's orbit) and, finally, the moon's axis of rotation, the ecliptic, and the lunar orbit are located in the same plane. Lagrange showed that because of this the surface of the moon is not spherical and the attraction of the earth gradually equalizes the period of the moon's own rotation and that of its revolution around the earth. Lagrange came close to explaining Cassini's last law, which d'Alembert had

[2]This quotation is taken from the English translation of the 1811 edition of *Mécanique Analytique* by Auguste Boissonnade and Victor N. Vagliente, Kluwer, Boston, 1997, p. 185. The principle of Forces Vives is now called the principle of kinetic energy.—*Transl.*

earlier failed to do, but made an estimation error. Only in 1780 did he finally succeed in establishing Cassini's theory.

The explanation of the inequalities in the motion of Jupiter's moons was chosen as the theme by the Paris Academy of Sciences in 1766. The solution of analogous questions for the moon brought fame in their time to Alexis Clairaut (1713–1765) and d'Alembert. Additional complications arise with Jupiter's moons, in particular because there are several moons and also because of the nearness of Saturn. Euler was surprised that Lagrange could handle this problem in the work that won the prize: "The irrational formula that expresses the distance from Jupiter to Saturn cannot be represented sufficiently by a convergent series, and this is a fundamental obstacle. I strongly doubt that it can be overcome. . . . Now it is all the more interesting for me to know how M. Lagrange overcame these difficulties in his work that obtained the prize, and since I do not have grounds to doubt the success of his solution I can flatter myself with the hope that theoretical astronomy has been brought in our time to the highest degree of perfection." When 24 years later Pierre-Simon Laplace (1749–1827) returned to the problem of Jupiter's moons in order to complete what Lagrange had begun, he spoke with delight of the results of his predecessor that were obtained with the help of "sublime analysis."

Visit to Paris

In 1766 Lagrange turned 30. This was an important dividing line in his life. Provincial Turin had become confining for Lagrange's scientific activities. In his personal life he was unassuming, was noted for his ill health, and his modesty in relations with people not infrequently took the form of shyness and even unsociability. But he knew how to appreciate and use his relationships with colleagues. At the beginning his contacts with friends in the Turin circle were satisfying to him. He invested much energy and spirit in the work of the circle but he had long since outgrown these colleagues. He no longer had regular contacts with Fagnano, who was elderly and died in 1766. Lagrange carried out an extensive correspondence, but when he traveled to Paris in 1763 he could see for himself the value of face-to-face contacts with scientists. He accompanied his friend the Marquis Caraccioli, who had been named ambassador to London. But Lagrange did not go to London. As Delambre recalled, "Dangerously falling ill after a meal where the Abbé Nollet entertained him with dishes prepared in the Italian style, Lagrange could not travel to London but remained to be treated in Paris, and hurried to return to Turin when he recovered."[3]

[3]Jean-Antoine Nollet was a clergyman and physicist.—*Transl.*

In Northern Italy castor oil, well fried in advance, was used to prepare meat. In Nollet's kitchen, where they decided to prepare the meal "à l'italienne," they used castor oil without the required preparation and it completely displayed its well-known laxative properties. However on the scientific front the illness bore fruit. Lagrange was able to form relationships with the best French mathematicians, d'Alembert, Clairaut, and the Marquis de Condorcet (1743–1794), but even among the less well-known scientists there were some who remained his friends for the rest of his life. Lagrange often said that these six months in Paris were the happiest time of his life.

In 1766 Euler moved from Berlin to St. Petersburg, freeing up the position of Director of the Physics-Mathematics section in the Berlin Academy of Sciences. He suggested to Friedrich II that Lagrange be his successor. D'Alembert, on whose opinion the king relied to the highest degree, energetically supported this candidacy. Lagrange was sent an invitation with this expressive justification: "the greatest geometer in Europe must live near its greatest king." Perhaps Friedrich was right about himself, but with Euler and d'Alembert alive and working Lagrange was hardly accepted as the greatest geometer in Europe. The king was probably worried a bit about his own wounded pride, since he was not able to get d'Alembert for his own academy and had to part from Euler.

All the same, there is no doubt that at his thirtieth birthday Lagrange was admitted to the mathematical Mt. Olympus. He had already taken shape as a mathematician: the foundation of everything he would do had been laid and the style of his work had become clear, as had his strong and weak points. Lagrange began his mathematical life as a student of Euler and d'Alembert in the highest sense of the word. He continued to work on the problems that they had begun and found new aspects that were unknown to his teachers. Their enthusiasm was testimony to this. Lagrange interpreted the creations of his teachers in a distinctive way: he mastered problems that Euler's intuitive genius had almost guessed, clarified them completely, and sharpened the necessary concepts and technical methods, which was more characteristic of d'Alembert. And in the future, Lagrange's strength would not be primarily in the discovery of new paths but rather in the astonishing ability to deepen, clarify, and supplement the necessary strokes of the painting which others had tried to paint before him. And no difficulties on this path were terrifying for Lagrange.

Lagrange in Berlin

The volume of *Mélanges de Turin* for 1766–1769 still contained work of Lagrange that had delighted Euler: he completely clarified the nature of the

addition formula for elliptic integrals that Euler had once conjectured. And as had happened before, Euler enthusiastically returned to a subject he had already left behind. By November, 1766 Lagrange was in Berlin, although the king of Sardinia[4] parted with his scholars reluctantly. This was not the best time for Lagrange to turn up at the academy. Euler, d'Alembert, and de Maupertuis were no longer there. But a very original mathematician, Johann Lambert (1728–1777), was working there. Lambert proved, in particular, that the number π is irrational. Lagrange and Lambert had much in common mathematically, as well as personally, and resembled each other. Their friendship was very important for each of them and lasted ten years until Lambert died. It was not easy for the reserved Lagrange to adapt to life at the Prussian court. But he, as opposed to Euler, could do this and avoid conflict. Lagrange led a measured life: outside obligations, meetings, and correspondence occupied most of the day but after the obligatory walk, every evening was devoted to working on science in quiet, behind closed doors. Lagrange got married and exchanged letters with d'Alembert about this. D'Alembert wrote, "I know that You have taken a dangerous leap. A great geometer must first of all calculate his own happiness and that having made the calculation you found the solution was marriage." Lagrange: "I do not know if I calculated well or poorly, or better—I did not calculate at all, because I would have acted like Leibniz, who could not decide whether to get married. I recognize that I never had an inclination toward marriage... one of my relatives had to do me a favor; someone had to take charge of me and my work." But it turned out that Lagrange soon had to look after his wife, who was dying of tuberculosis, and he did his duty without fault.

Analytical Mechanics

Lagrange stayed in Berlin for a little more than twenty years. This was the time of his maturation, the most productive period in his life. There are several great scientists who left behind one main book (e.g., Newton's *Principia*, Huygens' *Pendulum Clocks*). For Lagrange, *Analytical Mechanics* was such a book. It came out in 1788 when Lagrange was already in Paris. But the book kept to itself the important point that it was created in Berlin and conceived while the author was still in Turin.

The plan of the book can best be learned from the words of the author: "There already exist several treatises on mechanics, but the purpose of this one is entirely new. I propose to condense the theory of this science and the

[4] At the time, Turin was the capital of the Kingdom of Sardinia.—*Transl.*

method of solving the related problems to general formulas whose simple application produces all the necessary equations for the solution of each problem. I hope that my presentation achieves this purpose and leaves nothing lacking. In addition, this work will have another use. The various principles presently available will be assembled and presented from a single point of view in order to facilitate the solution of the problems of mechanics. Moreover, it will also show their interdependence and mutual dependence and will permit the evaluation of their validity and scope. . . . No figures will be found in this work. The methods I present require neither constructions nor geometrical or mechanical arguments, but solely algebraic operations subject to a regular and uniform procedure."[5]

In short, Lagrange intended to show that a purely analytical procedure was sufficient for solving mechanics problems (in order to emphasize this, he pointedly did not include pictures), that "uniform" (today we would say algorithmic) rules could be presented for considering such problems, and that there are simple general principles on which all of mechanics can be constructed. How original was this point of view? We can recall that Euler was the first, in his 1736 *Mechanics*, to turn away from Newton's purely geometric approach and use an analytic method based on changes in coordinates and on systems of differential equations (Lagrange calls this book "the first great work in which analysis was applied to the study of motion"). On the other hand, d'Alembert's 1743 *Dynamics* is prefaced by these words: "In the present essay I set a dual goal for myself: to broaden the field of mechanics and to make the approach to this science smooth and even. . . In a word, I tried to broaden the domain of application of the principles and at the same time to decrease their number." And Lagrange valued d'Alembert's treatise very highly: ". . .in it is presented a direct and general method, with the help of which we can solve, or in any case express in the form of equations, all the problems of mechanics that we have only been able to formulate."

What was Lagrange was thinking of that was so new? That he in turn had completed what his predecessors had dreamt of and turned their remarkable studies into a universal and working apparatus. He evaluated his own program rather modestly and did not in any way compare himself to Newton, "to whose lot fell the joy of explaining the world system." Lagrange thoroughly studied and presented this preceding work in his *Analytical Mechanics*. The historical pages are a decoration for his book. By the way, Lagrange was reproached for the fact that this survey contained definitions of basic concepts in mechanics that turned out to be insufficiently

[5]Ibid., p. 7.

worked through.

Thus, at the beginning of his mechanics Lagrange "collects" what others already did. Mechanics is divided into statics and dynamics. We have already talked about two starting points for statics: the principles of the lever and of the composition of motions. To this we add the principle of virtual (possible) velocities (now often called the principle of virtual displacement or virtual work) which goes back to Galileo and was worked out by Stevin, the Bernoulli brothers, and d'Alembert. The principle is that in equilibrium the work done by all forces is zero for all infinitesimal displacements that are compatible with the constraints placed on the elements of a mechanical system. Lagrange "only" wrote down this condition in the form of an analytic equation and tried to prove not only the ability of this principle to work, which had already been done by others, but most of all its universality and its sufficiency as a basis of all statics. "In obtaining this general formula Lagrange, with a talent that was almost characteristic of him alone and perhaps up to now had not been excellent, developed from this formula the general properties of the uniformity of forces and gave the solution to the foremost problems of statics. . ." (Aleksei Krylov). Another instructive aspect of the book is its substantiation of the principle using systems of pulleys.

Moving to dynamics, Lagrange exploited d'Alembert's idea on the reduction of dynamics to statics. It had been worked out in a somewhat different version in concrete problems by Jakob Hermann (1678–1733) and Euler. We are talking about the possibility of separating out those forces that do not produce any motion and are counterbalanced by the reactions of the constraints (d'Alembert spoke of the motive (force) that is lost to motion); these forces alone would cause the body to be in equilibrium. Starting from this, Lagrange obtained a fundamental equation for dynamics from the fundamental equation for statics. This is the emotional high point of the book. The goal of the rest was to demonstrate that all of mechanics can be derived from this fundamental equation (from a single formula!).

Realizing this program begins with deriving all the "starting points" of mechanics from the fundamental equation: the law of conservation of energy, the law of motion of the center of gravity, and the area principle. The culmination of this part is the derivation of the principle of least action from the fundamental equation. Lagrange understood that, in turn, his equation could be derived from the principle of least action and, possibly, that his earlier plans amounted to constructing analytical mechanics on the basis of this principle. Today exactly this method of construction is the most common, while Lagrange preferred to begin with the fundamental equation. Perhaps tactical considerations played a role here: his

contemporaries were still not ready to accept a variational presentation of mechanics.

Lagrange's next problem was to learn how to work with the fundamental equation. The biggest thing was to take into account the constraints placed on the points of the system. For this reason it is convenient to pass from the Cartesian coordinates of the points on which the relations have been placed to some sort of generalized coordinates that can change independently. These could be the angle of inclination of a pendulum or the width and length of a point moving on a sphere. Lagrange showed that for arbitrary independent coordinates the equation of motion is written in terms of the kinetic energy T and the potential energy U of the system, and that it is enough to use their difference $L = T - U$, the Lagrange function. These equations are now called the second-order Lagrange equations.

First-order equations correspond to the case where the constraints cannot be completely solved or where it is undesirable to solve them, i.e., some equations on the coordinates remain. Lagrange showed how to write the equation of motion using the equations on the constraints, where in these equations there are quantities than can be interpreted as reactive forces of various constraints. This is the first appearance of Lagrange multipliers, probably Lagrange's most well-known mathematical legacy (we will have more to say about them below).

The main part of the book is devoted to realizing the system he worked out in a series of important concrete situations: small oscillations, the motion of bodies under the action of mutual attraction (basically, celestial mechanics), constrained motion (in particular, pendulums), and the motion of a rigid body.

Lagrange realistically assessed the possibility of working out his program. He had no illusions that reducing mechanics problems to differential equations means that the problems are solved, since "they (the equations—S.G.) still have to be integrated, which often exceeds the abilities of the analysis that we know." In connection with this he worked out approximation methods and with great attention related them to special cases where the integration could be carried out explicitly. (This is very reminiscent of the point of view of modern mathematical physics.) From this standpoint he, following Euler, considered the problem of the rotation of a rigid body—of a "top."

Lagrange was completely focused on a proof that it was possible to turn mechanics into a chapter of analysis, to derive all of mechanics from a simple general principle. The idea of a deductive construction of mechanics in the form of Euclidean geometry was not new. It was not for nothing that Newton called his book *Principia* and his laws axioms. But no one before

had consistently carried out this program. Such a formal project necessarily has serious limitations, and seems curious after so much time has passed and the propositions proved already seem beyond doubt. Indeed, why did Lagrange completely avoid drawings or always "proceed genealogically" from the fundamental equation? But that is the logic with which science develops.

Those who continued his work could appreciate Lagrange better. The two sides of modern mechanics are associated with the names of Lagrange and William Rowan Hamilton (1805–1865). Here is what Hamilton wrote: "...Lagrange has perhaps done more than any other analyst, to give extent and harmony to such deductive researches, by showing that the most varied consequences respecting the motions of systems of bodies may be derived from one radical formula; the beauty of the method so suiting the dignity of the results, as to make of his great work a kind of scientific poem."[6]

A remarkable feature of Lagrange's constructions is that they found application far from mechanics. The Lagrangean equations appeared in the theory of electromagnetism. As Henri Poincaré (1854–1912) wrote, "To demonstrate the possibility of a mechanical explanation of electricity we need not trouble to find the explanation itself; we need only know the expression of the two functions T and U, which are the two parts of energy, and to form with these two functions Lagrange's equations, and then to compare these equations with the experimental laws."[7]

Lagrange's work was a model for James Clerk Maxwell (1831–1879) in his creation of the analytical theory of electricity: "The aim of Lagrange was to bring dynamics under the power of the calculus. He began by expressing the elementary dynamical relations in terms of the corresponding relations of pure algebraical quantities, and from the equations thus obtained he deduced his final equations by a purely algebraical process. Certain quantities (expressing the reactions between the parts of the system called into play by its physical connexions) appear in the equations of motion of the component parts of the system, and Lagrange's investigation, as seen from a mathematical point of view, is a method of eliminating these quantities from the final equations. In following the steps of this elimination the mind is exercised in calculation, and should therefore be kept free from the intrusion of dynamical ideas."[8]

[6]William R. Hamilton, "On a general method in dynamics," *Philos. Trans. Roy. Soc.*, Part II, 1834, pp. 247–308.

[7]Henri Poincaré, *Science and Hypothesis*, translated by W. J. Greenstreet, Walter Scott Publishing, London, 1905.

[8]James C. Maxwell, *Treatise on Electricity and Magnetism*, Vol. II, 3rd ed., Clarendon Press, Oxford, 1892, p. 199.

An especially effective means of extending Lagrange's ideas beyond the limits of mechanics was the principle of least action: "All reversible processes, whether mechanical, electrodynamic, or thermal in nature, are subject to one and the same principle giving a unique answer to all questions concerning the course of the process. This law is not the principle of conservation of energy which, although we apply it to all phenomena does not define their course uniquely; it is a more general principle, the principle of least action" (Max Planck (1858–1947)).

Lagrange saw his destiny to be the creation of a universal language of mechanics. Thanks to this he abstracted as much as possible from the specifics of the concrete problems that attracted his great predecessors. Later, Siméon-Denis Poisson (1781–1840) wrote: "It was desirable for geometers to review the fundamental questions of mechanics from the standpoint of physics. In order to discover the laws of motion and equilibrium, they needed to take a purely abstract point of view; and in these abstractions Lagrange went as far as one could imagine when he replaced physical constraints between bodies with equations that related the individual coordinates of their points; this was the essence of his analytical mechanics. But along with this remarkable conception one could now erect physical mechanics. . . ."

Lagrange left it for succeeding generations to fill up his design with concrete physical contents. The method he worked out turned out to be directly applicable to the solution of technical problems, which he also abstracted completely in creating analytical mechanics. Krylov enumerates some immediate applications of Lagrangean mechanics: Poncelet's theory of mechanisms, engineering calculations in construction (in particular for large iron bridges needed to develop the railroads), ballistics problems arising in the shift from smooth-bore to rifled weapons after the Crimean War, and the theory of gyroscopes. He concludes, "In 1805 at Trafalgar, Nelson's ships attacked from the distance of a pistol shot and were boarded. At Tsushima the firing was at a distance of about 7000 meters and at Jutland it was at about 14,000 to 18,000 meters.[9] Since that time the firing distance in battle has increased significantly, and at such distances a whole series of complex gyroscopic instruments are needed to achieve accuracy. The calculations are all done with the Lagrangean equations.

One can bring an uncountable set of examples from technology and physics, but this is enough to see the value that Lagrange's remarkable

[9]Trafalgar was a decisive battle in the Napoleonic Wars; Nelson won the battle for the British despite being boarded but died in the process. The Battle of Tsushima took place in 1905 during the Russo-Japanese War and the Battle of Jutland in 1916 during World War I.—*Transl.*

essay has had in the general development of science and technology in all its fields, and how right Lagrange was, without going into detail, to give his account the most general analytic form; therefore his methods are uniquely applicable to calculating the motion of heavenly bodies as well as the rocking of ships in rough seas, ship propeller screws, the flight of 16-inch shells, and the motion of electrons in the atom. From this we can judge the unusual genius of the creator of these methods—Joseph Louis Lagrange." These lines were written in 1936.

Celestial Mechanics

Among the several types of problems in mechanics that Lagrange considered, celestial mechanics no doubt took first priority. That was the system of values in 18th century mathematics and not one of the strongest mathematicians could bypass problems on the agreement of the law of universal gravitation with the results of direct astronomical observations. We have seen that Lagrange began to study such problems while still in Turin and energetically continued these studies in Berlin. All the fundamental problems of celestial mechanics were in Lagrange's field of vision. He worked out techniques for calculating the elements of planetary and comet orbits from three observations. And here is another characteristic detail: his presentation of the method was not accompanied by a single concrete calculation of an orbit. Lagrange saw his role only as solving a mathematics problem, after which the method would pass into the hands of those who would use it to calculate: "I refrain from [giving] all the details, but I flatter myself with the hope that not a single smart calculator will be unable to apply the theory set out in this work to a comet." This creates the impression that Lagrange had no appetite for concrete examples. The method, unproven in practice, had its weak points, despite all its depth. An essential practical adaptation of the method is associated with the name of Karl Friedrich Gauss (1777–1855), who was always computing orbits and needed to rush so that observers could find a lost asteroid or so that his calculations could be used to observe a comet directly. And the corresponding method, essentially created by Lagrange, bears Gauss' name.

The basic difficulty was that, as we have explained, a sufficiently precise description of the motion of heavenly bodies requires that we take into account the interactions among several bodies: for the motion of the moon we are in practice talking about its interaction not only with the earth but also with the sun; for the motion of the large planets Saturn and Jupiter we must include their mutual attraction. Moreover, comparisons of observations that began in ancient times revealed a steady deviation from

Kepler's laws—"inequalities." It was necessary to clarify whether in fact these "inequalities" could be explained within the framework of the law of universal gravitation by the "interference" of the three bodies. The enthusiasm generated by Newton's *Principia* lay not only in that he derived Kepler's laws from the law of universal gravitation but also in that he succeeded in explaining certain "inequalities" in the moon's motion within the framework of this law. Newton's baton was picked up by Euler, Clairaut, and d'Alembert. Explaining the inequalities turned out to be hard work, and more than once despairing scientists began to doubt the universality of the law of universal gravitation.

The most natural thing would obviously have been to solve the three-body problem: describe the motion of three bodies that interact with each other according to the law of universal gravitation. It soon became clear that it was impossible to do this, but in 1772 Lagrange cleared up the situation as much as possible. With great art he showed that the original system of 18th-order differential equations could be transformed to a sixth-order system, but the form of this system gave no hope of further success. Then he delineated the cases where the integration could be carried out: in one case all three bodies are collinear at the initial moment of time, and in another they are at the vertices of an equilateral triangle with special relations on the remaining parameters. Lagrange considered these equations purely out of curiosity, but the explanation called to mind that each of the asteroids in the Jupiter group makes a nearly equilateral triangle with Jupiter and the sun.

The next possibility lay in the fact that in a triangle the bodies are usually unequal, and it is natural to consider the interaction of two bodies with a perturbation owing to the third. Lagrange began systematically to work out the mathematical theory of perturbations, whose basics had already been set out by his great predecessors. In a perturbation it is natural to assume that an orbit remains elliptical but that its parameters vary somewhat. Two types of perturbations are singled out, periodic and secular. Periodic perturbations essentially depend on the position of the body in its orbit and are compensated for on average in due course. Secular perturbations are determined only by the mutual positions of the orbits as a whole, and can accumulate and lead to an instability in the solar system. The latter was the reason for the intent interest in secular perturbations. On the other hand, for the study of perturbations during comparatively short time intervals, necessary in the case of periodic perturbations, there was not yet sufficient observational data, and at the time the imprecise observations of the ancients were used in practice (this was true for the study of secular perturbations as well). The periods of the perturbations could

be greatly exceeded by the periods of revolution and long-period perturbations could be seen as secular. The most important problem was to learn how to distinguish them.

Lagrange, working on the problem of secular perturbations, departed from his usual habit and was constantly guided by obvious numerical examples. He worked on these problems in parallel with Laplace (1749–1827), who was younger but had already proved himself. Their styles of scientific research were extremely different. Laplace's reference points were completely specific problems of celestial mechanics, and a method was for him only the means to achieve concrete goals. He was never interested in the abstract development of a method outside the requirements of specific problems. The strong and weak sides of each of these great scientists were apparent in their work on related problems. Laplace showed that to first order there are no secular perturbations for the major semiaxes of the orbits of Jupiter and Saturn (candidates for this had been long-period perturbations with huge period). Laplace was certain that an analogous claim was valid for all the planets and although this would not prove the stability of the solar system (perturbations were considered only to the first order), it no doubt would be serious step in this direction. Laplace unsuccessfully tried to find a general proof but using his general method Lagrange obtained a proof "with the stroke of a pen," as Carl Jacobi (1804–1851) described it.

And here is a contradictory example. Lagrange made a great effort trying to explain the secular velocity of the moon's mean motion, which was found in 1693 by Galileo Galilei (1656–1742), the discoverer of a significant number of the "inequalities" that were known at the time. Lagrange tried to use his favorite trick with the imperfect sphericity of the moon and then with the analogous property of the earth. Trying all the possibilities that occurred to him, Lagrange arrived at the conclusion that either the ancient observations contained major flaws or that this effect could not at all be explained within the scope of the law of universal gravitation. Simultaneously he worked out a technique for accounting for higher-order terms in secular perturbations. He found that in the case of Jupiter and Saturn these terms are inessential and extrapolated this observation to all the remaining cases. Laplace, having much greater computational experience, understood that the situation with satellites could be completely different because of their faster rotation. He discovered at the start that the terms found by Lagrange made an essential contribution for the moons of Jupiter and then, carrying out the same calculations for the moon, obtained Galileo's velocity.

The fruitful scientific collaboration of Lagrange and Laplace did not develop into a quarrel only thanks to Lagrange's striking tact and restraint.

Laplace, who was ambitious and easily carried away, often acted insultingly with his unfounded claims and even incorrect behavior. A characteristic episode occurred in 1774 when Laplace became acquainted with Lagrange's work on secular perturbations, which had been sent before its publication to Paris, where Laplace was living. Laplace quickly saw additional possibilities and published an article of his own, in advance of Lagrange's article. Laplace prefaced his paper with the words, "I could not have undertaken this work if I had not read M. Lagrange's excellent work, which was sent to the Academy and is intended to appear in subsequent volumes." He added various arguments in favor of his haste and spoke of his desire to acquaint the public sooner with all the possibilities of Lagrange's method, but his lack of tact gave rise to doubt. And Lagrange... thanked Laplace for improving his method, since "from this science one can only win." In 1779 Lagrange wrote to Laplace, "I consider that quarrels are completely useless for scientific success and only lead to losing time and peace...." All his life he strictly followed this rule.

Arithmetic Works

Although mechanics was Lagrange's main activity during his Berlin period, other mathematical questions fell into his field of view, among these several problems in number theory. He doubtless studied them under Euler's influence. Nine small works in all were devoted to number theory. They have the character of independent studies and are little masterpieces that do not show the intention of creating a superhighway, which was characteristic of his work in mechanics. Perhaps these were exercises during his leisure time away from his main work. Thus, Lagrange followed in Euler's footsteps: he proved that quadratic irrationals, and only these, can be expanded as periodic continued fractions (a claim that Euler made without proof), he continued work on the Fermat–Pell equation, he studied quadratic residues, he made some advances in the proof of the law of quadratic reciprocity that Euler had stated. His proof of Wilson's theorem ($(p-1)!+1$ is divisible by p for any prime p) is instructive, being based on a connection to Fermat's Little Theorem and essentially using polynomials over a finite field. Lagrange's theorem that real numbers can be approximated by rationals was widely known. His best known result in number theory is that any natural number can be represented as the sum of no more than four squares. This claim arose with Fermat and Euler apparently tried to prove it.

Algebraic Reflections

Lagrange studied various aspects of problems on algebraic equations and systems of equations. Some of these problems were inspired by his research in celestial mechanics. He was interested in the approximate calculation and the extraction of roots, as well as the elimination of unknowns in systems of algebraic equations. But one of Lagrange's works, in the words of Augustin-Louis Cauchy (1789–1857), signified the beginning of a new era in algebra.

In 1770–1771 Lagrange published his memoir *Reflections on the Algebraic Solution of Equations*, which he had no doubt conceived while still in Turin. Strictly speaking, this is an entire book that takes up over 200 pages. Together with *Analytical Mechanics*, this is the high point of Lagrange's creativity.

The 16th century saw a succession of discoveries of formulas for the solution of third- and fourth-degree equations, and then for two centuries there was no success in finding a formula for the fifth-degree equation. A number of notable problems appeared that diverted mathematicians from this enigmatic problem and consoled them. However a few worthy mathematicians, among them Leibniz (1646–1716) and Euler did not give up hope. Everyone felt that it would be good instead to skillfully find a formula for every degree, i.e., actually to find a single method that would be valid for all degrees. Ehrenfried von Tschirnhaus (1651–1708) told his friend Leibniz that he had thought of a universal substitution to transform the general nth-degree equation into an equation with two terms, $y^n+a=0$, which would therefore would have a solution in radicals! This substitution gave the well-known formula for $n=3$ and was valid for $n=5$. Leibniz had to give his friend some bad news: in order to find the coefficients in the substitution for $n=5$ one had to solve an equation of degree greater than 5. Then Euler discovered that he could obtain a formula for $n=3$ and $n=4$ by making a substitution of the form $x=\sqrt[n]{A}+\cdots+\sqrt[n]{F}$, but could not advance further.

The situation no doubt required deeper reasoning and who but Lagrange could undertake this work. After all, he had already shown himself to be the unsurpassed master of getting to the deep nature of a problem and clarifying its general structure, where others saw disparate situations. He began by studying formulas for $n\le 4$, paying special attention to expressions under the nth root sign. For the quadratic equation $x^2+ax+b=0$ this is $\Delta=\frac{a^2}{4}-b$, and for the cubic $x^3+ax+b=0$ it is $\Delta_\pm=-\frac{b}{2}\pm\sqrt{(\frac{b}{2})^2+(\frac{a}{3})^3}$ (here $x=\sqrt[3]{\Delta_+}+\sqrt[3]{\Delta_-}$). The quantities $\Delta_\pm$ are the roots of a quadratic equation whose coefficients are rational expressions (i.e., use arithmetic

operations) in the coefficients of the original equation. Lagrange tried to express $\Delta_\pm$ in terms of the roots x_1, x_2, x_3 and noted that $\Delta = x_1 + x_2\varepsilon + x_3\varepsilon^2$, where ε is a root of the equation $y^3 = 1$ other than 1.

Here we have to stop and consider how Lagrange imagined this root. Today this question would present no difficulty since we have two complex roots $\varepsilon_\pm = -\frac{1}{2} \pm i\frac{\sqrt{3}}{2}$, but Lagrange was not able to work with complex roots (it was learned later how to do the necessary calculation). All the same, he worked decisively with "imaginary" roots in the firm belief that the cubic equation always had three roots (counting multiplicities). N. Bourbaki wrote, "...Lagrange, like Euler and all his contemporaries, does not hesitate to argue formally in a 'field of roots' of a polynomial (that is to say, in his language, to consider 'imaginary roots' of this polynomial); the Mathematics of his time had not supplied any justification for this type of argument. Also Gauss, deliberately hostile, from its beginnings, to the frantic formalism of the 18th century, raises himself powerfully, in his dissertation, against this abuse."[10]

Thus, the two roots of 1 give $\Delta_\pm$. In fact we are unable to distinguish the roots x_1, x_2, x_3 in advance but, as we have numbered them, the function $\Delta(x_1, x_2, x_3) = x_1 + x_2\varepsilon + x_3\varepsilon^2$ can only take the two values $\Delta_\pm$ under any of the $3! = 6$ substitutions. This was Lagrange's decisive observation! For the quadratic equation, $\Delta = (x_1 - x_2)^2$ and does not in general change when we rearrange the roots. In the case of a fourth-degree equation the expressions under the fourth-root sign have the form $x_1x_2 + x_3x_4$, where x_j are the roots, and can take only three distinct values under the $4! = 24$ ways of numbering the roots.

Here it is easy to verify that if we have a function that is a rational expression in the roots of an nth-degree equation and takes only q values under all possible permutations of the roots, then the function is a root of an equation of degree q whose coefficients can be expressed rationally in terms of the coefficients of the original equation. Lagrange called this observation a "true principle and, so to speak, metaphysical equations of the third and fourth degrees." For exactly this reason the solution of a cubic equation reduces to a quadratic and that of a quartic equation to a cubic.

We have to search for rational functions of the roots that take $q < n$ values for all possible permutations. But as n grows the rapid growth of the number of permutations becomes a big problem. First of all we have to note that the coefficients of the original equation are rational functions of the roots that do not in general change under rearrangement of the roots ($q = 1$), but we have to look for less trivial possibilities. Lagrange calls

[10] *Elements of the History of Mathematics*, p. 92.

expressions of the form $x_1 + x_2\varepsilon + \cdots + x_n\varepsilon^{n-1}$ resolvents, where $\varepsilon \neq 1$ is a root of unity like those in the formulas for the quadratic and cubic equations. Their absence for the quartic equation is naturally connected to the fact that the number 4 is not prime. One could have expected that the resolvents also had to appear in formulas for the equations of higher degree, but this is what the calculations show: the function $\Delta(x_1, \ldots, x_n)$ takes $(n-1)!$ values under all permutations. For $n \leq 3$, we have $(n-1)! < n$. Thus, Δ is a root of an equation of degree $(n-1)!$ with coefficients that can be expressed rationally in terms of the original ones.

When n is prime we can rewrite this claim: Δ is a root of an equation of degree $n-1$ whose coefficients are, in turn, roots of an equation of degree $(n-2)!$ with coefficients that are expressed rationally in terms of the original ones. In the case $n = 5$ the coefficients of the fourth-degree equation are roots of a sixth-degree equation. It becomes clear how the equations of larger degree arose in the constructions of Tschirnhaus and Etienne Bézout (1730–1783). Lagrange concluded, "From this it follows that it is highly doubtful the methods we have considered can give a complete solution to the equation of the fifth degree."

Furthermore, it is natural not to restrict ourselves to resolvents and to explain whether there are not other functions of the roots that take no more than q values. Because of this Lagrange investigated the group of permutations, essentially laying the foundations for group theory. As soon as the terminology of groups appeared, a series of Lagrange's assertions automatically turned into theorems of group theory. Let a function $\delta(x_1, \ldots, x_n)$ of the roots take q values under rearrangement; then there is a subset (subgroup!) of $\frac{n!}{q}$ permutations that the function δ does not change. From this it follows, in particular, that q divides $n!$, and so it is essential to study subgroups of the group of permutations. If we describe all "largest" subgroups of this group, namely subgroups of $\frac{5!}{q}$ elements, where $1 < q < 5$, then we will describe all functions of the roots that take $q < 5$ values. Here Lagrange stopped.

He did not doubt that this is the only way to obtain a formula, but did not obtain definitive results: "Here, if I am not mistaken, are the true principles of the solution of equations and the analysis that is most suitable to arrive at a solution; as we have seen, everything reduces to some calculation of combinations with whose help the results which we must await are obtained a priori."

The group of permutations was studied in detail by Cauchy. Paolo Ruffini (1765–1822) proved the absence of nontrivial functions of the roots of the fifth-degree equation that take fewer than five values, being certain that he had proven the unsolvability of the fifth-degree equation in radicals.

However it remained to prove that the existence of such functions is indeed necessary for the existence of the formula we need. A complete unsolvability proof was given by Niels Abel (1802–1829). Before this there was Gauss' work on constructing regular polygons with compass and straightedge or, equivalently, on expressing the roots of the equation $y^n - 1 = 0$ using square roots. In it, headache-inducing tricks with permutations of the roots allowed him to solve a two thousand-year-old problem, the construction of the regular 17-gon, that is discussed in the chapter on Gauss. The problem of the solvability of algebraic equations found its definitive solution in the theory of Évariste Galois (1811–1832). But Lagrange was the first.... Incidentally, the connection between roots and rearrangements was discovered at approximately the same time by Alexandre-Théophile Vandermonde (1735–1796). Although he accomplished less, Vandermonde saw the major point and it is not right that Lagrange has overshadowed him in the history of mathematics.

Crisis

Mathematics was Lagrange's only passion, and it was enough to fill his entire life and to bring him many happy moments. Everything was devoted to his scientific work. Delambre tells us of Lagrange's relation to music: "I love it because it isolates me; I hear the first three notes, but at the fourth I discern nothing, I give myself up to my thoughts, nothing interrupts me, and then I solve the most difficult problems." It was characteristic of Lagrange that the great goals of knowing the truth and of world harmony were not bound up with his personal ambitions, with the desire to compete and outstrip his contemporaries. If he learned that someone had successfully investigated a problem on which he himself had been working, he immediately stopped thinking about it with the sincere feeling of "freedom from obligation." Thanks to this Lagrange enjoyed unusual emotional stability, which gave him the strength to overcome the difficulties of life and not curtail his intense work.

Only one thing could upset Lagrange—losing his orientation, being unsure that he had chosen the right goals. This feeling started to appear soon after he moved to Berlin. In 1772 he wrote to d'Alembert, "Does it not seem to You that higher geometry is rather approaching a decline, and that only You and Euler maintain it?" This was written by a scholar who was at the peak of his powers (he was 36 years old), whose *Analytical Mechanics* had begun to take shape, and who had just published an memoir on algebra that would determine the development of algebra for the next 100 years!

This statement deserves some thought. It seems that Lagrange saw

what he would work on for the coming 10–15 years, but a longer perspective seemed dubious to him. Possibly the particular style of his work had begun to tell on him. He had mapped out his basic directions in his youth, followed them with a share of his well-known conservatism, and not without reason hoped to complete the problems he had posed in the foreseeable future. A feeling of the end of mathematics probably could not have arisen with Euler, who throughout his long scientific life actively sought out new problems and went from one problem to another without being afraid to leave much unfinished. We should pay attention to the fact that Lagrange did not put himself in the same rank as Euler and d'Alembert. This was not the appearance of formal modesty. It was also characteristic that he envied his contemporaries who could easily find new problems, such as Gaspard Monge (1746–1818): "This devil Monge is always full of new and daring ideas" or "This scamp with his theory of the generation of surfaces is headed for immortality."

This feeling of the decline of mathematics did not abandon Lagrange. On September 21, 1781 he again wrote to d'Alembert: "Furthermore, I am beginning to feel that my inertial force is increasing little by little, and I cannot say whether I will still be doing Geometry ten years from now. It also seems to me that the mine is almost too deep already, and that unless we discover new veins we will have to abandon it sooner or later.

Physics and Chemistry now offer more brilliant riches and are easier to work; also the century's taste appears to have completely turned in that direction, and it is not impossible that the positions of Geometry in the Academies will one day become the way the chairs of Arabic are now in the Universities."[11]

It is natural to be puzzled by this. In celestial mechanics what was outlined was coming to an end while in algebra the language was only being worked out. Rough results were being obtained but the program was still not sufficiently defined and work had to be redirected. But the psychological laws of scientific creation are like that: one person cannot advance infinitely far on a difficult problem. The material has to settle and needed results such as Gauss', which confirmed the great effectiveness of working with permutations of roots. For Abel and Galois both Lagrange's and Gauss' work were key.

In Paris

His premonitions did not deceive Lagrange. In 1787, soon after the death of Friedrich II, he moved to Paris and essentially stopped his mathemat-

[11] *Oeuvres de Lagrange*, Vol. 13, p. 368.

ical activity. Lagrange was 51. In the single year of 1783 both Euler and d'Alembert had passed away. Lagrange was enthusiastically welcomed by the French scientists; now he was without a doubt "the leading geometer in Europe" and only Laplace could seriously compete with him. The court was not indifferent to Lagrange. He was diverted unusually easily from geometry towards work in philosophy, chemistry, history, and medicine. Could Lagrange have hoped to begin a new scientific life? The Paris scene was one of a range of scientific activities. Scientific circles flourished and contacts between scientists of different specialties were popular. The chemist Antoine Lavoisier (1743–1794) was particularly active in establishing such connections. Scientists were actively interested in social problems and the role of science in the life of the state.

Lagrange did not leave mathematics: his works would still appear, he would be actively interested in the works of others, he would still talk about his pedagogical activities and his own textbooks, but the peak of his scientific activities had already passed. And the time would soon come when the majority of French scientists (perhaps excluding Laplace) would interrupt their usual work.

First there was the revolution, in which scientists took a most active part. They had never before been able to influence the life of the country directly. They were in the Municipality and the Constituent and Legislative Assemblies. The astronomer Jean Bailly (1736–1793) became the mayor of Paris, the mathematician Lazare Carnot (1753–1823) was in charge of the defense of France (he was called the "organizer of victory"), and Monge became the Naval Minister. Also, scientific work aimed at solving practical problems sharply increased.

Lagrange kept to the sidelines of politics. The law of 1793 required foreigners to leave France, but a special decree of the Committee of Public Safety made an exception for Lagrange. During the most difficult days he did not leave France, sharing the fate of his colleagues. Their participation in political life cost Bailly and Condorcet their lives. Lavoisier was executed as a "tax farmer."[12] Lagrange watched these happenings intently. Delambre preserved the words Lagrange spoke after Lavoisier was guillotined: "It took one moment to remove this head and, perhaps one hundred years will not be enough for a similar one to appear."

As a scientist, Lagrange conscientiously fulfilled all assignments. A large number of commissions and offices gradually accumulated in which it was customary to include scientists. He worked on measuring distance at sea and estimating the supply of bread and meat in the country in or-

[12]Tax collector.—*Transl.*

der to estimate the probability of famine. One of his writings included a calculation of the explosive power of the powder in a gun barrel (it was not published during the author's life—perhaps this was one of the first classified scientific works).

Scientists were especially included in the work of the Commission of Measures and Weights. Today it is not easy to make sense of why, in a time of hunger and devastation and with a constant danger of war, such enormous attention was paid to reforming the system of measures and weights. A lack of coordination in the system of measures explained many difficulties and people spoke with great emotional heat about how imperfections in measures were a means for exploiting the populace. Another aspect of the issue was the inconvenience of the system of measures, which was an international problem, and a successfully conceived system could have served to increase the prestige of the Revolution in the international arena. From this point of view it was important to choose units that were not tied to any national tradition. The bishop of the city of Autun and future Napoleonic diplomat, Charles Maurice de Talleyrand, proposed to use an idea that had come from Huygens and to take as the basis for length a pendulum whose period of oscillation was equal to one second. But the idea of taking a portion of a meridian as the unit of length won out.

The work was thought through at the highest level. Lavoisier and Juste Haüy (1743–1822) measured the weight of water. Geodesic measurements were begun for which there were no methods, and relations with Spain as well as the situation in places in France itself interfered with them. But the revolutionary Convention was impatient to introduce a system of measurements "for all time and all peoples" (the motto later engraved on the standard meter). Problems of the metric system were adjudicated by the Convention in 1793 together with the sharpest questions. The Commission was accused of being slow and some of its members were removed for "insufficient revolutionary virtue and hatred of tyranny." Such an accusation could be enough to send one to the guillotine!

Lagrange's duties on the Commission were not of such a sharp, theoretical nature. He studied the choice of a basis for the new system and proposed taking the prime number 11 as its foundation. He thought it important that no part of the fundamental unit turn into an independent unit over time. In the end everything was constructed on the basis of the decimal system.

The Academy was closed at the time and was reborn as the Institute of France, and Lagrange stood at the head of the Physics–Mathematics section.

Pedagogical Activities

Revolutionary France paid a lot of attention to educational reform during the stormy and rich changes of 1793–1795. "After bread, education is the most important need of the people," said Georges Danton. The education of the people was thought of no less than supplying the people with bread. The École Normale (Normal School) was organized for the preparation of teachers and the École Polytechnique (Polytechnic School) for that of military engineers. The École Polytechnique was first called the Central School of Public Works. Never before having taught, Lagrange enthusiastically gave lectures in both schools. With his interest in thinking through the foundations, his lectures were an occasion for rethinking modern mathematics, its fundamental ideas, and the connections among different fields. Two books were born from his lectures, *The Theory of Analytic Functions* in 1797 and *Lectures on the Calculus of Functions* in 1801.

A fundamental conception of Lagrange eloquently characterizes the full title of the first book: "The Theory of Analytic Functions containing the Principles of differential Calculus, freed from all consideration of the infinitely small, of vanishing quantities, of limits and fluxions, and reduced to the algebraic analysis of finite quantities." The point is that almost two centuries of mathematics had resolutely used infinitesimals but this concept remained fuzzy and there was no convincing foundation for the rules for working with them. However there was no doubt that the formalism that had been worked out allowed one to obtain correct results that had not been successfully obtained by other means, and to refrain from the language of infinitesimals (as was initially proposed) was already impossible. A situation that had gone on for an impermissibly long time remained confused.

In 1784 the Berlin Academy proposed as the theme for a competition to construct "a clear and precise theory of what in mathematics is called infinite. It is known that higher geometry constantly deals with the infinitely large and the infinitely small. However the ancient geometers and even analysts carefully avoided everything that touched the infinite, and the great contemporary analysts recognize that the expression 'infinite quantity' is contradictory. The Academy therefore wishes to obtain an explanation of how from a contradictory assumption so many true theorems have been derived, and that a true, clear—in a word authentic mathematical principle can be expressed which would replace the infinite, not making too difficult or long the investigations produced using this means." The initiator of the competition was no doubt Lagrange.

His point of view consisted of the fact that the idea of infinitely small is

in fact contradictory, but the calculus had been constructed so successfully that the errors that arose compensated for each other and a correct answer was always obtained. As early as Volume II of *Mélanges de Turin*, Lagrange wrote in 1760–1761 that "calculus itself corrects the false assumptions made in it." As Felix Klein (1849–1925) wrote, he "repudiated analysis as a general discipline, understanding it simply as a collection of formal rules related to particular special functions" and "such self-restriction removed for that time a whole series of difficulties." Thus, Lagrange's point of view was that in principle one could not make infinitesimal calculus meaningful, that one needed to look at it formally, convince oneself somehow that the errors in fact cancel out, and then calmly use it.

We have again run into the readiness of 18th century mathematicians to deal with purely formal procedures (we already spoke about work with "imaginary" roots of equations). In the 20th century an analogous point of view arose in the realm of Hilbert's program on the foundations of mathematics, in which infinities are taken as formal objects and we only need to convince ourselves that the rules of handling them are not inconsistent in order to be sure that the statements about finite objects obtained from them are correct.

Lagrange's prognosis was not borne out. A meaningful foundation of analysis on the basis of limits had long been advanced by d'Alembert. But Lagrange, as is evident from the title of his book, rejected this basis along with everyone else. His thought process was very interesting. He remarked that there are no problems in constructing the rules of differentiation of polynomials, and in the same algebraic language one could construct differential calculus for functions that can be expanded in infinite series. Lagrange was sure, as were his predecessors, that every function admits such an expansion (only Cauchy disputed this opinion) Lagrange relied on his intuition in practical analysis, which suggested that all the functions encountered in applications admit a series expansion. Within a century Karl Weierstrass (1815–1897) would construct the theory of analytic functions of a complex variable along this path, but as a means of grounding the analysis of real functions this program turned out to be untenable. Bourbaki wrote, "The monumental work of Lagrange represents an attempt to base analysis on one of the most arguable of the Newtonian concepts, that which confuses the notions of arbitrary functions and that of functions which can be expanded in a power series, and to draw from that[13] (by consideration of the coefficient of the term of the first order in the series) the notion of differentiation. Of course, a mathematician of the quality of Lagrange could

[13]From the series.—S.G.

not fail to obtain in this case important and useful results, as for example (and in a way which was in fact independent of the point of departure that we have just indicated) the general proof of the formula of Taylor with the expression for the remainder in the form of an integral, and its evaluation by the theorem of the mean; in any case the work of Lagrange is at the origin of the method of Weierstrass in the theory of functions of a complex variable, as well as the modern algebraic theory of formal series. But, from the point of view of its immediate objective, it represents a retreat rather than progress."[14]

It is indicative that Lagrange never confused the problem of the foundations of analysis with a particular construction of analysis and its applications. In the preface to the second edition of *Analytical Mechanics* (1811) Lagrange wrote, "I have kept the ordinary notation of the differential calculus because it fits the system of infinitesimals adopted in this treatise. Once the spirit of this system has been grasped well and the accuracy of its results established by either geometrical methods or by the analytical method of derived functions, the infinitesimal calculus can then be applied as a certain and manageable tool to shorten and simplify the demonstrations."[15]

Lagrange's remarkable method for finding a constrained extremum first appeared in the pages of *The Theory of Analytic Functions*. In finding the largest and smallest values of a function of several variables, say $f(x, y)$, the problem inevitably arises of finding an extremum under some condition on the variables, e.g., $\varphi(x, y) = 0$, where it is not always convenient to pass to the smallest number of parameters. Finding an extremum of a function of one variable on an interval reduces to equating the values of the function at the interior stationary points and at the endpoints. To find an extremum in a domain D of several variables we must compare the values of f at the interior stationary points with the values on the boundary, but the boundary no longer consists of two points and the question arises of the constrained extremum on the boundary. However this is only one of the many situations where a constrained extremum arises.

Lagrange noted that the problem cited above reduces to finding those λ for which the function $f + \lambda\varphi$ has stationary points when $\varphi = 0$. This gives rise to a system of equations for finding these points. Analogously, we consider the case of any number of parameters and constraints. "The method of Lagrange multipliers" grew out of Lagrange's results on mechanical systems with constraints. In applications, Lagrange multipliers frequently have an interesting interpretation. Today the sphere of applications of Lagrange's idea has broadened. In particular, linear programming

[14] *Elements of the History of Mathematics*, p. 196.
[15] *Analytical Mechanics*, p. 8.

developed from it and in applications to economics problems the Lagrange multipliers can often be interpreted in the language of prices.

The Last Years

Under the Directory (1795–1799) and Consulate (1799–1804), Lagrange's situation was strengthened. During the Empire that followed he became a count, a senator, and a chevalier of the Order of the Legion of Honor. Napoleon was not indifferent to mathematics and understood Lagrange's true value very well. The Emperor's daily life left him little time for scientific patronage. He limited himself to dispensing awards with short testimonials intended for history. He called Lagrange "the Cheops pyramid of science."

Lagrange died on April 10, 1813. Delambre recalled the surprising peace with which he met his last hour: "I felt that I was dying; my body weakened little by little, my mental and physical abilities faded imperceptibly; I noted with pleasure the very gradual decrease in my strength and I arrived at the end without pain, without regrets, and along a very gentle path. . . . I made my career; I acquired some fame in Mathematics. I hated no one, I harmed no one, and I must finish it well. . . ."[16]

In his stormy century Lagrange was able to lead a measured life. His contemporaries worked to recall details that they could use to enliven his biography. They did not tell stories about him as they did for Laplace. Krylov remarked that the story of the Italian-style meal in Paris told above was perhaps the only adventure in Lagrange's life. They recalled that he was able to improve Lambert's position in Berlin, that during the terrible year of 1793 he was not afraid to defend Delambre, who was going to be fired from the Commission on Measures, that he touchingly looked after Poisson when the latter was his student at the École Polytechnique, and that surprisingly he knew how to listen to his interlocutors. And sometimes a small but significant trait came out: Lagrange's whole being "was imbued with quiet irony."

And unexpectedly precisely this modest man came to be taken as the image of a great scientist and person, and not just by mathematicians. Goethe wrote: "The mathematician is perfect only in so far as he is a perfect man, in so far as he senses in himself the beauty of truth; only then will his work be thorough, transparent, prudent, pure, clear, graceful, indeed elegant. All this is needed, in order to resemble Lagrange."[17] Elsewhere he wrote

[16] *Oeuvres de Lagrange*, Vol. 1, p. xliv.

[17] Thanks to my colleague Jens Kruse for help in locating the original German and rendering this in English.—*Transl.*

about Lagrange, "He was a good man, and on that very account a great man. For when a good man is gifted with talent, he always works morally for the salvation of the world—as poet, philosopher, artist, or what not."[18]

Today, Euler and Lagrange are considered to be the greatest mathematicians of the 18th century, teacher and student whose gifts strikingly complemented one another. Euler, striving to see as far ahead as possible, spoke about things for which there was still no appropriate language and left a legacy of problems that served as reference points for a long time. Lagrange, reaching for deep structure in everything and trying to create a picture with no blank spots, passed along to succeeding generations the language and methods that would be enough to solve new problems for many years.

[18] *Conversations of Goethe with Johann Peter Eckermann*, translated by John Oxenford, Da Capo Press, New York, 1998, p. 291.

Pierre-Simon Laplace

> Chancellor of the imperial Senate, receiving over 100,000 pounds annually, being no less diligent than a simple academician, Laplace tried to tie up all the irregularities and perturbations in the motion of the luminaries with the principle of universal gravitation, to extend the method of mathematical analysis to the phenomena of terrestrial physics, and to subject the phenomena of public life to his formulas, where the average person sees a mystery or blind luck. François Arago

On March 5, 1827 at nine o'clock in the morning the Marquis Laplace died, a peer of France, one of the first chevaliers of the Legion of Honor and worthy of its highest decoration, the Grand Cross. "What we know is nothing in comparison with what we do not know" were his last words. Laplace was called "the French Newton" and he died exactly one hundred years after Newton, who had been his idol.

Laplace's eulogies were delivered with some embarrassment. In his speech Fourier said, "I could also have, perhaps I should have, recalled the high political positions in which he was invested; but this recounting would only indirectly relate to the object of this discourse. It is the great geometer whose memory we celebrate. We have separated the immortal author of 'Celestial Mechanics' from all the accidental facts that involve neither his glory nor his genius. In effect, Sirs, what does it matter to posterity, which will have so many other details to forget, to learn or not learn that Laplace was for some moments a minister of a great state?"[1] It was embarrassing to his associates that Laplace was able to remain a republican and a monarchist, an atheist and a Catholic, and to receive

[1] Joseph Fourier, "Éloge historique de Laplace," *Mémoires de l'Académie des Sciences*, Paris, 1831.

Pierre-Simon Laplace.

honors under the Empire and after the Restoration. Incidentally, the former Jacobin Fourier also became a baron later on.

Beaumont-Paris, 1749–1789

The future marquis was born on March 23, 1749 into a peasant family in the little town of Beaumont in Normandy. Later he spoke of his childhood reluctantly and never saw his parents after he turned 21. Thanks to an unknown patron he completed his studies at a Benedictine college. At the age of 17 he was already teaching mathematics at the École Militaire (Military School).

Laplace began an intensive study of mathematics and mechanics. In 1770, armed with a letter of recommendation to the great d'Alembert, he set off for Paris. For a long time he was unable to start on the path of the recommendation until an idea fortunately came into his head—to lay out his understanding of mechanics in writing. The originality of the young man's idea made a strong impression on d'Alembert: "You recommended yourself to me, and this completely sufficient. My assistance is at Your service."

With d'Alembert's help it was arranged for Laplace to teach at the École Militaire and later he took the position of examiner at the Royal Artillery Corps, a position that was freed up after Bézout's death. In 1784 the young Napoleon Bonaparte brilliantly passed his examination. Laplace had the

opportunity to recall this in 1804: "I want to add my own greetings to the greetings of the people to the Emperor of France, whose career I had the happy privilege of launching twenty years ago, a career that has brought such glory to him and such happiness to France."

In 1772 Laplace was nominated for the Academy of Sciences[2] as an adjunct member (with a lower salary) in geometry[3] but he was not elected. Evidently, one reason for this was that the French scholars did not have a very favorable opinion about their young colleague. Lagrange took a more indulgent and optimistic position: "I am somewhat surprised at what You are writing to me about Laplace: he is boasting about his first successes—a shortcoming characteristic, in general, of very young people. However as they know more they usually become less presumptuous" (from a letter to Condorcet, the Permanent Secretary of the Academy of Sciences). Laplace was already thinking of moving to Berlin, to Lagrange, but in 1774 he obtained a position as adjunct member in mechanics.

Almost all of Laplace's scientific activities were devoted to celestial mechanics (see below). But his interests were significantly broader than that.

Thus during 1779–1784 he collaborated with Lavoisier on the most diverse questions (defining specific heat, the problem of phlogiston, atmospheric electricity): "I truly do not know how I got involved in working in physics, and You know, perhaps, that I would have done better by refraining from this; but I could not withstand the persistence of my friend Lavoisier, who put as much pleasantness and intelligence into this joint work as I could wish for. Moreover, since he is very rich he begrudged nothing in order to give the experiments the precision that was needed for such delicate research." Laplace also took part in public life: he joined a committee of the Academy of Sciences that investigated hospitals for the poor and sanitation in the city slaughterhouses. Laplace's prestige grew. In 1784 he became an academician in mechanics.

The path of the Beaumont peasant was not unique. At the end of the 18th century almost half the members of the Academy of Sciences came from simple family backgrounds. For example, Monge was the son of a village grinder, Fourier that of a tailor, and Poisson that of a soldier. The

[2]At that time, there were five academies in France. We note among these the Académie Française founded in 1635 by Cardinal Richelieu for the perfection of the French language and the composition of a dictionary, and the Académie des Sciences, founded in 1666. The Académie Française consisted of 40 lifetime members. New members were chosen to replace ones who died. The members of the Académie Française were called "immortals." The Académie des Sciences might more precisely have been called the Academy of Natural Sciences.

[3]In the 17th century, all of mathematics was called geometry. Up to that time, the Mathematics section of the Académie des Sciences was called the Geometry section.

participation of the upper class in science was usually limited to patronage and honorary membership in the Academy; d'Alembert regretted, "In our times so many patrons have been bred that it is impossible to extol and thank them all as one should."

In 1788 Laplace married. Within a year a son was born. The measured, prosperous life was interrupted by the events that decisively changed the life of the country.

Revolution, Empire, Restoration

A significant portion of the French scientific community was caught up in the revolutionary events. The astronomer Bailly, Laplace's friend, was the first mayor of Paris, Condorcet was a member of the Municipality, and the outstanding mathematician Monge was Naval Minister. In 1791 a number of academicians put forth their candidacies for the Legislative Assembly, including Condorcet and Lavoisier. In connection with this Jean-Paul Marat published an impassioned pamphlet called *Modern Charlatans*. At the same time it touched Laplace: "Among the best mathematics academicians are Laplace, Monge, and Cousin;[4] a species of automatons, accustomed to following well-known formulas and applying them to the blind, like a mill horse that is used to making a certain number of circles before stopping."

Laplace, together with Lagrange, Monge, and Lavoisier, joined the work of the commission on the metric system, whose goal was to create a single system of measures. During the Jacobean dictatorship Laplace was removed from the commission because of "insufficient revolutionary virtue and hatred of tyranny." In 1799 he returned to the commission and the standard meter and kilogram were made under his supervision.

In the summer of 1793, following an appeal by the Committee of Public Safety, a large group of scholars conducted scientific research on the organization of a defense against an expected attack. Laplace was not among them. He moved to the quiet town of Melun, where he began work on the multivolume *Celestial Mechanics*, the major work of his life.

In 1793 the Convention abolished the existing academies. In 1793–1794 several former academicians met their end at the guillotine. Condorcet was sentenced to death along with the Girondist deputies. The tax farmer Lavoisier was also condemned, by the "Law of Suspect." Bailly, whom Laplace tried to hide in his home at Melun, died on the scaffold.

Laplace returned to Paris after the Thermidor revolution in the autumn of 1794. Along with Lagrange and several of the best scientists, he took

[4] Jacques Cousin (1739–1800).

a position as professor at the École Normale. This new kind of educational institution had been conceived before the Convention. Enlisting top scientists in the role of teachers was something new. Later, the École Polytechnique was created to train engineers at the same high level. Laplace gave lectures in both schools. He became president of the Commission of Measures and Weights and worked actively with the Bureau of Longitude, which was created to oversee astronomical, geodesic, and time measurements.

In 1795 the Directory founded the Institut National des Sciences et Arts, (later called the Institut de France). The Institute was divided into sections, the first of which was called the Classe des Sciences Mathématiques et Physiques.

General Bonaparte supported all sorts of contacts with the Institute and took an active part in the work of the geometry section. During his Egyptian campaign he signed his proclamations "Bonaparte, Supreme Commander, member of the Institute."

The first two volumes of *Celestial Mechanics* appeared in 1799 and Laplace, literally within several days of the revolution of the 18th of Brumaire (November 12), gave the first volume to Napoleon. In his response the general said, "With gratitude I accept, citizen, the copy of Your excellent work that You sent me. The first six months that I can devote to it will go to reading Your excellent work."

After the establishment of the Consulate Napoleon decided to give the position of Minister of Internal Affairs to a scientist. The choice fell on Laplace, probably in view of his great fame and personal acquaintance with Napoleon. However Laplace's work as minister was not so successful. As opposed to his Cabinet colleagues Talleyrand and Joseph Fouché, Laplace did not know how to find his bearings and understand the Consul's thinking as a patron of the sciences. He went after royalism and religion, not without naivete: "Do not miss an opportunity to prove to your fellow citizens that superstition will profit from the changes stemming from the 18th of Brumaire no more than royalism will" (from a circular by Minister Laplace). After a few more months passed, Napoleon's brother Lucien replaced Laplace. In his memoirs, written on the island of St. Helena, Napoleon wrote, "The first-class geometer soon showed himself to be only a mediocre administrator; his first steps in this area convinced us that we had made a mistake with him. It is remarkable that not one question from practical life had arisen in Laplace's true sphere. He looked for accuracy and detail everywhere; his ideas were notable for being enigmatic; finally, he was completely filled with the spirit of the 'infinitely small', which he even carried over to administration."

Nevertheless, this change did not cut short the amiable relationship of Bonaparte and Laplace. On becoming First Consul, Bonaparte named Laplace as "gardien" of the Senate. Incidentally, this Senate played no role at all in public life. Laplace became chancellor of the Senate in 1803. Among the few acts of the Senate was a change in the revolutionary calendar according to a report written by Laplace. The order of the Legion of Honor was established and Laplace was among its first chevaliers. In 1808 he became a count of the Empire.

At the same time Laplace continued working on *Celestial Mechanics*. The third volume appeared in 1802, dedicated to Napoleon, "to the hero and pacifier of Europe, to whom France owes its prosperity, its greatness, and the most brilliant epoch of its glory." In his reply, Napoleon said "I truly regret that the force of circumstances kept me from the field of science." Somewhat later Emperor Napoleon wrote, "It seems to me that *Celestial Mechanics* increases the magnificence of our century." On August 12, 1812, at the Battle of Smolensk,[5] Napoleon received *The Analytical Theory of Probability* and again regretted, "At another time, having leisure time, I would have read Your 'Theory of Probability' with interest." And later, "Spreading and perfecting the mathematical sciences is closely joined to the welfare of the state."

Napoleon was actively involved in the affairs of the Institute. In 1801 a mandatory structure was introduced for members of the Institute. The members lined up after Mass at the Tuileries palace to be presented to the emperor. At that time they could present him with scientific works and obtain his "paternal" directions. While he patronized the exact sciences he mistrusted the humanities. In 1803 Napoleon eliminated the section of Moral and Political Sciences at the Institute. When he got word that discussions about politics were taking place in the section of French Language and Literature (the Second Section), he announced to Ségur,[6] "You preside over the Second Section of the Institute. I order You to tell it that I do not want them to talk about politics at their sessions. If the section will not obey, I will break it like a worthless stick."

Before the fall of Paris in 1814, the Senate took an unexpected action: at Talleyrand's initiative, it called for the return of the Bourbon dynasty. Laplace was one of the first to sign this decision. During the Hundred Days,[7] he did not leave the provinces.

[5]The city in Russia where the Russian and French armies fought a major battle during Napoleon's march toward Moscow.

[6]Count Louis-Philippe Ségur was a French soldier and diplomat under the king who became a member of Napoleon's entourage.—*Transl.*

[7]With the Bourbon return, Napoleon abdicated and went into exile on the island of Elba.

Under the Restoration the sections of the Institute were called academies again. The Academy of Sciences meekly removed the Monge and Carnot from their ranks for having been disloyal to the monarchy. Laplace was showered with honors. In the first year of the reign of Louis XVIII he became a marquis and peer of France, and received the Grand Cross of the Legion of Honor. In 1816 he became the president of the Bureau of Longitude and chairman of the commission for reorganizing the École Polytechnique, and was elected to the "Academy of Immortals," a rare distinction for someone in the exact sciences. Laplace's speeches in the Chamber of Peers were rare, colorless, and uncompromisingly monarchial. When part of the Institute protested against the introduction of censorship by Charles X, Laplace dissociated himself from the protest in print. Henri de Saint-Simon was indignant: "Sirs, who learn disorganized material, infinitely small quantities, algebra, and arithmetic! Who gave you the right to be on the front lines? . . . You brought from science only one observation, namely, that he who flatters the world's greats takes advantage of their goodwill and generosity."

Many stories about Laplace's behavior in the Academy of Sciences have been preserved. Here are two of them.

François Arago (1786–1853) and Poisson were competing for the same position in the Academy. Laplace announced that he had to give his preference to Poisson, who was older. A sharp exchange of opinions took place:

Lagrange: "But You yourself, Monsieur de Laplace, were elected as a member of the Academy when you had not yet done anything outstanding and showed only hope, and all Your great discoveries were made later."

Laplace: "And I still consider that the title of academician should be shown to young persons as a future reward, in order to stimulate their powers."

Hallé[8]**:** "You are like a coachman who attaches a wisp of hay to the end of a pole on his carriage to entice the horses. This kind of ruse ends in the horse becoming exhausted and dying."

Laplace had to concede.

Another time, in 1822, Fourier and Jean-Baptiste Biot (1774–1862) were running for the position of Permanent Secretary. Laplace took two ballots instead of one. The person next to him saw that he had written Fourier's name on both. Laplace put the ballots into a hat and asked his neighbor to choose one, tore up the other one, and announced loudly that he did not know for which candidate he had voted.

The Hundred Days was the period between his return to power and his final defeat at Waterloo.—*Transl.*

[8]Jean-Noël Hallé (1754–1822).

After Lagrange's death in 1813 Laplace's influence in the Academy of Sciences became especially strong. In 1826, a year before Laplace died, the young Abel appeared in Paris. He wrote, "Thus, *Celestial Mechanics* is complete. The author of such a work can look with satisfaction at the path he took in science." Elsewhere he wrote, "It is obvious that any theory of Laplace is much greater than what any mathematician of lesser stature could create. It seems to me that if you want to achieve anything in mathematics you need to study the masters, not the apprentices."

Celestial Mechanics

The start of Laplace's scientific activities came at a complex time. The great era of constructing infinitesimal analysis had ended. There were no problems on which the best mathematicians were concentrating their powers. It seemed to many that the days of pure mathematics were numbered. Even the multifaceted Lagrange, whose algebraic work defined his time, had stopped doing mathematics at some point. He had written to d'Alembert in 1772, "Does it not seem to You that higher geometry is rather approaching a decline, and that only You and Euler maintain it?"

Under these conditions the center of interest was located in applied mathematics, where undisputed priority went to the problem of building a theory of motion for heavenly bodies based on the law of universal gravitation.

Here is the prehistory of this problem. At the beginning of the 17th century Kepler, taking the viewpoint of Copernicus and with Tycho Brahe's scrupulous observations in mind, formulated three laws that were obeyed by the motion of the planets around the sun. Newton's genius of a conjecture was that these laws are consequences of a single universal law of universal gravitation that controls the interaction of heavenly bodies and the attraction of the earth. Terrestrial and celestial mechanics were united. Under the law of gravitation one could explain the motion of the moon, the ebb and flow of the tides, the precession of the equinoxes, and other effects. But it was not easy for Newton's theory to gain recognition. Huygens and Leibniz did not believe in it. Johann Bernoulli worked hard to explain the ellipticity of the orbits without using the law of gravitation. In France Newton was opposed by Descartes' followers, who took a contrary point of view on many issues. For example, in Newton's considerations it was important that the earth was flattened at the poles while the measurements of the French geodesists (which turned out to be wrong) showed that it was stretched out there instead. In 1727 Voltaire joked, ". . .in Paris they think the earth is stretched at the poles, like an egg, while in London it is compressed, like a pumpkin."

There was one way in which Newton's opponents had a strong position. A thorough analysis of observations showed that Kepler's laws were satisfied only approximately, and that small deviations could accumulate over time and sharply upset the stability of the solar system. Newton did not see a way to get around these "secular" perturbations: "...hardly noticeable inequalities, which can come from the interaction of the planets and comets..., probably, will increase over a rather long time until that time when, finally, the system will require the hands of the Creator to bring it into order." In reply, Leibniz noted "Newton and his adherents have an extraordinarily amusing idea of divine creation. From their point of view God must from time to time wind his world clock.... God created such an imperfect machine that he must at times clean it from dirt and even repair it, like a watchmaker corrects his work." The mathematical difficulty was that in deducing Kepler's laws from Newton's law the two-body problem (sun and planets) must be dealt with. The desire to account for the influence of still another object lead to the three-body problem, which to this day has not been solved in the general case.

Euler, Clairaut, and d'Alembert continued Newton's work. Euler studied perturbations in the motions of Jupiter and Saturn. All three gave their own variants of the theory of the moon's motion. Clairaut deduced equations for the three-body problem but left it with the words, "Let he who can, integrate." The most effective result was Clairaut's prediction of the exact time that Halley's comet would return. It was expected in 1758 but Clairaut's calculations showed that because of the influence of Jupiter's attraction it would be "held back" more than a year. Euler and Clairaut constructed a theory of motion for the earth that took the perturbed motion of the other planets into account.

In the 1770s Lagrange became interested in problems about the anomalies of the solar system. The young Laplace also started to become interested in them. Euler and d'Alembert sorted out a series of effects associated with the mutual attraction of Jupiter and Saturn, but one phenomenon remained unexplained—the so-called "great inequalities," discovered by Halley in 1676 when comparing modern and ancient observations with those of the ancients. It seemed that Jupiter was slowly but systematically accelerating, while Saturn was slowing down.

Laplace, like Euler and Lagrange before him, sought an approximate solution to the three-body problem, considering an infinite series of perturbation terms. To obtain an approximation formula he had to decide how many terms to take in the series and how large an error was made by discarding the remaining ones. For simple series students can do such exercises. But it was not understood how to proceed with the perturbation

series. Laplace thought he could succeed by taking the necessary number of terms and steadily comparing the result obtained with the given observations: "The extraordinary difficulty of the problems relating to the system of the world forced geometers to resort to approximations, which can always save them as long as the discarded terms do not turn out to have much influence. When the observations indicated such an influence to them they turned again to their analysis; in checking they always found a reason for the deviations they noted; they determined their laws and discovered the inequalities which had not yet been indicated by the observations. Thus one can say that nature itself assists the analytic perfection of theories based on the principle of universal gravitation." In the case of Jupiter and Saturn the noted anomalies arise from the fact that after every five revolutions of Jupiter and three revolutions of Saturn the planets occupy almost the same position and the perturbations build up. All the same, as Laplace's calculations showed, the perturbations do not accumulate without bound; they are not "secular" but rather periodic, with a huge period of 913 years. Thus, although the compensation takes place utterly slowly, the time comes when Jupiter begins to slow down and Saturn begins to speed up.

Halley's conjecture about the "great inequalities" was settled in 1784. "When I explained these inequalities and determined those that had already been computed with more attention than had been given up to now, I was convinced that all observations, ancient and modern, were presented by my theory in all their accuracy. Before they had seemed unexplainable by the law of universal gravitation; now they serve as one of its most striking confirmations. Such is the fate of this brilliant discovery: every difficulty that arose here turned into its triumph and this is the truest sign that it corresponds to the real system of nature."

Euler, d'Alembert, and Clairaut put a lot of effort into constructing a theory of motion for the moon that agrees with observations. The main effect that was necessary to explain is the rapid (41° per year) displacement of elliptical orbits. All three calculated displacements of no more than 20°. It was only in 1849 that Clairaut succeeded in making calculations accurate enough to obtain the needed displacement (there had been serious thought of putting correction terms into Newton's law). However still one more "detail" remained, one that was even noted by Halley in 1693. In analyzing Ptolemy's *Almagest* and medieval data on eclipses, he positively showed that the moon's motion was accelerating.

Laplace resolved this in 1787. The cause of the acceleration turned out to be a long-period oscillation that had been discovered earlier in the eccentricity of the earth's orbit; when the eccentricity decreases (the orbit becomes more like a circle), the mean velocity of the moon increases. This

was one more perturbation that seemed to be "secular" but turned out to have a long period!

Laplace did not propose even one conjecture in astronomy. He had the right to say, "Posterity will probably see with gratitude that the latest geometers did not leave even one astronomical phenomenon whose laws and principles were not discovered." He showed that Saturn's rings cannot be solid (Herschel confirmed this by observations during Laplace's lifetime). Laplace made the theory of the tides more precise in an essential way, and showed with the theory of perturbations how lunar observations can be used to determine the astronomical unit (the distance from the earth to the sun) and to determine the shape of the earth more accurately.

It stands to reason that Laplace did not bypass problems about the moons of Jupiter, which had been traditional for all the great astronomers since Galileo had discovered them. In 1774 this problem was chosen by the Academy of Sciences as the theme for its prize. In 1789 Laplace constructed a theory of motion for Jupiter's moons, taking into account the influence of the sun and their interaction.

The main problem that occupied Lagrange and Laplace during 1773–1784 was the problem of the stability of the solar system as a whole. Perturbations for all the planets had been systematically studied, and although a strict proof of stability had not been obtained there were no arguments about the agreement of all the anomalies with the theory of gravitation. Confidence in the theory of perturbations was such that when unexpected deviations were found in the motion of Uranus, Urbain Leverrier resolved to explain them by the existence of a new planet.

"Five geometers, Clairaut, Euler, D'Alembert, Lagrange, and Laplace, shared between them the world of which Newton had disclosed the existence. They explored it in all directions, penetrated into regions which had been supposed inaccessible, pointed out there a multitude of phenomena which observation had not yet detected; finally, and it is this which constitutes their imperishable glory, they reduced under the domain of a single principle, a single law, everything that was most refined and mysterious in the celestial movements. Geometry had thus the boldness to dispose of the future; the evolutions of ages are scrupulously ratifying the decisions of science."[9]

Laplace's publications are divided into two stages: immediate communications about the results he obtained in the 1770s and 1780s and their systematization and amplification in the five volumes of *Celestial Mechanics*. It was characteristic of Laplace that he forced his way through with

[9]François Arago, *Biographies of Distinguished Scientific Men*, translated by W. H. Smyth et al., Longman, Brown, Green, Longmans, and Roberts, London, 1857, p. 201.

unbelievable strength to the solution of specific problems without being diverted to the formation and systematization of an apparatus. Lagrange was the opposite, spending much effort on going from the method to a formalism that was suitable for solving a wide range of problems. Therefore modern texts in theoretical mechanics prominently include the name of Lagrange, while Laplace's name can basically be found in historical essays.

"Whether it was a question of the libration of the moon or a problem in number theory, Lagrange for the most part saw only the mathematical side of the issue; therefore he gave great significance to elegant formulas and generalizations of methods. For Laplace, to the contrary, mathematical analysis was a tool that he adapted to the most varied problems, always fitting a given special method to the essence of the question. Perhaps posterity will say that one was a great geometer, while the second was a great philosopher who tried to apprehend nature, compelling the highest geometry to serve it" (Poisson).

Relations between Laplace and Lagrange were not simple. Laplace's ambition to be the leading mathematician in France continually ran up against the much greater prestige of Lagrange, who had arrived in Paris in 1788. According to many accounts of their contemporaries, it was painful for Laplace to hear Lagrange praised. Lagrange's behavior in the most difficult situations was irreproachable, while many of Laplace's acts were criticized. Maintaining correct relations between Laplace and Lagrange was to a great degree the result of Lagrange's patience. Characteristically, Fourier's eulogy of Laplace said nothing about his moral qualities; at the same time it said, not strangely, much about Lagrange's superlative human qualities.

A hurried style with no attempt to find the internal workings can deceive even the specialist. As a curiosity we can put forth the opinion of Poisson, Laplace's student: "Laplace never saw the truth unless by chance. It hides from this vain man who says only murky things about it. However he tries to turn this murkiness into insight, and by his efforts he gives a noble form to a necessary concern, like a man who is afraid to say too much and to give away a secret that he never had." There are legends about how often Laplace said "it is easy to see." Biot, reading the page proofs of *Celestial Mechanics* and Nathaniel Bowditch (1773–1838), its English translator, told of the hours and days they needed to fill the gaps. According to Biot, Laplace himself often needed to think hard to do this.

The System of the World

While in Melun, Laplace wrote his popular book *The System of the World*, which appeared in 1796. In this book, Laplace lays out his hypothesis

about the origins of the solar system. Laplace, a follower of Newton, "not contriving a hypothesis," proposed his considerations "with the care appropriate for everything that does not present the result of observations or calculations." Laplace describes the development of the solar system as a closed process, not requiring the intervention of outside forces.

There is a well-known story about a conversation between Napoleon and Laplace, who gave him his book:

Napoleon: "Citizen Laplace, Newton talks about God in his book. In your book, which I have already looked over, I did not find God's name even once."

Laplace: "Citizen First Consul, I did not require this hypothesis."

Laplace's words are often taken as a proof of his atheism, although evidently here we are talking about a concrete situation where Laplace's construction did not require external factors either in the hypothesis about the origin of the solar system or in the question of its stability.

According to Laplace's hypothesis everything began with a gaseous cloud revolving around an axis. The cloud, cooling down, first flattened out along the equatorial plane and then spread out in a ring where the planetary orbits are now (because of the balancing of centrifugal and gravitational forces). Various instabilities in the motion of the parts of the ring and their mutual attraction led to the parts coming together into planets. The formation of the system of planetary moons took place analogously, where the example of Saturn shows that sometimes the parts of the ring did not stick together. The basic points of Laplace's model are: all the revolutions take place in the same direction (corresponding to the the direction of the original revolution of the cloud), the trajectories are nearly circular and their planes are near to the equatorial plane of the cloud, and the period of revolution increases with the distance from the center.

The first blows to Laplace's hypothesis were delivered by Herschel, during Laplace's lifetime: moons were found around Uranus that revolved in the "reverse" direction and whose orbital planes were almost perpendicular to the orbital plane of the planet. The number of contradictions began to grow quickly. Scientists tried to correct the hypothesis in many ways and to include more complex constructions in it.

Laplace's hypothesis played an enormous role in the history of cosmology. It was the first hypothesis that rested on a great number of specific facts in mechanics and astronomy (the hypotheses of Georges de Buffon (1707–1788) and Immanuel Kant that came before did not satisfy these conditions, although there are many points of contact between Laplace's hypothesis and that of Kant, which he did not know about). Even at the beginning of the 20th century Poincaré wrote about Laplace's hypothesis

that "for its age, it is not too wrinkled."

"Common Sense Reduced to Calculation"

This is how Laplace graphically described probability theory. It was his second scientific love all through the entire course of his scientific career, starting with his first works in 1774.

The style of Laplace's work in this area was different from what was characteristic of the author of *Celestial Mechanics*. Here there was no one big problem and a lot of time was devoted to trying to understand what had been done earlier, starting with problems about the division of stakes that were among the sources of probability theory.

At the centre of attention were Jacob Bernoulli's Law of Large Numbers, which said that for a large number of trials the frequency of an event approaches its probability in some sense. Starting off with a result of Abraham de Moivre (1667–1754), Laplace obtained an estimate of the probability that the difference between the frequency and probability of an event is large. This is one of the central theorems of probability theory and is called the de Moivre–Laplace theorem. Its proof uses the methods of mathematical analysis, which was a novelty in probability theory.

Laplace appreciated and applied to the sciences the results of the English minister Thomas Bayes (1702–1761) on estimating the probabilities of competing hypotheses when the results of checking them are known.

Laplace's results were summed up in his *Analytical Theory of Probabilities*, which came out during his life in three editions, the first in 1812. A lot of space is devoted to creating the machinery for probability and first of all to the method of generating functions, which now find applications far from probability theory. Laplace was responsible for the "classical definition" of probability, where events are defined as sets of equally likely outcomes: "The theory of probability consists of reducing all events of the same type to some number of equally likely cases, that is cases about whose realization we are equally uninformed, and of defining the number of such cases which are favorable for the event whose probability we seek."

Together with a book for "experts" Laplace wrote a book for the general public. This is his *A Philosophical Essay on Probabilities*, taken from lectures given at the École Normale in 1795 and incorporated in the second edition of *Analytical Theory of Probabilities* in 1814.

Laplace was one of the first authors who gave examples in a book on probability theory not only of games of chance but of real statistics. For example, he presented data showing that the number of letters in France that were not delivered because of a missing address practically did not change from year to year.

Laplace's point of view was that probabilistic considerations are only needed where part of the information is unknown: "We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which could comprehend all the forces by which nature is animated and the respective situation of the beings who compose it—an intelligence sufficiently vast to submit these data to analysis—it would embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence."[10] The hypothetical case discussed in this quotation is now called Laplace's demon.

Laplace's thinking about probability theory was to a significant extent stimulated by his work in astronomy and cosmogony. But he was also concerned about the role of chance in public life. Most often his statements about this contained no concrete calculations. Here is an example: "Let us not offer in the least a useless and often dangerous resistance to the inevitable effects of the progress of knowledge; but let us change only with an extreme circumspection our institutions and the usages to which we have already so long conformed. We should know well by the experience of the past the difficulties which they present; but we are ignorant of the extent of the evils which their change can produce. In this ignorance the theory of probability directs us to avoid all change; especially is it necessary to avoid the sudden changes which in the moral world as well as in the physical world never operate without a great loss of vital force."[11]

There was one question whose formalization Laplace expected—the application of probability theory to legal procedures. It grew out of the viewpoint that a legal decision whose correctness is absolutely beyond doubt is impossible, and that one should only care that the probability that the decision is correct be as large as possible. This goes back to Condorcet and is closely tied to practical legal questions during the Revolution. Laplace's position was more careful, and all the same he thought that one needed to calculate the probability "that the decision of a tribunal which can condemn only by a given majority will be just, that is to say, conform to the true solution of the question proposed. . ."[12] and since "the majority of our opinions being founded on the probability of proofs it is indeed important to submit

[10] Pierre-Simon Laplace, *A Philosophical Essay on Probabilities*, translated by Frederick Wilson Truscott and Frederick Lincoln Emory, Wiley, New York, 1902, p. 4.

[11] Ibid., p. 108.

[12] Ibid., p. 135.

it to calculus."[13] It was proposed to include the political sympathies of the judges, the intricacies of the cases, the intellectual characteristics of the judges, etc. Life showed the errors and social dangers of such calculations.

In 1899, at the time of Alfred Dreyfus' retrial in a military court, "proofs" of his guilt were presented that were based on the probabilistic calculations of a certain Bertillon.[14] Henri Poincaré reached this conclusion about their reliability: "...Even if these computations were accurate, in any case the conclusions would not be valid because the application of probabilistic calculations to the moral sciences are a scandal for mathematics, since Laplace and Condorcet, who knew how to compute, arrived at results lacking in common sense!"

During the 1930s in the Soviet Union, the prosecutors of the Vyshinsky[15] school also talked about the probabilities of crimes, but it seems that they did not get into calculating probabilities.

We have been able to touch on only the most important directions in Laplace's scientific work. Much remains that is outside the limits of our story: his work on capillary action, sound and light, the mathematical results leading to what are now called "Laplace transforms" and "Laplace's equation," etc.

Scholars have recently come to appreciate Laplace's foresight once again. In *The System of the World*, there is a proof that "the gravitational attraction of a heavenly body can be so great that no light can come from it." This would happen if the body had the density of the earth but diameter equal to 250 times the diameter of the sun. In other words, the first cosmic velocity[16] in the gravitational field of this body would exceed the speed of light. Thus Laplace was the first to pay attention to the possible existence of black holes.

To a great extent, Laplace's life reflects the complexity of the times in which he lived. However, throughout his life he carried a belief in science and did not interrupt his work under any circumstances. It is difficult to overestimate Laplace's role in the history of science.

"...Laplace was born to make everything deeper, to cross all boundaries, in order to to solve what seemed unsolvable. He completed the science of the heavens, if that science can be completed" (Fourier).

[13]Ibid., p. 109. Here "probability of proofs" means the probability that the testimony of witnesses is correct.—*Transl.*

[14]The Dreyfus affair was a famous anti-Semitic scandal in French society of the 1890s. Alphonse Bertillon was a police official and anthropologist who analyzed a handwriting sample in the case.—*Transl.*

[15]Andrei Vyshinsky was the chief prosecutor in the Stalin show trials and later became the Soviet Union's Foreign Minister and Ambassador to the United Nations.—*Transl.*

[16]The velocity required to orbit the body.—*Transl.*

Prince of Mathematicians

Nihil actum reputans si quid superesset agendum. (Judging that nothing was done if something was left undone.) Gauss[1]

In 1854, the health of Privy Councillor Gauss, as his colleagues at the University of Göttingen called him, worsened decisively. There was no question of continuing the daily walks from the observatory to the literary museum, a habit of over twenty years. They managed to convince the professor, who was nearing eighty, to go to the doctor! He improved during the summer and even attended the opening of the Hannover–Göttingen railway. In January 1855, Gauss agreed to pose for a medallion by the artist Heinrich Hesemann. After the scientist's death in February 1855, a medal was prepared from the medallion, by order of the Hannover court. Beneath a bas-relief of Gauss, these words were written: *Mathematicorum princeps* (Prince of Mathematicians). The story of every real prince should begin with his childhood, embroidered with legends. Gauss is no exception.

1 Gauss' Debut

"The obstinacy with which Gauss followed a path once chosen, the youthful impetuosity with which he regularly and recklessly took the steepest way towards his goal—these hard tests strengthened his powers and made him capable of striding recklessly over all obstacles, even when they had already been removed by earlier investigations. And to this praise of independent activity I would like to add another: the praise of youth. What I want to

[1] As translated in Felix Klein, *Development of Mathematics in the* 19*th Century*, translated by Michael Ackerman, Math Sci Press, Brookline, MA, 1979, p. 8.—*Transl.*

The Young Gauss (1803).

say perhaps means only that the laws which underlie the development of mathematical genius are the same as those for any other creative gift: in the early years, when a person has just reached full physical growth, great revelations may hurry in upon him; it is then that he creates what he has to bring into the world as his own new value, even though his ability to express them may not yet be equal to his abundant flow of ideas" (Felix Klein).[2]

Braunschweig, 1777–1795

Gauss did not inherit his title, although his father Gebhard Dietrich was no stranger to mathematics. A jack-of-all-trades, primarily a fountain builder but also a gardener like his father before him, Gebhard Dietrich was known for his talent as an accountant. Merchants made use of his services during fairs in Braunschweig and even Leipzig, and he was also regularly employed by the largest burial fund in Braunschweig (a position he bequeathed to his son by his late wife, Johann Georg, a retired soldier).

[2]Ibid., pp. 31–32.

Carl Friedrich was born on April 30, 1777, in house number 1550 on the Wendengraben canal in Braunschweig. According to his biographers, he inherited good health from his father's side and a brilliant mind from his mother's. He was closest to his uncle Johann Friederich Benze, a skillful weaver in whom, in his nephew's words, "an innate genius perished." Gauss said of himself that he "could count before he spoke." The earliest mathematical legend about him claims that at the age of three he followed his father's calculations with a bricklayer, unexpectedly corrected him, and turned out to be right.

At seven, Carl Friedrich entered the Catharineum school. Since they only learned to count in the third grade, little Gauss did not attract any special attention for the first two years. Students usually entered the third grade at age ten and studied there until confirmation at fifteen. The teacher, whose name was J. G. Büttner, had to work simultaneously with children of different ages and with different preparation. Thus he usually gave some of his students long computations to do so that he could talk to the others. Once a group of students, including Gauss, was asked to sum the integers from 1 to 100. (Different sources name different numbers.) As soon as they were done, the students were to put their slates on the teacher's table, and the order of the slates counted towards their grades. The ten-year-old Gauss put down his slate when Büttner had hardly finished dictating the problem. To everyone's surprise, his answer was even correct. The secret was simple: While the problem was being dictated, Gauss rediscovered the formula for the sum of an arithmetic progression! The child-genius' fame spread throughout little Braunschweig.

In the school where Gauss studied, the teacher's assistant, whose chief duty was to cut quills for the pupils' pens, was a man named Martin Bartels. Bartels was interested in mathematics and owned a few mathematics books. He and Gauss began to study together; they learned about Newton's binomial formula, infinite series, etc.

What a small world it turned out to be! After sometime, Bartels received a chair in pure mathematics at the University of Kazan, and was the teacher of Nikolai Lobachevsky (1792–1856).

In 1788, Gauss entered the gymnasium. Mathematics was not studied there, but rather classical languages were. Gauss happily studied languages and was so successful that he did not even know what he wanted to become—a mathematician or a philologist.

Word of Gauss reached the court, and in 1791, he was presented to Carl Wilhelm Ferdinand, the duke of Braunschweig. The boy lived at court and amused the courtiers with his feats of calculation. Thanks to the duke's patronage, Gauss was admitted to the University of Göttingen in

October 1795. At first he attended lectures on philology and almost none on mathematics. But he continued to study mathematics anyway.

Here is a comment by Felix Klein, the noted mathematician who studied Gauss' scientific work at length: "A natural interest, I might even say a certain childlike curiosity, first led the boy to mathematical questions, independently of any outside influence. Indeed, it was simply the art of calculating with numbers that first attracted him. He calculated continually, with overpowering industry and untiring perseverance. By this incessant exercise in manipulating numbers (for example, calculating decimals to an unbelievable number of places) he acquired not only the astounding virtuosity in computational technique that marked him throughout his life, but also an immense memory stock of definite numerical values, and thereby an appreciation and overview of the realm of numbers such as probably no one, before or after him, has possessed. Aside from arithmetic he was occupied with numerical operations on infinite series. From his activity with numbers, and thus in an inductive, 'experimental' way, he arrived quite early at a knowledge of their general relations and laws.... It was not so rare in the eighteenth century—for example, with Euler—but stands in sharp contrast to the normal practice of today's mathematicians.... All these early intellectual games, devised solely for his own pleasure, were first steps towards a great goal that became conscious only later. It is part of the anticipatory wisdom of genius to place the pick-axe precisely on the rock vein where the gold mine lies concealed, and to do this even in the half-playful first testings of its powers, unconscious of its deeper meaning. We now come to the year 1795, of which we have more detailed evidence.... Then, still before his Göttingen period, a passionate interest in the integers seized him, even more tenaciously than before, as is vividly evidenced by the preface to the *Disquisitiones Arithmeticae*. Unacquainted with the literature, he had to create everything for himself. Here again it was the untiring calculator who blazed the way into the unknown. Gauss set out huge tables: of prime numbers, of quadratic residues and nonresidues, and of the fractions $\frac{1}{p}$ for $p = 1$ to 1000 with their decimal expansions carried out to a complete period, and therefore sometimes to several hundred places! With this last table Gauss tried to determine the dependence of the period on the denominator p. What researcher of today would be likely to enter upon this strange path in search of a new theorem? But for Gauss it was precisely this path, followed with such unheard of energy—he himself maintained that he differed from other men only in his diligence—that led to his goal.... In the autumn of 1795 he moved to Göttingen, where he must have devoured the works of Euler and Lagrange, presented to him for the first time."[3]

[3]Ibid., pp. 29–30.

A Discovery after Two Thousand Years

On June 1, 1796, the following notice appeared in the newspaper *Jenenser Intelligenzblatt*: "Every beginner in geometry knows that it is possible to construct different regular polygons [with compass and straightedge], for example triangles, pentagons, 15-gons, and those regular polygons that result from doubling the number of sides of these figures. One had already come this far in Euclid's time, and it seems that since then one has generally believed that the field for elementary geometry ended at that point, and in any case I do not know of any successful attempt to extend the boundaries beyond that line.

Therefore it seems to me that this discovery possesses special interest, *that besides these regular polygons, a number of others are geometrically constructible, for example the* 17*-gon*."[4]

Beneath the notice was the signature, *C. F. Gauss, Braunschweig*, Mathematics Student at Göttingen.

This is the first communication of a discovery by Gauss. Before discussing it in detail, let us refresh our memories about what "every beginner in geometry knows."

Constructions with Straightedge and Compass

Suppose we are given an interval of length 1. Using a straightedge and compass, we can construct new intervals whose lengths are obtained from those we already have by addition, subtraction, multiplication, division, and extracting square roots.

By successively carrying out these operations, we can construct, using straightedge and compass, any interval whose length can be expressed in terms of 1 *by a finite number of these operations. We call these numbers quadratic irrationals. One can prove that these are the only intervals that can be constructed with straightedge and compass*.

It is easy to see that the problem of constructing a regular n-gon is equivalent to the problem of dividing a circle of radius 1 into n equal parts. The chords of the arcs into which we divide the circle are the sides of a regular n-gon, and their lengths are all equal to $2\sin(\frac{\pi}{n})$. Thus *for each* n *for which* $\sin(\frac{\pi}{n})$ *is a quadratic irrational, we can construct a regular* n*-gon with straightedge and compass. This condition is satisfied, for example, by* $n = 3, 4, 5, 6,$ *and* 10. This is well known for $n = 3, 4,$ and 6.

[4]This appears in English in Tord Hall, *Carl Friedrich Gauss*, translated by Albert Froderberg, M.I.T. Press, Cambridge, MA, 1970, p. 24.—*Transl.*

We will show that $\sin(\frac{\pi}{10})$ is a quadratic irrational. Consider an isosceles triangle ABC whose vertex angle B equals $\frac{\pi}{5} = 36^\circ$ and whose base AB has length 1. Let AD bisect angle A. Then $x = AC = AD = BD = 2\sin(\frac{\pi}{10})$. We have

$$\frac{BD}{DC} = \frac{AB}{AC}; \qquad \frac{x}{1-x} = \frac{1}{x}, \quad x = \frac{\sqrt{5}-1}{2}.$$

This is a quadratic irrational number, so we can construct the sides of a regular 10-gon.

Furthermore, if we can divide a circle into p_1p_2 equal parts then we can of course divide it into p_1 equal parts. (In particular, we can construct a regular pentagon.) The converse is in general not true, but we note two special cases when it does hold:

(1) *If we can divide a circle into p equal parts, then we can divide it into 2^kp equal parts, for any k*. This is true because we can bisect any angle with straightedge and compass.

(2) If we can divide a circle into p_1 equal parts and into p_2 equal parts, and p_1 and p_2 are relatively prime (e.g., if p_1 and p_2 are distinct primes), then we can divide the circle into p_1p_2 equal parts. This follows from the fact that the greatest common divisor of the angles $\frac{2\pi}{p_1}$ and $\frac{2\pi}{p_2}$ is $\frac{2\pi}{p_1p_2}$, and the greatest common divisor of two commensurable angles can be found by straightedge and compass. In particular, $\frac{2\pi}{15} = \frac{1}{2}(\frac{2\pi}{3} - \frac{2\pi}{5})$, which implies the possibility of constructing a regular 15-gon.

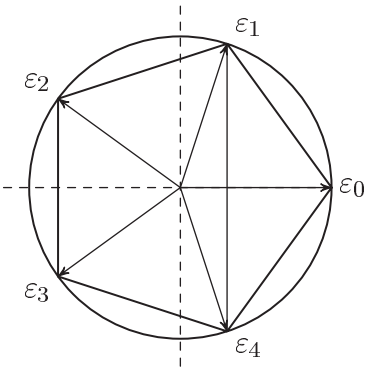
A Few Words about Complex Numbers

We need to know just a bit about complex numbers, namely, their basic operations and geometric interpretation. Recall that a complex number $z = a + ib$ corresponds to a point with coordinates (a, b) and to the vector from the origin $(0, 0)$ to this point. The length of the vector, $r = \sqrt{a^2 + b^2}$ is called the *modulus* of the given number z, and is denoted by $|z|$. We can write z in trigonometric form: $z = a + ib = r(\cos\varphi + i\sin\varphi)$. The angle φ is called the *argument* of z.

Addition of complex numbers corresponds to vector addition. In multiplying two complex numbers, we multiply their moduli and add their arguments. This implies that the equation $z^n = 1$ has exactly n roots, which are usually denoted by

$$\varepsilon_k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}, \quad k = 0, 1, \ldots, n-1. \tag{1}$$

As vectors, the ε_k end at the vertices of a regular n-gon. If we prove that the ε_k are quadratic irrationals (i.e., that their real and imaginary parts a



and b are quadratic irrationals), then this will prove that the regular n-gon can be constructed with straightedge and compass.

Regular n-Gons and Roots of Unity

We rewrite $z^n = 1$ as

$$z^n - 1 = (z-1)(z^{n-1} + z^{n-2} + \cdots + z + 1) = 0.$$

We obtain two equations, $z = 1$ and

$$z^{n-1} + z^{n-2} + \cdots + z + 1 = 0. \tag{2}$$

The roots of equation (2) are the numbers ε_k, for $1 \le k \le n-1$, and we will work with this equation below.

For $n = 3$ we obtain the equation $z^2 + z + 1 = 0$, with roots

$$\varepsilon_1 = -\frac{1}{2} + i\frac{\sqrt{3}}{2}, \qquad \varepsilon_2 = -\frac{1}{2} - i\frac{\sqrt{3}}{2}.$$

For $n = 5$ the situation is more complicated, since we have the fourth-degree equation

$$z^4 + z^3 + z^2 + z + 1 = 0, \tag{3}$$

with four roots ε_1, ε_2, ε_3, and ε_4. Although we have Ferrari's formula for solving the general fourth-degree equation, it is impossible to use in practice. In our case there is a special form of equation (3) that is useful. In order to solve (3), we first divide by z^2, obtaining

$$z^2 + \frac{1}{z^2} + z + \frac{1}{z} + 1 = 0,$$

or

$$\left(z+\frac{1}{z}\right)^2+\left(z+\frac{1}{z}\right)-1=0.$$

We make the substitution $w=z+\frac{1}{z}$:

$$w^2+w-1=0, \tag{4}$$

with roots

$$w_{1,2}=\frac{-1\pm\sqrt{5}}{2}.$$

We can also find ε_k from the equations

$$z+\frac{1}{z}=w_1, \qquad z+\frac{1}{z}=w_2, \tag{5}$$

but this is unnecessary. To find them it is enough to know that twice the real part of ε_1 equals

$$2\cos\left(\frac{2\pi}{5}\right)=\varepsilon_1+\varepsilon_4=\varepsilon_1+\frac{1}{\varepsilon_1}=w_1=\frac{-1+\sqrt{5}}{2}.$$

Since w_1 is a quadratic irrational, so are ε_1 and ε_4. The argument for ε_2 and ε_3 is similar.

Thus, for $n=5$ the solution to our problem can be reduced to the successive solution of two quadratic equations: first solve (4), whose roots are the sums $\varepsilon_1+\varepsilon_4$ and $\varepsilon_2+\varepsilon_3$ of the symmetric roots of (3), and then find these roots of (3) from (5).

This is the very method Gauss used to construct the regular 17-gon, singling out groups of roots whose sums are found successively from quadratic equations. But how can we look for these "good" groups? Gauss found a surprising way to answer this question.

Constructing the Regular 17-Gon

"And on March 30, 1796 he underwent his conversion on the road to Damascus.... For a long time Gauss had been busy with grouping the roots of unity $x^n=1$ on the basis of his theory of 'primitive roots.' Then suddenly one morning, still in bed, he saw clearly that the construction of the 17-gon follows from his theory. As already mentioned, this discovery marked a turning point in Gauss' life. He decided to devote himself entirely to mathematics, not philology" (Felix Klein).[5]

[5] *Development*, p. 30.

Let us go into a bit more detail about the path that Gauss took. One of the young Gauss' mathematical games was to divide 1 by various primes p and write down the successive decimal digits, impatiently waiting for them to begin to repeat. Sometimes he had a long wait. For $p = 97$ the repetition began with the 97th digit, and for $p = 337$ with the 337th. Gauss was not confused by long lines of computations, but rather entered the mysterious world of numbers with their help. He was not too lazy to consider all $p < 1000$ (see the earlier quote from Klein).

It is known that Gauss did not immediately try to prove the periodicity of these fractions in general ($p \neq 2, 5$), but the proof probably gave him no difficulty. Indeed, it is enough just to note that we must keep track of the remainder rather than the quotient. The digits of the quotient begin to repeat when at the preceding step the remainder equals 1 (why?). This means that we must find some k such that $10^k - 1$ is divisible by p. Since there are only a finite number of possible remainders (they lie between 1 and $p - 1$), for some $k_1 > k_2$ the numbers 10^{k_1} and 10^{k_2} leave the same remainder after dividing by p. But then $10^{k_1 - k_2} - 1$ is divisible by p (why?).

It is a little harder to show that we can always take $p - 1$ as k, i.e., that $10^{p-1} - 1$ is always divisible by p when $p \neq 2, 5$. This is a special case of the result known as Fermat's Little Theorem. When Fermat (1601–1655) discovered it, he wrote that he "was illuminated by a clear light." Now the young Gauss had rediscovered it. He would always value this result highly: "This theorem is worthy of attention both because of its elegance and its great usefulness."[6]

Gauss was interested in the smallest value of k for which $10^k - 1$ is divisible by p. Such a value of k always divides $p - 1$, and sometimes it equals $p - 1$, e.g., for $p = 7, 17, 19, 23, 29, 97$, and 337. It remains unknown whether the number of such primes p is finite or infinite.

Gauss replaced 10 by any number a for which p does not divide a, and asked when $a^k - 1$ is not divisible by p for $k < p - 1$. In this case we say a is a *primitive root* mod p. This is equivalent to saying that all the nonzero remainders $1, 2, \ldots, p - 1$ occur when we divide $1, a, a^2, \ldots, a^{p-2}$ by p (why?).

Gauss did not know at the time that Euler (1707–1783) had been interested in primitive roots. Euler conjectured (but could not prove) that *each prime number has at least one primitive root*. Legendre (1752–1833) gave the first proof of Euler's conjecture, and Gauss gave a very elegant proof. But this was later, and for the time being Gauss worked with concrete examples. He knew, for instance, that $a = 3$ is a primitive root mod $p = 17$. In

[6] *Disquisitiones Arithmeticae*, translated by Arthur A. Clarke, Yale University Press, New Haven, CT, 1966, p. 32.

the first row of the table below we show the values of k, and below them the remainders of 3^k after dividing by 17. Note that the second row contains all the remainders from 1 to 16, which implies that 3 is primitive mod 17.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	3	9	10	13	5	15	11	16	14	8	7	4	12	2	6

These calculations lay at the heart of grouping the roots of

$$z^{16} + z^{15} + z^{14} + \cdots + z + 1 = 0 \tag{6}$$

in order to reduce its solution to a chain of quadratic equations. Gauss' idea was that another way to number the roots was needed. We will give a root ε_k the number l and denote it by $\varepsilon_{[l]}$ if 3^l divided by 17 leaves the remainder k. We can use the table to go from one numbering to the other, finding k in the second row and the corresponding l above it in the first, but it is more convenient to use Figure 1, where the old numbers appear outside the circle and the new ones inside. This is the numbering Gauss used to divide the roots of equation (6) into groups and reduce its solution to a chain of quadratic equations.

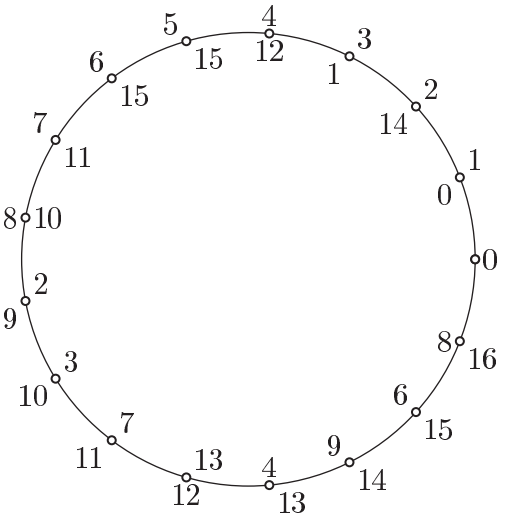


Figure 1.

At the first stage, let $\sigma_{2,0}$ and $\sigma_{2,1}$ be the sums of the roots $\varepsilon_{[l]}$ for even and odd l, respectively. Each is the sum of the eight roots for which dividing l by 2 leaves the remainder 0 and 1, respectively. These sums turn out to

be the roots of a quadratic equation with integer coefficients. Next, let $\sigma_{4,0}$, $\sigma_{4,1}$, $\sigma_{4,2}$, and $\sigma_{4,3}$ be the sums of the four roots $\varepsilon_{[l]}$ for which dividing l by 4 leaves the remainder 0, 1, 2, and 3, respectively. We will show that these quantities are the roots of quadratic equations whose coefficients can be expressed arithmetically in terms of $\sigma_{2,0}$ and $\sigma_{2,1}$. Finally, form the sums $\sigma_{8,i}$ of the two roots $\varepsilon_{[l]}$ for which dividing l by 8 leaves the remainder i. We write quadratic equations for them whose coefficients are expressed in a simple way in terms of the $\sigma_{4,j}$. We have $\sigma_{8,0} = 2\cos(\frac{2\pi}{17})$, and the quadratic irrationality of $\sigma_{8,0}$ will imply that it is possible to construct the regular 17-gon with straightedge and compass. It is instructive to write down the decomposition of the roots in the old numbering. You should agree that it is impossible to guess the decomposition in this form! We will now carry out the scheme we have just described.

Detailed Calculations

We will now prove that the seventeenth-order roots of unity are quadratic irrationals. Note that $\varepsilon_k\varepsilon_l = \varepsilon_{k+l}$ (if $k + l \geq 17$, then replace $k + l$ by its remainder after division by 17) and $\varepsilon_k = (\varepsilon_1)^k$. We first remark that

$$\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{16} = \varepsilon_{[0]} + \varepsilon_{[1]} + \cdots + \varepsilon_{[15]} = -1.$$

(For example, consider this as the sum of a geometric progression.)

Let $\sigma_{m,r}$ denote the sum of $\varepsilon_{[k]}$ for those k leaving remainder r after division by m. We obtain

$$\sigma_{2,0} = \varepsilon_{[0]} + \varepsilon_{[2]} + \varepsilon_{[4]} + \cdots + \varepsilon_{[14]},$$
$$\sigma_{2,1} = \varepsilon_{[1]} + \varepsilon_{[3]} + \varepsilon_{[5]} + \cdots + \varepsilon_{[15]}.$$

Clearly,

$$\sigma_{2,0} + \sigma_{2,1} = \varepsilon_{[0]} + \varepsilon_{[1]} + \cdots + \varepsilon_{[15]} = -1.$$

One can show that[7]

$$\sigma_{2,0} \cdot \sigma_{2,1} = 4(\varepsilon_{[0]} + \varepsilon_{[1]} + \cdots + \varepsilon_{[15]}) = -4.$$

Now, by Vieta's theorem, we can construct a quadratic equation with roots $\sigma_{2,0}$ and $\sigma_{2,1}$:

$$x^2 + x - 4 = 0, \quad x_{1,2} = \frac{-1 \pm \sqrt{17}}{2}.$$

[7] We can see this by carrying out the multiplication directly, using $\varepsilon_k\varepsilon_l = \varepsilon_{k+l}$ and Figure 1. However, below we will learn a method for avoiding these tedious calculations.

In order to distinguish $\sigma_{2,0}$ from $\sigma_{2,1}$, we again use Figure 1. In each sum, the roots appear together with their conjugates. It is clear that $\sigma_{2,0} > \sigma_{2,1}$, since $\sigma_{2,0}$ is twice the sum of the real parts of ε_1, ε_2, ε_4, and ε_8, while for $\sigma_{2,1}$ we repeat this with ε_3, ε_5, ε_6, and ε_7. Thus

$$\sigma_{2,0} = \frac{\sqrt{17}-1}{2}, \qquad \sigma_{2,1} = \frac{-\sqrt{17}-1}{2}.$$

Consider the roots taken four at a time:

$$\begin{aligned}
\sigma_{4,0} &= \varepsilon_{[0]} + \varepsilon_{[4]} + \varepsilon_{[8]} + \varepsilon_{[12]},\\
\sigma_{4,1} &= \varepsilon_{[1]} + \varepsilon_{[5]} + \varepsilon_{[9]} + \varepsilon_{[13]},\\
\sigma_{4,2} &= \varepsilon_{[2]} + \varepsilon_{[6]} + \varepsilon_{[10]} + \varepsilon_{[14]},\\
\sigma_{4,3} &= \varepsilon_{[3]} + \varepsilon_{[7]} + \varepsilon_{[11]} + \varepsilon_{[15]}.
\end{aligned}$$

We have $\sigma_{4,0} + \sigma_{4,2} = \sigma_{2,0}$ and $\sigma_{4,1} + \sigma_{4,3} = \sigma_{2,1}$. One can also show that $\sigma_{4,0} \cdot \sigma_{4,2} = \sigma_{2,0} + \sigma_{2,1} = -1$, which means that $\sigma_{4,0}$ and $\sigma_{4,2}$ are roots of $x^2 - \sigma_{2,0}x - 1 = 0$. Solving this equation and noting that $\sigma_{4,0} > \sigma_{4,2}$ (see Figure 1 again), after some simple transformations we obtain

$$\begin{aligned}
\sigma_{4,0} &= \frac{1}{4}\left(\sqrt{17} - 1 + \sqrt{34 - 2\sqrt{17}}\right),\\
\sigma_{4,2} &= \frac{1}{4}\left(\sqrt{17} - 1 - \sqrt{34 - 2\sqrt{17}}\right).
\end{aligned}$$

Analogously,

$$\begin{aligned}
\sigma_{4,1} &= \frac{1}{4}\left(-\sqrt{17} - 1 + \sqrt{34 + 2\sqrt{17}}\right),\\
\sigma_{4,3} &= \frac{1}{4}\left(-\sqrt{17} - 1 - \sqrt{34 + 2\sqrt{17}}\right).
\end{aligned}$$

We now go to the next step. Set

$$\begin{aligned}
\sigma_{8,0} &= \varepsilon_{[0]} + \varepsilon_{[8]} = \varepsilon_1 + \varepsilon_{16},\\
\sigma_{8,4} &= \varepsilon_{[4]} + \varepsilon_{[12]} = \varepsilon_4 + \varepsilon_{13}.
\end{aligned}$$

We could have considered the six other expressions of this kind but we do not need to, since it suffices to prove that $\sigma_{8,0} = 2\cos(\frac{2\pi}{17})$ is a quadratic irrational, thus permitting the construction of the regular 17-gon. We have $\sigma_{8,0} + \sigma_{8,4} = \sigma_{4,0}$ and $\sigma_{8,0} \cdot \sigma_{8,4} = \sigma_{4,1}$. Figure 1 shows that $\sigma_{8,0} > \sigma_{8,4}$, so

$\sigma_{8,0}$ is the largest root of $x^2 - \sigma_{4,0}x + \sigma_{4,1} = 0$, i.e.,

$$\begin{aligned}\sigma_{8,0} = 2\cos\left(\frac{2\pi}{17}\right) &= \frac{1}{2}\left(\sigma_{4,0} + \sqrt{\sigma_{4,0}^2 - 4\sigma_{4,1}}\right) \\ &= \frac{1}{8}\left(\sqrt{17} - 1 + \sqrt{34 - 2\sqrt{17}}\right) \\ &\quad + \frac{1}{4}\sqrt{17 + 3\sqrt{17} - \sqrt{170 + 38\sqrt{17}}}.\end{aligned}$$

We have somewhat transformed the expression we obtained directly for $\sqrt{\sigma_{4,0}^2 - 4\sigma_{4,1}}$, but we will not tire the reader by reproducing these simple calculations.

Using the formula this gives for $\cos(\frac{2\pi}{17})$, we can complete the construction of the regular 17-gon by using elementary rules for constructing expressions that are quadratic irrationals. This is a rather awkward procedure, and these days rather compact construction methods are known. We will present one in an appendix, without proof. In one respect the formula for $\cos(\frac{2\pi}{17})$ leaves no doubt, but it would have been impossible to arrive at it within the confines of the traditional geometric ideas of Euclid's time. Gauss' solution belonged to a different era in mathematics. Note that the most interesting claim is the possibility, in principle, of constructing the regular 17-gon; the procedure itself is not important. To prove the construction is possible it is enough to see that at each step we have quadratic equations whose coefficients are quadratic irrationals, without writing down explicit expressions for them (this becomes particularly important when the number of sides is larger).

In the solution we have described for equation (6), we have left completely unexplained why the numbering $\varepsilon_{[l]}$ for partitioning the roots turned out to be successful. How could we have guessed it would be the basis for the solution? We will now essentially repeat the solution, exposing the key idea—symmetry in the set of roots.

Symmetry in the Set of Roots of Equation (6)

First of all, the problem of the roots of unity is closely connected to the arithmetic of remainders after division by n (arithmetic modulo n). Indeed, if $\varepsilon^n = 1$, then ε^k is also an nth root of unity and its value depends only on the remainder left after dividing k by n. Set $\varepsilon = \varepsilon_1$ (see formula (1)); then $\varepsilon_k = \varepsilon^k$, so $\varepsilon_k \cdot \varepsilon_l = \varepsilon_{k+l}$, where the sum is taken modulo n (the remainder after dividing by n). In particular, $\varepsilon_k \cdot \varepsilon_{n-k} = \varepsilon_0 = 1$.

Problem 1. If p is prime and δ is any complex pth root of unity, then the powers δ^k, $k = 0, 1, \ldots, p-1$, contain all the pth roots of unity.

Remark. We must prove here that for each m with $0 < m < p$, all the numbers $0, 1, \ldots, p-1$ are contained among the remainders after dividing km by p, for $k = 0, 1, \ldots, p-1$.

We denote exponentiation to the kth power by T_k: $T_k\varepsilon_l = (\varepsilon_l)^k = \varepsilon_{lk}$.

Problem 2. Prove that if $n = p$ is prime, then each transformation T_k, $k = 1, 2, \ldots, p-1$, is a one-to-one transformation of the set of roots onto itself, i.e., the set $\{T_k\varepsilon_0, T_k\varepsilon_1, \ldots, T_k\varepsilon_{p-1}\}$ coincides with the set of roots $\{\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_{p-1}\}$.

Problem 1 shows that for each l with $1 \le l \le p-1$, $\{T_0\varepsilon_l, T_1\varepsilon_l, \ldots, T_{p-1}\varepsilon_l\}$ is the set of all roots. Problems 1 and 2 imply the following conclusion: *construct a table whose entry in the kth row and lth column is $T_k\varepsilon_l$*, for $1 \le k, l \le p-1$. *Then each row and each column contain all the roots $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{p-1}$ in some order, without repetition*. Note that $T_{p-1}\varepsilon_l = \varepsilon_{-l} = (\varepsilon_l)^{-1}$. Those who know the definition of a group may verify that the transformations T_k form a group with respect to the multiplication $T_k \cdot T_l = T_{kl}$.

Now consider the case $p = 17$. We call a set M of roots invariant with respect to T_k if $T_k\varepsilon_l \in M$ for each $\varepsilon_l \in M$. The only set that is invariant with respect to all the T_k is the set $\{\varepsilon_1, \ldots, \varepsilon_{16}\}$ of all the roots.

The underlying conjecture is that *a group of roots is "better" when more transformations leave it invariant*.

We introduce another numbering $T_{[l]}$ for T_k, as we did for ε_k: $T_{[l]} = T_k$ when $k = 3^l$. In the new notation,

$$T_{[k]}\varepsilon_{[l]} = \varepsilon_{[k+l]},$$
$$T_{[m]}(T_{[k]}\varepsilon_{[l]}) = T_{[m+k]}\varepsilon_{[l]},$$

where the sum in brackets is taken modulo 16. The reader will of course discover an analogy with logarithms, which is not surprising since $\varepsilon_{[l]} = \varepsilon_{3^l}$.

Problem 3. Prove that if some set of roots is invariant with respect to some $T_{[k]}$, where k is odd, then this set is invariant with respect to all $T_{[m]}$, i.e., if it is nonempty then it is the set of all roots.

Remark. It suffices to show that if k is odd, then there is some m such that 16 divides km with remainder 1.

On the other hand, there are two groups of roots that are invariant with respect to $T_{[k]}$ for all even k: the roots $\varepsilon_{[l]}$ for even l and for odd l. We denote their sums by $\sigma_{2,0}$ and $\sigma_{2,1}$.

Clearly, $\sigma_{2,0}+\sigma_{2,1}=-1$. Consider $\sigma_{2,0}\cdot\sigma_{2,1}$. This is the sum of pairwise products $\varepsilon_{[k]}\cdot\varepsilon_{[l]}$, where k is even and l is odd, each equal to some root $\varepsilon_{[m]}$—64 terms in all. We will show that each root $\varepsilon_{[0]}, \varepsilon_{[1]}, \ldots, \varepsilon_{[15]}$ occurs among them exactly four times and that as a result, $\sigma_{2,0}\cdot\sigma_{2,1}=-4$. We will use the fact that $T_{[k]}$ preserves each group of roots when k is even and transforms one into the other when k is odd. Each term of $\sigma_{2,0}\cdot\sigma_{2,1}$ can be uniquely represented in the form $\varepsilon_{[m]}\varepsilon_{[m+r]}$, where $0\le m\le 15$ and $r=1,3,5,7$ (proof!). Group together terms with the same value of r. We obtain a sum of the form

$$\begin{aligned}&\varepsilon_{[0]}\varepsilon_{[r]}+\varepsilon_{[1]}\varepsilon_{[r+1]}+\cdots+\varepsilon_{[15]}\varepsilon_{[r+15]}\\&\quad=T_{[0]}(\varepsilon_{[0]}\varepsilon_{[r]})+T_{[1]}(\varepsilon_{[0]}\varepsilon_{[r]})+\cdots+T_{[15]}(\varepsilon_{[0]}\varepsilon_{[r]})\\&\quad=T_{[0]}\varepsilon_{[r]}+T_{[1]}\varepsilon_{[r]}+\cdots+T_{[15]}\varepsilon_{[r]}\\&\quad=\varepsilon_{[0]}+\varepsilon_{[1]}+\cdots+\varepsilon_{[15]}=-1.\end{aligned}$$

We have used the fact that

$$T_{[m]}\varepsilon_{[k]}\cdot T_{[m]}\varepsilon_{[l]}=T_{[m]}(\varepsilon_{[k]}\varepsilon_{[l]}),$$

and the properties of $T_{[m]}$ we have already mentioned.

The values of $\sigma_{2,0}$ and $\sigma_{2,1}$ were found above.

We now go to the next step. We want to introduce new, smaller, groups of roots that are invariant with respect to some $T_{[k]}$. By analogy with Problem 3, one can show that k must be divisible by 4. Therefore there are four groups of roots, invariant with respect to all $T_{[4l]}$ and smaller than the ones already considered. Let the sums of the roots in these groups be $\sigma_{4,0}$, $\sigma_{4,1}$, $\sigma_{4,2}$, $\sigma_{4,3}$. We have already noted that $\sigma_{4,0}+\sigma_{4,2}=\sigma_{2,0}$ and $\sigma_{4,1}+\sigma_{4,3}=\sigma_{2,1}$.

We will calculate the product $\sigma_{4,0}\cdot\sigma_{4,2}$, which is the sum of sixteen terms of the form $\varepsilon_{[4k]}\varepsilon_{[4l+2]}$. Each term can be written uniquely in the form $\varepsilon_{[2m]}\cdot\varepsilon_{[2m+2r]}$, where $r=1,3$ and $m=0,1,2,3,4,5,6,7$. We group together terms with the same r and note that $\varepsilon_{[0]}\varepsilon_{[2]}=\varepsilon_1\varepsilon_9=\varepsilon_{10}=\varepsilon_{[3]}$ and $\varepsilon_{[0]}\varepsilon_{[6]}=\varepsilon_1\varepsilon_{15}=\varepsilon_{16}=\varepsilon_{[8]}$. For $r=1$, we obtain the sum

$$T_{[0]}\varepsilon_{[3]}+T_{[2]}\varepsilon_{[3]}+\cdots+T_{[14]}\varepsilon_{[3]}=\sigma_{2,1}.$$

For $r=3$ we have the sum $\sum_k T_{[2k]}\varepsilon_{[8]}=\sigma_{2,0}$, i.e., $\sigma_{4,0}\cdot\sigma_{4,2}=\sigma_{2,0}+\sigma_{2,1}=-1$. By solving the quadratic equations, we found $\sigma_{4,0}$ and $\sigma_{4,2}$.

In the last step we consider groups of roots invariant with respect to $T_{[8]}$; there are eight. In particular, $\sigma_{8,0}+\sigma_{8,4}=\sigma_{4,0}$. We will compute $\sigma_{8,0}\cdot\sigma_{8,4}$. Taking into account $\varepsilon_{[0]}\cdot\varepsilon_{[4]}=\varepsilon_1\varepsilon_{13}=\varepsilon_{14}=\varepsilon_{[9]}$, we obtain $\sigma_{8,0}\cdot\sigma_{8,4}=T_{[0]}\varepsilon_{[9]}+T_{[4]}\varepsilon_{[9]}+T_{[8]}\varepsilon_{[9]}+T_{[12]}\varepsilon_{[9]}=\sigma_{4,1}$. This allowed us to find $\sigma_{8,0}=2\cos(\frac{2\pi}{17})$ and thus to complete the solution.

We have seen that Gauss' argument is completely based on transformations that rearrange the roots. Lagrange (1736–1813) was the first to consider the role of such transformations in questions about the solvability of equations. Gauss was probably not familiar with Lagrange's work at the time. Later, Évariste Galois (1811–1832) placed the study of these transformations at the foundation of the remarkable theory that now bears his name. Essentially, Gauss constructed Galois theory full-blown for the equation of the division of the circle.

Possible Generalizations and Fermat Primes

If we do not try to obtain an explicit expression for the roots but only try to prove they are quadratic irrationals, then we can almost entirely omit the calculations and consider only the idea of invariance. Namely, $\sigma_{2,0} \cdot \sigma_{2,1}$ is the sum of certain roots $\varepsilon_{[l]}$, and since this sum is transformed into itself by the action of each transformation $T_{[k]}$, all the roots it contains occur the same number of times. Thus $\sigma_{2,0} \cdot \sigma_{2,1}$ is an integer. Analogously, $\sigma_{4,0} \cdot \sigma_{4,2}$ does not change under all transformations of the form $T_{[2k]}$ and is thus a combination of the $\sigma_{2,j}$. Also, $\sigma_{8,0} \cdot \sigma_{8,4}$ is preserved by all $T_{[4k]}$ and is thus a combination of the $\sigma_{4,j}$.

This abbreviated argument allows us to specify those primes p to which we can generalize Gauss' proof to show that the pth roots of unity are quadratic irrationals. An analysis shows that we have only used the fact that $p - 1 = 2^k$ (at each step the groups were divided in half) and the numbering of the roots, which relied on 3 being a primitive root of the prime 17. We could have used any primitive root for the numbering. As we have already noted, every prime p has at least one primitive root. Incidentally, one can show that 3 is primitive for all p of the form $2^k + 1$ (prove it!). We also remark that if $p = 2^k + 1$ is prime, then $k = 2^r$. Thus, *we have proved it is possible to construct a regular p-gon with straightedge and compass for all primes p of the form* $p = 2^{(2^r)} + 1$.

Primes of this form have their own history, and are called *Fermat primes*. Fermat proposed that all such numbers are prime. Indeed, for $r = 0$ we obtain 3, for $r = 1$ we have 5, and for $r = 2$ we get 17. For $r = 3$ we obtain 257 and $r = 4$ yields 65,537, both of which are prime. For $r = 5$ we obtain 4,294,967,297. Fermat found no prime divisors of this number, but Euler explained that Fermat had "overlooked" the divisor 641. We now know that the Fermat numbers are composite for $r = 6$, 7, 8, 9, 11, 12, 15, 18, 23, 36, 38, 73 (for example, $5 \cdot 2^{75} + 1$ is a prime divisor of $r = 73$). It has been conjectured that there are only a finite number of Fermat primes.

As for regular n-gons with composite n, properties (1) and (2) on p. 268

immediately imply that the desired construction is possible for all $n > 2$ of the form $n = 2^k p_1 p_2 \dots p_l$, where $p_1, p_2, \dots, p_l$ are distinct Fermat primes. Remarkably, there are no other values of n for which the construction is possible. Gauss did not publish a proof of this claim: "The limits of the present work exclude this demonstration here, but we issue this warning lest anyone attempt to achieve geometric constructions [i.e., with straightedge and compass] for sections other than the ones suggested by our theory (e.g., sections into 7, 11, 13, 19, etc., parts) and so spend his time uselessly."[8] Gauss' result implies the possibility, in principle, of constructing a regular p-gon for $p = 257$ and 65,537 but calculating the roots, let alone describing the construction explicitly, requires a colossal but completely mechanical effort. It is remarkable that people were found who wanted to carry this out not only for $p = 257$ (Friedrich Richelot (1808–1875) did it in an 80-page paper, and there is reason to believe that Gauss himself also did) but also for $p = 65,537$ (the solution obtained by Johann Hermes (1846–1912) is in Göttingen, in a trunk of considerable proportions). The English mathematician John Littlewood (1885–1977) once joked about this: "A too-persistent research student drove his supervisor to say, 'Go away and work out the construction for a regular polygon of 65,537 [$= 2^{16} + 1$] sides.' The student returned twenty years later with a construction (deposited in the Archives of Göttingen)."[9]

Concluding Remarks

We have already noted that March 30, 1796, the day Gauss found the construction of the regular 17-gon, decided his destiny. Felix Klein writes, "On this date begins the diary.... We see the proud series of great discoveries in arithmetic, algebra and analysis parade before us.... Among these traces of the burgeoning of a mighty genius one finds, touchingly, little miniatures of school exercises, which even a Gauss was not spared. Here we find a record of conscientious exercises in differentiation; and just before a section on the division of the lemniscate, there are totally banal integral substitutions, such as every student must practice."[10]

Gauss' work has long stood as an unattainable model of mathematical discovery. One of the founders of non-Euclidean geometry, Bolyai János[11] (1802–1860), called it "the most brilliant discovery of our time and even of all time." But it was difficult to comprehend! Thanks to letters written

[8] *Disquisitiones*, p. 459.

[9] *A Mathematician's Miscellany*, Methuen and Company, London, 1953, p. 42.

[10] *Development*, p. 30.

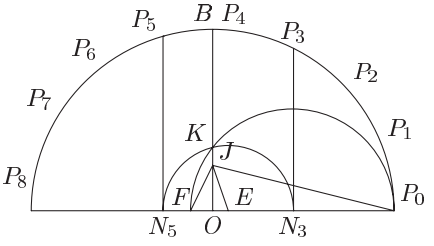
[11] Also known as Johann Bolyai.—*Transl.*

to his homeland by the great Norwegian mathematician Niels Abel (1802–1829), who proved that the general fifth-degree equation is not solvable in radicals, we know the difficult path he followed in studying Gauss' theory. In 1825, Abel wrote from Germany: "Even if Gauss is the greatest genius, he evidently did not try to have everything understood all at once." He decided not to meet with Gauss, but later wrote from France: "I finally succeeded in lifting the veil of mystery that has so far surrounded the theory of the division of the circle created by Gauss." Gauss' work inspired Abel to construct a theory in which "there are so many remarkable theorems that it is simply unbelievable." He then planned to go to Germany, to "take Gauss by storm." Gauss also undoubtedly influenced Galois.

All his life, Gauss retained a touching love for his first discovery: "It is said that Archimedes willed that a monument be placed over his grave in the form of a sphere and cylinder, in memory of his having found that the ratio of the volumes of a cylinder and a sphere inscribed in it is 3:2. Like Archimedes, Gauss expressed the wish that the 17-gon be immortalized in a monument on his grave. This shows what significance Gauss himself placed on his discovery. This picture is not on Gauss' tombstone, but a monument erected to Gauss in Braunschweig stands on a seventeen-sided pedestal, although this is hardly noticeable to the observer" (Heinrich Weber (1795–1878)).

Addendum

Here is an extract from Coxeter's *Introduction to Geometry*[12] containing a recipe by Herbert Richmond (1863–1948) for constructing the regular 17-gon:



Join P_0 to J, one quarter of the way from O to B. On the diameter through P_0 take E, F, so that $\angle OJE$ is one quarter of OJP_0 and $\angle FJE$ is 45°. Let the

[12]H. S. M. Coxeter, *Introduction to Geometry*, Wiley, New York, 1969, p. 27. Reprinted with permission.

circle with diameter FP_0 cut OB in K, and let the circle with center E and radius EK cut OP_0 in N_3 (between O and P_0) and N_5. Draw perpendiculars to OP_0 at these two points, to cut the original circle in P_3 and P_5. Then the arc P_3P_5 (and likewise P_1P_3) is $\frac{2}{17}$ of the circumference. (The proof involves repeated application of the principle that the roots of the equation $x^2 + 2x \cot 2C - 1 = 0$ are $\tan C$ and $-\cot C$.)

II The Golden Theorem

> . . .I chanced on an extraordinary arithmetic truth. . . since I considered it so beautiful in itself and since I suspected its connection with even more profound results, I concentrated on it all my efforts in order to understand the principles on which it depended and to obtain a rigorous proof. When I succeeded in this I was so attracted by these questions that I could not let them be. Gauss[13]

Gauss' diary, the chronicle of his remarkable discoveries, begins on March 30, 1796, the day he constructed the regular 17-gon. The next entry appears on April 8th, and talks of the proof of what he called the *theorema aureum* (golden theorem). Fermat, Euler, and Lagrange had proved special cases of this assertion. Euler stated the general conjecture, and Legendre gave an incomplete proof. On April 8th, Gauss found a complete proof of Euler's conjecture. Incidentally, Gauss did not yet know of the work of his great predecessors. He had traveled the difficult road to the *theorema aureum* independently!

It all began with schoolchild-like observations. Sometimes, looking at a very large integer, we can immediately say that its square root is not an integer. For example, we can use the fact that squares of integers cannot end in 2, 3, 7, or 8. Sometimes we can use the fact that the square of an integer is either divisible by 3 or has remainder 1, but never 2. Both properties are of the same type, since the last digit is the remainder after division by 10. Gauss was interested in the general problem: what are the possible remainders when we divide a square by a prime? We will now consider this question.

Quadratic Residues

We will assume below that p is a prime number not equal to 2. When we divide an integer by p, we may have a "deficit" or a "surplus," i.e., we can

[13] *Disquisitiones Arithmeticae*, p. xviii.

consider either positive or negative remainders. We will agree to choose the remainder that is smallest in absolute value.

It is not hard to prove that if p is odd, then every integer n can be written uniquely as

$$n = pq + r, \quad |r| \le \frac{p-1}{2}, \tag{7}$$

where q and r are integers.

We will call r the remainder after dividing n by p, or the residue of n modulo p, denoted as[14]

$$n \equiv r \pmod{p}.$$

Let us write the residues for the first few primes $p > 2$ in Table 1.

Table 1. *Residues* (*remainders*) mod p.

p	$k = \frac{p-1}{2}$	r
3	1	-1 0 1
5	2	-2 -1 0 1 2
7	3	-3 -2 -1 0 1 2 3
11	5	-5 -4 -3 -2 -1 0 1 2 3 4 5
13	6	-6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6
17	8	-8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8

We are interested in the possible residues (remainders) of squares of integers. We will call these quadratic residues, and the others quadratic nonresidues.

The numbers n^2 and r^2, where r is the remainder of n modulo p, have the same remainder after division by p. Therefore, if we want to find quadratic residues, it suffices to square only residues, i.e., integers r with $|r| \le k = \frac{1}{2}(p-1)$. It is enough to consider $r \ge 0$.

We will carry out the calculations for the primes in Table 1 and construct a new table (Table 2) in which the boldface numbers are the quadratic residues.

Let us try to find some patterns and see how general they are. First, *in each row there are exactly $k+1$ boldface numbers*. We will show that this holds for all primes $p > 2$. It follows from what we said earlier that each odd p (nonprime as well) has at most $k+1$ quadratic residues. We can show

[14] What we are calling the residue (remainder) is usually called the smallest absolute residue (remainder). We have shortened its name since we will not be dealing with other residues. The notation $n \equiv r \pmod{p}$ is also used in a more general situation: it means that p divides $n - r$.

Table 2. *Quadratic residues and nonresidues* mod p

p	$k=\frac{p-1}{2}$	r
3	1	**−1** 0 1
5	2	−2 **−1** 0 1 2
7	3	**−3** −2 −1 0 1 2 3
11	5	−5 −4 −3 **−2** −1 0 1 2 3 4 5
13	6	−6 −5 −4 **−3** −2 **−1** 0 1 2 3 4 5 6
17	8	**−8** −7 −6 −5 **−4** −3 **−2** **−1** 0 1 2 3 4 5 6 7 8

there are exactly $k+1$ if we see that the numbers r^2, for $0 \le r \le k$, have distinct remainders after division by p. If $r_1 > r_2$ and r_1^2, r_2^2 have the same remainder, then p divides $r_1^2 - r_2^2$. Since p is prime, it divides either $r_1 + r_2$ or $r_1 - r_2$, which is impossible since $0 < r_1 \pm r_2 < 2k < p$. This is the first time we have used the fact that p is prime (show our claim is not true for composite numbers).

Fermat's Theorem and Euler's Criterion

Furthermore, 0 and 1 are clearly boldface in each row. No pattern is immediately visible for the boldface numbers in the other columns. Begin with $a = -1$. It is boldface for $p = 5, 13, 17, \ldots$, but not for $p = 3, 7, 11, \ldots$. You may have noticed that the primes in the first group have remainder 1 after division by 4, while those in the second have -1 (for primes $p \neq 2$, no other remainders are possible). Thus we can propose that -1 is a quadratic residue for primes of the form $p = 4l + 1$ and a quadratic nonresidue for $p = 4l - 1$. This pattern was first noted by Fermat, but he left no proof. Try to prove it yourself! You will see that the main difficulty lies in finding how to use the assumption that p is prime. It is not at all clear how to do this, and without this assumption the claim becomes false.

After several unsuccessful attempts, the first proof was found by Euler in 1747. In 1755 he found a different, quite elegant proof, using *Fermat's Little Theorem*: *If p is prime, then for each integer a, $0 < |a| < p$,*

$$a^{p-1} \equiv 1 \pmod{p}. \tag{8}$$

Proof. For $p = 2$ the assertion is obvious, and so we may assume p is odd. Consider the p numbers $0, \pm a, \pm 2a, \pm 3a, \ldots, \pm ka$, where $k = \frac{1}{2}(p-1)$. These have distinct remainders after division by p, since otherwise p divides $r_1 a - r_2 a$ for some $r_1 > r_2$, $|r_1| \le k$, $|r_2| \le k$. But p divides neither a nor $r_1 - r_2$, since $0 < r_1 - r_2 < p$. Multiply these numbers together, except for 0,

to get $(-1)^k(k!)^2a^{p-1}$. Since all nonzero residues are among the remainders of these factors, and by the rule for the remainder of a product, we find that the product has the same residue as $(-1)^k(k!)^2$, i.e., p divides $(k!)^2(a^{p-1}-1)$. Since p does not divide $k!$ $(0 < k < p)$, it divides $a^{p-1}-1$, which completes the proof. □

Corollary (Euler's criterion for quadratic residues). *A residue $b \neq 0$ is quadratic if and only if*

$$b^k \equiv 1 \pmod{p}, \quad k = \frac{p-1}{2}. \tag{9}$$

Proof. It is easy to establish that condition (9) is necessary. By (8), if $a^2 \equiv b \pmod{p}$, $0 < a < p$, then $a^{2k} = a^{p-1}$ and b^k must each have residue equal to 1. It is more complicated to prove sufficiency, and we will deduce it from the following lemma. □

Lemma 1. *Let $P(x)$ be a polynomial of degree l and let p be prime. If there are more than l distinct residues r modulo p for which*

$$P(r) \equiv 0 \pmod{p}, \tag{10}$$

then (10) *holds for all residues.*

Proof. We proceed by induction on l. For $l = 0$ the assertion is obvious. Suppose it is true for all polynomials of degree at most $l-1$. Let $r_0, r_1, \ldots, r_l$, $0 \leq r_j < p$, satisfy $P(r) \equiv 0 \pmod{p}$. We represent $P(x)$ in the form $P(x) = (x - r_0)Q(x) + P(r_0)$, where $Q(x)$ is a polynomial of degree $l-1$ and p divides $P(r_0)$. Then p divides $(r_j - r_0)Q(r_j)$ for $1 \leq j \leq l$. Since p cannot divide $r_j - r_0$ it must divide $Q(r_j)$, and so by the induction assumption p divides $Q(r)$ for all r. Thus, p divides $P(r)$ for all r. □

We will apply the lemma to $P(x) = x^k - 1$. Then the k nonzero quadratic residues satisfy (10). But there is a residue $(r = 0)$ not satisfying (10), so by the lemma no quadratic nonresidue can satisfy (10), and thus (9) is also sufficient.

Remark. If b is a quadratic nonresidue, then $b^{(p-1)/2} \equiv -1 \pmod{p}$. Indeed, if $b^{(p-1)/2} \equiv r \pmod{p}$, then $r^2 \equiv 1 \pmod{p}$, so $r = -1$. (Only the residues $r \equiv 1 \pmod{p}$ and $r \equiv -1 \pmod{p}$ satisfy the congruence $r^2 \equiv 1 \pmod{p}$.)

Euler's criterion allows us to determine instantly the primes p for which the residue -1 is quadratic. Substituting $b = -1$ into (9), we obtain that (9) holds when $p = 4l + 1$ (k is even) but not when $p = 4l - 1$ (k is odd). The conjecture we stated above is now a theorem.

Problem 4. Prove that if $p \neq 2$ is a prime divisor of $n^2 + 1$, then $p = 4l + 1$.

Thus we have proved that -1 is a quadratic residue for $p = 4l + 1$ and a quadratic nonresidue for $p = 4l - 1$.

Let us consider several aspects of this proof. The assertion consists of two parts: a negative assertion for $p = 4l-1$ and a positive one for $p = 4l+1$. For the former, it is natural to try to find some property that quadratic residues satisfy but -1 does not, which is what Euler did. The property we found turned out to be characteristic, i.e., we proved the latter at the same time. If you try to prove this independently, you would probably try to construct a specific number n^2 whose remainder after division by $p = 4l + 1$ is -1. Euler's proof is not efficient in the sense that it does not explicitly construct n in terms of p, but only confirms its existence. In other words, it guarantees that if we go through the numbers $1, 2, \ldots, 2l$, divide their squares by p and take the remainders, then sooner or later we will get -1. The question remains whether there is a more specific construction of n and p that does not use this exhaustive procedure. In 1773, Lagrange (1736–1813) gave a positive answer, using the following theorem.

Wilson's Theorem.[15] *If $p = 2k + 1$ is prime, then*

$$(-1)^k (k!)^2 \equiv -1 \pmod{p}. \tag{11}$$

We use Lemma 1 to prove this theorem. Let $P(x)$ be the product $(x^2 - 1)(x^2 - 4) \cdots (x^2 - k^2)$ and $Q(x) = x^{2k} - 1$. Then $R(x) = P(x) - Q(x)$ is a polynomial of degree at most $2k - 1$, which is divisible by p for $x = \pm 1, \pm 2, \ldots, \pm k$ (since this is true for P and Q). By the lemma, $R(x) \equiv 0 \pmod{p}$ for all x. The only really new fact is that $R(0) \equiv 0 \pmod{p}$. Since $R(0) = (-1)^k (k!)^2 + 1$, we obtain (11).

Lagrange's Corollary. *For $p = 4l + 1$, $[(2l)!]^2 \equiv -1 \pmod{p}$.*

Problem 5. Prove that if (11) is true, then p is prime.

This problem gives us an excuse to note that in Lagrange's construction, it is essential that p be prime.

Having explained when $a = -1$ is a quadratic residue, Euler, using an enormous amount of data, tried to find analogous conditions for other values of a. He noticed that for $a = 2$ everything depends on the remainder after dividing p by 8. In fact, 2 turns out to be a quadratic residue for primes $p = 8l \pm 1$ and a nonresidue for $p = 8l \pm 3$. (The remainder after dividing an odd prime by 8 can only be ± 1 or ± 3.) Moreover, 3 is a quadratic residue

[15] John Wilson (1741–1793) was a lawyer who studied mathematics at Cambridge.

for $p = 12l \pm 1$ and a quadratic nonresidue for $p = 12l \pm 5$. Euler conjectured that in general, everything is determined by the remainder after dividing p by $4a$.

Euler's Conjecture.[16] *Let $0 < r < 4a$. Either a is a quadratic residue for all primes in the arithmetic progression $4aq + r$, $q = 0, 1, 2, \ldots$, or it is a quadratic residue for none.*

Clearly, if $4a$ and r have a common divisor $s > 1$, then there will be no primes in this progression. If they are relatively prime, then, by a theorem of Lejeune Dirichlet (1805–1859), the progression contains infinitely many primes. (This generalizes the theorem that there are infinitely many primes in the sequence of all natural numbers.)

Let us return to Euler's conjecture. It turned out that his criterion, which served us well for $a = -1$, spurns us even for $a = 2$. Euler was unable to work through this case. Apart from $a = -1$, he was only able to prove his conjecture for $a = 3$. Lagrange, whom we have already mentioned, later proved it for $a = 2, 5, 7$. In 1785 Legendre proposed a general proof, but it contained gaps in essential places.

Gauss' proof. Gauss, like his predecessors, first treated the case $a = -1$ and then, already guessing the general result, examined case after case and advanced further than the others: He considered $a = \pm 2$, ± 3, ± 5, and ± 7. The general case (Euler's conjecture) did not yield to the first attack: "This theorem bothered me for a whole year and did not yield to the most strenuous efforts." Note that this was where Gauss "caught up to" the mathematics of his time: The efforts of the best mathematicians, who were trying to prove Euler's conjecture, were fruitless.

Finally, on April 8, 1796, he found a general proof, which Leopold Kronecker (1823–1891) quite aptly called "a test of the power of Gauss' genius." The proof was by double induction on a and p. Gauss had to conceive eight essentially different arguments for eight different cases! He must have been not only astonishingly inventive but also surprisingly courageous to continue on this path. Gauss later found six other proofs of the *theorema aureum* (about fifty are now known). As often happens, once the theorem was proved many simpler proofs were found. We will present here a proof that differs only a little from Gauss' third proof. At its heart lies a key lemma, proved by Gauss no earlier than 1808. □

Lemma 2. *Let $p = 2k + 1$ be prime, a be an integer, $0 < |a| \leq 2k$, $r_1, r_2, \ldots, r_k$ be the residues of $a, 2a, \ldots, ka$, and v be the number of negative residues among*

[16] This is what Gauss called *theorema aureum*.

them. Then

$$a^k \equiv (-1)^\nu \pmod{p}. \tag{12}$$

Applying Euler's criterion, we obtain the following corollary.

Gauss' Criterion for Quadratic Residues. *A residue is quadratic if and only if ν is even, where ν is the number in Lemma* 2.

***Proof of Lemma* 2.** Note that the absolute values of the residues $r_1, r_2, \ldots, r_k$ are all distinct. This follows from the fact that p does not divide the sum or difference of any two of them: $r_i \pm r_j = (i \pm j)a$, $i \neq j$, $|i \pm j| < p$, $|a| < p$. Thus, the moduli $|r_1|, \ldots, |r_k|$ are the numbers $1, \ldots, k$ in some order. So $a \cdot 2a \ldots \cdot ka = a^k k!$ has the same remainder after division by p as $r_1 \ldots r_k = (-1)^\nu k!$. Since the prime p does not divide $k!$, we obtain (12). □

Proof of Euler's Conjecture. We note that the fact that p is prime is no longer used in the argument, but it was used in full measure in Gauss' lemma. If $a > 0$, we mark off the points $\frac{mp}{2}$ on a number line, while if $a < 0$ we mark off $-\frac{mp}{2}$, where $m = 0, 1, 2, \ldots, |a|$ (Figure 2(a)–(b). They divide the line into intervals, which we number according to their left endpoints. Now we mark the points $a, 2a, \ldots, ka$ with crosses. Since a is an integer not divisible by p, the crosses cannot fall at the points we marked but fall within the intervals ($\frac{|a|p}{2} > |a|k$). It is easy to see that the number ν in the lemma is *the number of crosses in the odd-numbered intervals* (prove it!).

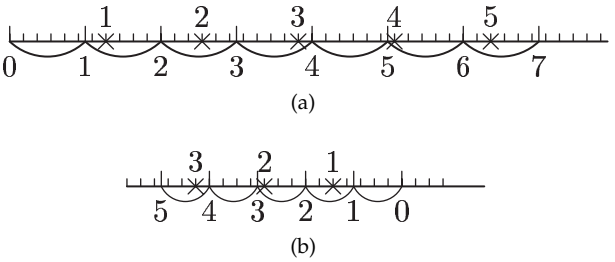


Figure 2. (a) $p = 11$ ($k = 5$), $a = 7$, $\nu = 3$. (b) $p = 7$ ($k = 3$), $a = -5$, $\nu = 2$.

We now apply a similarity transformation with coefficient $\frac{1}{a}$ to our picture (Figure 2 becomes Fig 3). The points $\frac{mp}{2}$ are mapped to points that divide $[0, \frac{p}{2}]$ into $|a|$ equal parts, and the crosses are mapped to the integers $1, 2, \ldots, k$.

The numbering of the intervals will now depend on the sign of a. They are numbered according to their left endpoints for $a > 0$, and according to their right endpoints for $a < 0$; ν is the number of integers in the odd-numbered intervals. If we increase p by $4al$, then we add exactly $2l$ integers

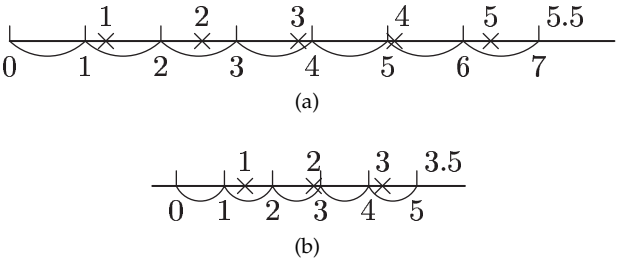


Figure 3.

to each interval. This follows because translating an interval by an integer does not change the number of integers it contains, and every interval (closed or not) of integer length n with noninteger endpoints contains exactly n integers (prove it!). This, in passing from p to $p + 4al$ the value of ν changes by an even number, and $(-1)^\nu$ does not change. This means that $(-1)^\nu$ is the same for all p in the arithmetic progression $p = 4aq + r$, which proves Euler's conjecture. □

At the same time, this gives a method for deciding if a is a quadratic residue for p. Divide p by $4a$ and let r by the remainder (positive for convenience). Divide $(0, \frac{r}{2})$ into $|a|$ parts, numbered according to their left (right) endpoints if a is positive (negative). Count the number ν of integers in the odd-numbered intervals. Then a is a quadratic residue if and only if ν is even.

We will carry out the calculations for $a = 2$ to confirm Euler's observation on p. 286, that everything depends on the remainder after dividing p by 8. Let $a = 2$; then it suffices to consider $r = 1, 3, 5, 7$, since in all other cases the arithmetic progression will contain no primes. As Figure 4 shows, 2 is a quadratic residue for $p = 8q + 1$ and $p = 8q + 7$, i.e., for $p = 8q \pm 1$.

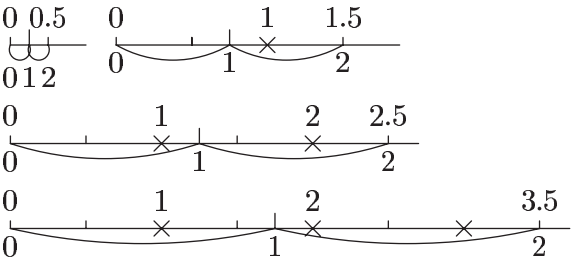


Figure 4. $r = 1$, $a = 2$, $\nu = 0$; $r = 3$, $a = 3$, $\nu = 1$; $r = 5$, $a = 2$, $\nu = 1$; $r = 7$, $a = 2$, $\nu = 2$.

Exercise. Show that -2 is a quadratic residue for $p = 8q+1$ and $p = 8q+3$.

We treat $a = \pm 3$ analogously. This table shows the values of ν:

	$r=1$	$r=3$	$r=5$	$r=7$
$a=3$	0	1	1	2
$a=-3$	0	1	2	3

Thus, 3 is a quadratic residue for $p = 12l \pm 1$ and a nonresidue for $p = 12l \pm 5$, and -3 is a quadratic residue for $p = 12l + 1$ and $p = 12l - 5$.

For $a = 2, 3$ you have of course noted yet another pattern: Primes whose remainders after dividing by $4a$ have the same absolute value are either both quadratic residues or both quadratic nonresidues. Euler did not fail to notice this, and he stated his conjecture in a stronger form than we have.

Supplement to Euler's Conjecture. *Let p and q be primes with $p + q = 4a$. Then a is a quadratic residue either for both p and q or for neither.*

Proof. We carry out the construction in the proof of Euler's conjecture for the intervals $(0, \frac{p}{2})$ and $(0, \frac{q}{2})$, with $a = \frac{p+q}{4}$. For convenience we arrange the intervals so that they are directed oppositely from 0, i.e., we reverse $(0, \frac{q}{2})$, as in Figure 5. Let $\nu(p)$ and $\nu(q)$ be the number of integers in the respective odd-numbered intervals. It suffices to prove that $\nu(p) + \nu(q)$ is even. Let $\nu_j(p)$ and $\nu_j(q)$ be the number of integers in the respective jth intervals. It is easy to see that $\nu_j(p) + \nu_j(q) = 2$ for $j > 0$, which will imply the result we need. □

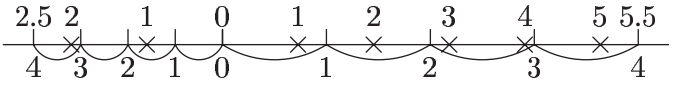


Figure 5. $p = 11$, $q = 5$, $a = \frac{p+q}{4}$, $\nu(p) = 2$, $\nu(q) = 2$.

Indeed, there are $2j$ integer points in the interval between the jth left and right points ($j > 0$) since, as we have already noted, there are $2j$ integers in any interval of length $2j$ with noninteger endpoints.

The Law of Quadratic Reciprocity

In 1798, Legendre[17] pointed out a very convenient statement equivalent to Euler's conjecture, the *law of quadratic reciprocity*. We introduce the following notation, known as the Legendre symbol:

$$\left(\frac{a}{p}\right) = \begin{cases} +1 & \text{if } a \text{ is a quadratic residue modulo } p, \\ -1 & \text{if } a \text{ is a quadratic nonresidue.} \end{cases}$$

By Euler's criterion and the remark following it (p. 284),

$$\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}. \tag{13}$$

This immediately implies the multiplicative property of the Legendre symbol:

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right). \tag{14}$$

We also note that the definition of the Legendre symbol may be extended to all a not divisible by p so that (13) and (14) continue to hold, by setting

$$\left(\frac{a+p}{p}\right) = \left(\frac{a}{p}\right). \tag{15}$$

We can now state the law of quadratic reciprocity.

The Law of Quadratic Reciprocity. *If p and q are odd primes, then*

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\cdot\frac{q-1}{2}}. \tag{16}$$

In other words, $(\frac{p}{q})$ and $(\frac{q}{p})$ have opposite signs if $p = 4l+3$ and $q = 4m+3$ and coincide otherwise.

This is called a "reciprocity" law because it establishes a reciprocity between p as a quadratic residue modulo q and q as a quadratic residue modulo p.

Proof. In all cases, either $p - q = 4a$ or $p + q = 4a$.

[17] Euler was apparently the first to conjecture the law of quadratic reciprocity, but both his and Legendre's proofs were incomplete. Gauss was the first to supply a correct proof. See Morris Kline, *Mathematical Thought from Ancient to Modern Times*, Oxford University Press, New York, 1972, pp. 611–612, 814–815.—*Transl.*

Case 1. Suppose $p - q = 4a$, i.e., p and q have the same remainder after dividing by 4. Then $(\frac{p}{q}) = (\frac{q+4a}{q}) = (\frac{4a}{q}) = (\frac{a}{q})$, by (15), (14), and the fact that $(\frac{4}{q}) = 1$ for all q. Moreover, $(\frac{q}{p}) = (\frac{p-4a}{p}) = (\frac{-4a}{p}) = (\frac{-1}{p})(\frac{a}{p})$. By Euler's conjecture, which we have already proved, $(\frac{a}{p}) = (\frac{a}{q})$, i.e., $(\frac{p}{q}) = (\frac{q}{p})$ when $(\frac{-1}{p}) = 1$ and $(\frac{p}{q}) = -(\frac{q}{p})$ when $(\frac{-1}{p}) = -1$. It remains to recall that $(\frac{-1}{p}) = 1$ for $p = 4l + 1$ and $(\frac{-1}{p}) = -1$ for $p = 4l + 3$.

Case 2. Suppose $p + q = 4a$, i.e., p and q have different remainders after dividing by 4. We have $(\frac{p}{q}) = (\frac{4a-q}{q}) = (\frac{a}{q})$. Analogously, $(\frac{q}{p}) = (\frac{a}{p})$. By the supplement to Euler's conjecture, $(\frac{a}{p}) = (\frac{a}{q})$, i.e., $(\frac{p}{q}) = (\frac{q}{p})$.

The proof is complete. □

It is not hard to see that these arguments can be reversed to deduce Euler's conjecture and its supplement from the quadratic reciprocity law (do it!). Also, formulas (14)–(16) give us a way to compute $(\frac{p}{q})$ that is substantially simpler than the combinatorial method described above. We illustrate it by an example:

$$\left(\frac{59}{269}\right) = \left(\frac{269}{59}\right) = \left(\frac{59 \cdot 4 + 33}{59}\right) = \left(\frac{3}{59}\right) \cdot \left(\frac{11}{59}\right) = -1,$$

since $(\frac{3}{59}) = -(\frac{59}{3}) = -(\frac{2}{3}) = 1$ while $(\frac{11}{59}) = -(\frac{59}{11}) = -(\frac{4}{11}) = -1$. It is easy to show that the computation of the Legendre symbol can always be reduced to the case where p or q equals 2.

Exercise. Compute $(\frac{37}{557})$ and $(\frac{43}{991})$.

In conclusion, we note that the problem of quadratic residues served as the starting point for a great and fruitful mathematical endeavor. Gauss' many attempts to obtain new proofs of the law of quadratic reciprocity were not motivated primarily by a desire to simplify its proof. The thought never left him that he had not really uncovered the deep patterns that gave rise to the law. These patterns were fully revealed only later, as part of the theory of algebraic numbers. Gauss spent a great deal of effort on generalizing the quadratic law to the cubic and biquadratic cases and obtained remarkable results. This research has continued, and the study of various reciprocity laws remains one of the central problems of number theory to this day.

III Royal Days

We have described in detail Gauss' first two great discoveries, made in Göttingen during a ten-day period, a month before he turned nineteen. The second belongs entirely to arithmetic (number theory), while the first essentially depended on arithmetical considerations. Number theory was Gauss' first love.

The Favorite Science of the Greatest Mathematicians

This was one of the many names Gauss gave to arithmetic, i.e., number theory. At the time, arithmetic had already turned from a collection of isolated observations and assertions into a science.

Later, Gauss would write: "The really profound discoveries are due to more recent authors like those men of immortal glory P. De Fermat, L. Euler, L. Lagrange, A. M. Legendre (and a few others). They opened the door to what is penetrable in this divine science and enriched it with enormous wealth."[18]

One of the most surprising sides of the "Gauss phenomenon" is that in his earliest work he only rarely relied on the achievements of his predecessors, but instead quickly rediscovered what had been done in number theory during a century and a half of work by the best mathematicians.

Gauss used his stay in Göttingen to study the classic works, rethink them, and compare them with what he had himself discovered. In his view, the results of this activity should be summed up in a comprehensive work. Gauss set out to write such a book after returning to Braunschweig in 1798, after completing his university studies. It was to include his own results, which had remained unpublished if we do not count the newspaper notice, which incidentally promised: "This discovery is really only a corollary of a theory with greater content, which is not complete yet, but which will be published as soon as it is complete."[19] It took four years of strenuous work to realize this immense plan.

In 1801, Gauss' famous *Disquisitiones Arithmeticae* (*Investigations of Arithmetic*) appeared. This huge book (over 500 pages in large format) contains his fundamental results: the law of quadratic reciprocity, the problem of dividing a circle, and the question of representing integers in the form $am^2 + bmn + cn^2$ (in particular, as a sum of squares). The book was published with the support of the duke and was dedicated to him. As published, it consisted of seven parts. There was not enough money for an

[18] *Disquisitiones Arithmeticae*, p. xviii.
[19] Hall, *Gauss*, p. 24.

eighth, which was to be devoted to a generalization of the reciprocity law to powers greater than two, and in particular to the law of biquadratic reciprocity. Gauss did not find a complete proof of the latter until October 23, 1813, when he noted in his journal that this coincided with the birth of his son.

Klein wrote, "In the *Disquisitiones Arithmeticae* Gauss created modern number theory in its true sense and fixed its whole subsequent development. Our amazement at this achievement increases when we consider that Gauss created this whole world of thought purely out of himself and by himself, without any outside stimulus."[20]

Apart from *Disquisitiones Arithmeticae*, Gauss essentially discontinued his work in number theory. He only thought through and completed what he had conceived in those years. For example, he thought of six more proofs of the quadratic reciprocity law. *Disquisitiones* was far ahead of its time. Gauss had no serious mathematical contact while writing it, and for a long time after the book appeared it was not intelligible to any of the German mathematicians. It was not brought to France, where he could have counted on the interest of Lagrange, Legendre, etc. The bookseller who was supposed to distribute it there went bankrupt, and a great many copies disappeared. As a result, Gauss' students later had to copy out parts of the book by hand. The situation in Germany began to change only in the 1840s, when Dirichlet studied *Disquisitiones* in depth and lectured on it. The book came to Kazan, to Bartels and his students, in 1807.

Disquisitiones turned out to have an enormous influence on the development of number theory and algebra. Starting from Gauss' work on dividing the circle, Galois successfully analyzed the solvability of equations in radicals. Reciprocity laws occupy a central position in algebraic number theory to this day.

The Helmstadt Dissertation

In Braunschweig, Gauss did not have the books he needed for his work on *Disquisitiones*. Thus he often went to neighboring Helmstadt, where there was a good library. Here in 1798, Gauss prepared a dissertation devoted to a proof of the Fundamental Theorem of Algebra, the assertion that every polynomial with complex (and, in particular, real) coefficients has a complex root. If we want to remain within the domain of real numbers, then the fundamental theorem takes the form: *every polynomial with real coefficients can be decomposed into the product of first- and second-degree polynomials*.

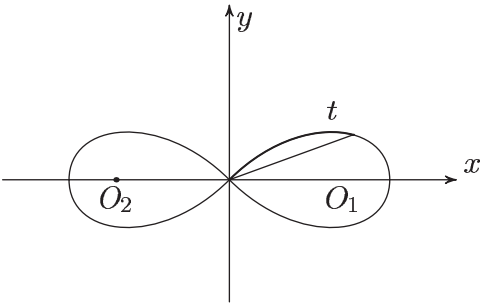
[20] *Development*, p. 24.

Gauss critically reviewed all previous attempts at a proof and pursued an idea of D'Alembert with great care. Gauss' proof was not flawless, since a rigorous theory of continuity was still lacking. He later thought of three more proofs, the last one in 1848.

The Lemniscate and the Arithmetic–Geometric Mean

We will discuss one more direction in Gauss' work that he began in his youth.

In 1791, when Gauss was fourteen years old, he played the following game. He took two numbers a_0, b_0 and constructed their arithmetic mean $a_1 = \frac{a_0+b_0}{2}$ and geometric mean $b_1 = \sqrt{a_0 b_0}$. Then he calculated the means of a_1 and b_1, i.e., $a_2 = \frac{a_1+b_1}{2}$ and $b_2 = \sqrt{a_1 b_1}$, etc. Gauss computed both sequences to a large number of decimal places. Very soon he could no longer distinguish a_n from b_n—all the digits he calculated were the same. In other words, both sequences quickly approached the same limit $M(a_0, b_0)$, called the *arithmetic–geometric mean*.



In those years, Gauss tinkered a lot with a curve known as a *lemniscate* (or Bernoulli's lemniscate), the set of points for which the product of the distances from each of two fixed points O_1, O_2 (the foci) is constant and equals $(\frac{O_1O_2}{2})^2$. Gauss began a systematic study of the lemniscate in 1797. For a long time he tried to find its length, before guessing that it equals $\frac{2\pi}{M(\sqrt{2},2)} O_1O_2$. We do not know how Gauss arrived at this, but we do know that it was on May 30, 1799, and that, not having a proof at first, he computed both quantities to eleven places (!). Gauss thought of a function for the lemniscate, analogous to the trigonometric functions for the circle. For example, for the lemniscate with foci $\sqrt{2}$ units apart, the lemniscate sine $\mathrm{sl}\, t$ is just the length of the chord corresponding to an arc of length t. Gauss spent the last years of the eighteenth century on constructing a theory of lemniscate functions. He obtained addition and multiplication theorems, analogous to theorems for the trigonometric functions.

From lemniscate functions, Gauss turned to their generalization—elliptic functions. He understood that this was "a completely new domain of analysis." After 1800, he could no longer give elliptic functions the time needed to bring the theory to a state that was satisfactorily complete and rigorous. From the very beginning he avoided regular publication, hoping to publish everything at once, as he had done with his work in number theory. But other matters never left him the necessary time.

In 1808, he wrote to his friend and student Heinrich Schumacher (1780–1850): "We now understand the circular and logarithmic functions as well as one times one, but the magnificent golden spring, guarding the secret of the higher functions, is still almost *terra incognita*. I worked very much on this before, and in time will present my own great work on it, which I alluded to in my *Disquisitiones Arithmeticae*. Come and be amazed at the extraordinary richness of the new and most interesting truths and relationships possessed by these functions."

Gauss believed that he did not have to hurry to publish his results. It went on like this for thirty years, and then in 1827 two young mathematicians, Abel and Carl Jacobi (1804–1851), published much of what he had obtained.

"Jacobi's results represent a part of my own great work, which I intended to publish at some time. It will be an exhaustive work on this subject, if only heaven will be pleased to prolong my life and give me strength and peace of mind" (from a letter to Schumacher).

"Mr. Abel anticipated many of my ideas and made my problem easier by about a third, setting forth results with great rigor and elegance. Abel followed the same path that I did in 1798, so there is nothing improbable in the fact that we obtained so similar results. To my surprise, this resemblance even extends to form, and here and there to notation, therefore many of his formulas seem copied from mine. But so that no one should misunderstand me, I must add that I do not recall a single instance when I talked about this research with any outsider" (from a letter to Wilhelm Bessel (1784–1846)).

Finally, in a letter to August Crelle (1780–1855): "Since Abel has demonstrated such insight and such elegance in the problems in his account, I feel that I may completely refrain from publishing the results I obtained" (May 1828).

We should note that Gauss' remark in *Disquisitiones* that the theory of dividing a circle could be extended to the lemniscate turned out to have a great influence on Abel.

With the beginning of the new century, Gauss' scientific interests shifted away from pure mathematics. He returned to it periodically and each time obtained new results worthy of his genius. In 1812, he published a work on

the hypergeometric function. (This function depends on three parameters, and by giving them specific values we can obtain most of the functions occurring in mathematical physics.) Gauss' contribution to the geometric interpretation of complex numbers is widely known. We will discuss his work in geometry below. But mathematics would no longer be the main work of his life. A characteristic external sign is that in 1801 he stopped making regular entries in his journal, although there are notes until 1814. We rarely realize how short Gauss' "mathematical century" was—less than ten years—and a great part of this time was devoted to work that was unknown to his contemporaries (elliptic functions).

Asteriods

We will now talk about Gauss' new inclinations. Many biographers have argued about the reasons why Gauss began to study astronomy. We must first keep in mind that, beginning with Kepler, Galileo, and Newton, astronomy was the most striking area in which to apply mathematics. This tradition was continued in the work of Euler, D'Alembert, Clairaut, Lagrange, and Laplace. In predicting and explaining celestial phenomena, mathematicians felt as if they were admitted to the mysteries of the universe. Gauss, with his early interest in concrete calculations, could of course not help but test his powers in this traditional arena.

There were also prosaic reasons. Gauss earned a meager living as a *privat-dozent* in Braunschweig, receiving 6 thalers a month. A pension of 400 thalers from his patron duke did not improve his situation enough to support a family, and he was contemplating marriage. It was not simple to obtain a chair in mathematics somewhere, and Gauss was not very attracted to teaching. The broadening net of observatories made a career as an astronomer more accessible.

Gauss had begun to be interested in astronomy while in Göttingen. He carried out some observations in Braunschweig, and used part of his pension from the duke to buy a sextant. He searched for a worthy computational problem, solving minor ones in the meantime. For example, he published a simple method for computing the dates of Easter and other cyclical holidays, in place of the extremely confusing techniques that had been used earlier. The idea for a real problem arose in 1801, under the following circumstances.

On January 1, 1801, the astronomer Giuseppe Piazzi (1746–1826), who was creating a star catalog, discovered an unknown star of the eighth magnitude. He observed it for forty days and then asked the leading astronomers to continue the observations. For various reasons, his request

was not met. In June the information reached Franz Zach (1754–1832), who published the only journal in astronomy at the time. Zach conjectured that this was a "long supposed, now probably discovered, new major planet of our solar system between Mars and Jupiter." Zach's conjecture seemed plausible, and an urgent search for the "missing" planet was required. But this demanded the calculation of its trajectory. Determining an elliptical trajectory from the 9° arc that Piazzi had found was beyond the limits of the computational ability of the astronomers. In September 1801, Gauss dropped everything else and undertook to compute the orbit. In November the calculations were complete. They were published in the December issue of Zach's journal, and on New Year's Eve, exactly a year after Piazzi's sighting, the famous German astronomer Heinrich Olbers (1758–1840), using the trajectory calculated by Gauss, found the asteroid now known as Ceres. It was a genuine sensation!

On March 25, 1802, Olbers discovered another asteroid, Pallas. Gauss quickly computed its orbit, showing that it lay between Mars and Jupiter. The astronomers unhesitatingly accepted the validity of Gauss' computational methods.

Recognition came to Gauss, including election as a corresponding member of the St. Petersburg Academy of Sciences. He was soon invited to take the position of director of the observatory there. Gauss wrote that it was flattering to be invited to the city where Euler had worked, and that he was seriously considering the move. In his letters, Gauss wrote that the weather was often bad in St. Petersburg, so that he would not be too busy with observation and would have time for study. The thousand rubles he would receive were worth more than the 400 thalers he now had, but it was more expensive to live there.

At the same time, Olbers began efforts to keep Gauss for Germany. In 1802 he proposed that the curators of the University of Göttingen offer Gauss the position of director of the newly organized observatory there. Olbers wrote that Gauss "has a positive aversion to a chair in mathematics." It was agreed on, but Gauss did not move until the end of 1807. During that time he married ("life seems to me springlike, always with bright new colors.") In 1806 the duke, to whom Gauss evidently felt a sincere attachment, died from a wound. Now there was nothing to keep him in Braunschweig.

Gauss' life in Göttingen did not proceed sweetly. In 1809, his wife died after giving birth to a son, and then the child died, too. In addition, Napoleon placed a heavy tax on Göttingen, and Gauss himself had to pay the unbearable amount of 2000 francs. Olbers tried to pay the money for him, and so did Laplace, who was right in Paris. Both times Gauss proudly

refused. Another benefactor was found, this time anonymous, and there was no one to whom Gauss could return the money. It was learned much later that this was the Elector of Mainz, a friend of Goethe. "Death is dearer to me than such a life," wrote Gauss, between notes on the theory of elliptic functions. Those around him did not appreciate his work and considered him an eccentric at the least. Olbers calmed Gauss, saying he should not count on people's understanding: "You must take pity on them and serve them."

The year 1809 also saw the appearance of the famous *Theoria Motus Corporum Coelestium* (*Theory of Motion of the Heavenly Bodies, Moving About the Sun in Conic Sections*),[21] which Gauss had completed in 1807. The delay occurred partly from the publisher's fear that there would be no demand for the book in German, and for patriotic reasons Gauss refused to publish it in French. They compromised by publishing it in Latin. This was Gauss' only book on astronomy, although he also published a few articles.

Gauss set out his methods for computing orbits. To persuade the reader of the power of his technique, he repeated the calculation of the orbit of the comet in 1769, which Euler had found at the time after three days of intensive computation (losing sight of it afterwards, according to some sources). Gauss required one hour. The book contained the method of least squares, which remains to this day one of the most widely used methods for working with observational data. Gauss indicated that he had known of this method since 1794, and had used it systematically since 1802. (Legendre had published the method of least squares two years before the appearance of Gauss' *Theoria Motus*.)

Gauss received many honors in 1810: a prize from the Academy of Sciences of Paris and the Gold Medal of the Royal Society of London, as well as election to several academies.

In 1804, the Paris Academy had chosen the theory of perturbations of Pallas as the theme for its grand prize (a gold medal weighing 1 kilogram). The deadline was twice extended (to 1816) in the hope that Gauss would submit an entry. Gauss was assisted in his calculations by his student Friedrich Nicolai (1793–1846), ("a young man, tireless in carrying out calculations"), but they were still not carried through to the end. Gauss stopped working on them, falling into a deep depression.

Gauss continued his regular astronomical activities almost up to his death. The famous comet of 1812 (which "presaged" the fire of Moscow!) was observed everywhere, thanks to his calculations. On August 28, 1851, he observed a solar eclipse. He had many astronomy students

[21] There is an 1857 English translation of this book by Charles Henry Davis (reprinted by Dover, New York, 1963).—*Transl.*

(Schumacher, Gerling, Nicolai, Struve). The excellent German geometers Möbius and Staudt studied not geometry but astronomy with him. He conducted an active correspondence with many astronomers, regularly read articles and books on astronomy, and published reviews. We also learn much about his mathematical activities from his letters to astronomers. How different the image of Gauss the astronomer is from the inaccessible recluse that he seems to mathematicians!

Geodesy

By 1820, the center of Gauss' practical interests had shifted to geodesy. As early as the beginning of the century, he had tried to use measurements of a meridian arc, made by French geodesists to establish a standard of length (the meter), to find the true amount of the earth's flattening.[22] But the arc turned out to be too small. Gauss dreamt of measuring a sufficiently large arc, but was only able to begin in 1820. Although the measurements stretched out over two decades, Gauss could not fully realize his idea. His research on the treatment of results of measurements, carried out in connection with geodesy, were of great importance (his fundamental publications on the method of least squares belong to this time), as well as various geometric results related to the need for making measurements on the surface of an ellipsoid.

In the 1820s the question arose of Gauss moving to Berlin, where he would become the head of an institute. The most promising young mathematicians were to be invited there, above all Jacobi and Abel. The negotiations were drawn out over four years. The disagreement was over whether Gauss would deliver lectures and how much he would be paid—1200 or 2000 thalers a year. The negotiations were unsuccessful, but not completely: In Göttingen, Gauss was paid the salary he wanted in Berlin.

The Inner Geometry of a Surface

We are obliged to geodesy for the fact that, for a comparatively short time, mathematics became one of Gauss' main activities again. In 1816 he thought of generalizing the basic problem of cartography, mapping one surface onto another "so that the image is similar to the original in the smallest parts." Gauss advised Schumacher to select this problem for the Copenhagen Scientific Society prize competition, which was declared in 1822. In the same

[22] In other words, the extent to which its shape is not spherical.—*Transl.*

year Gauss submitted his memoir, in which he introduced parameters[23] allowing a complete solution of the problem, special cases of which had been studied by Euler and Lagrange (mapping a sphere or surface of revolution onto a plane). Gauss described the conclusions of his theory in detail for many concrete cases, some of which arise in geodesy problems.

In 1828, Gauss' fundamental geometry memoir, *Disquisitiones Generales circa Superficies Curvas* (*General Investigations of Curved Surfaces*), appeared.[24] It is devoted to the inner geometry of a surface, i.e., to what is associated with the very structure of the surface and not to its position in space.

Figuratively speaking, the inner geometry of a surface is what we can learn about its geometry "without leaving it." On the surface we can measure length, by stretching a string so that it lies completely on the surface. The resulting curve is called a geodesic (analogous to a line on a plane). We can measure angles between geodesics and study geodesic triangles and polygons. If we deform the surface (thinking of it as an unstretchable and untearable film), then distances between points will be preserved, geodesics will remain geodesics, etc.

It turns out that, "without leaving the surface," we can tell whether or not it is curved. A deformation can never turn a "really" curved surface into a plane. Gauss proposed a numerical measure for the degree of curvature of a surface.

Consider a neighborhood of area ε near a point A of a surface. At each point of this neighborhood we take a normal (a vector perpendicular to the surface) of length one. All normals of a plane will be parallel, but for a curved surface they will diverge. We displace the normals so that their origins are at a single point. Then their endpoints form some figure on the unit sphere. Let $\varphi(\varepsilon)$ be the area of this figure. Then $k(A) = \lim_{\varepsilon \to 0} \frac{\varphi(\varepsilon)}{\varepsilon}$ measures the curvature of the surface at A. It turns out that $k(A)$ is the same for all deformations. In order for a piece of the surface to turn into a plane, $k(A)$ must be zero at all points A of the piece. This measure of curvature is related to the sum of the angles of a geodesic triangle.

Gauss was interested in surfaces of constant curvature. A sphere is a surface of constant positive curvature (at each of its points $k(A) = \frac{1}{R}$, where R is the radius). In his notes, Gauss mentions a surface of revolution with constant negative curvature. Later this was called a *pseudosphere*, and Eugenio Beltrami (1835–1900) discovered that its inner geometry is a hyperbolic non-Euclidean geometry.

[23] Curvilinear coordinates.—*Transl.*

[24] There is an English translation by Adam Hiltebeitel and James Morehead, Raven Press, Hewlett, NY, 1966.—*Transl.*

Non-Euclidean Geometry

According to some information, Gauss had been interested in the parallel postulate as early as 1792, while in Braunschweig. In Göttingen he often discussed this problem with Bolyai Farkas,[25] a Hungarian student. We know from a 1799 letter to Bolyai how clearly Gauss understood that there are many assertions which, if we accept them, allow us to prove the fifth postulate. "I have certainly achieved results which most people would look upon as proof. . . ." And, ". . .the way in which I have proceeded does not lead to the desired goal but instead to doubting the validity of geometry." It is only one step from here to understanding the possibility of non-Euclidean geometry, but that step still had not been taken. This sentence is often erroneously taken as evidence that Gauss had arrived at non-Euclidean geometry as early as 1799.

Gauss' words that he was not able to devote enough time to this problem deserve notice. It is typical that there is no mention in his journal of the problem of parallel lines. It was evidently never at the center of Gauss' attention. In 1804, he rejected Bolyai's attempts to prove the parallel postulate. His letter ends, "However I still hope that at some time, and before my end, these submerged rocks will allow us to pass over them." These words seem to indicate a hope that a proof would be found.

Here is more testimony: "In the theory of parallel lines we are now no further than Euclid was. This is the *partie honteuse* [shameful part] of mathematics, which sooner or later must get a very different form" (1813). "We have not advanced beyond the place where Euclid was 2000 years ago" (1816). But in that very year of 1816 he speaks of "the gap which cannot be filled," and in 1817 we read in a letter to Olbers: "I am coming more and more to the conviction that the necessity of our geometry cannot be proved, at least not *by* human intelligence and not *for* human intelligence. Perhaps we shall arrive in another existence at other insights into the essence of space, which are not unattainable for us. Until then one would have to rank geometry not with arithmetic, which stands a priori, but approximately with mechanics."[26]

At about the same time, Ferdinand Schweikart (1780–1859), a jurist from Königsberg, arrived at the notion that the fifth postulate was impossible to prove. He proposed that an "astral geometry," in which the parallel postulate does not hold, exists alongside Euclidean geometry. Gauss' student Christian Gerling (1788–1864), who was working in Königsberg, wrote to

[25]Sometimes known as Wolfgang Bolyai.—*Transl.*

[26]This translation is taken from G. Waldo Dunnington, *Carl Friedrich Gauss: Titan of Science*, Exposition Press, New York, 1955, p. 180.

his teacher about Schweikart's idea and attached a note from him. In reply, Gauss wrote: "Almost everything is copied from my soul." Schweikart's activities were continued by his nephew Franz Taurinus (1794–1874), with whom Gauss exchanged some letters beginning in 1824.

In his letters Gauss stressed that his statements were of an especially partial nature and must in no case be made public. He did not believe these ideas could be grasped, and feared the interest of hordes of dilettantes. Gauss had spent more than a few difficult years and greatly valued the opportunity to work quietly. He warned Gerling, who had planned only to mention that the parallel postulate could turn out to be false, "but the wasps, whose nest you stir up, will fly around your head." He gradually reached a decision to write down his results but not publish them: "Probably, I will not be able to work out my space research on this question soon enough so that they can be published. Perhaps it will not happen during my lifetime, since I fear the Boeotians[27] cries if I were to express my opinion fully" (from an 1829 letter to Bessel). In May of 1831, Gauss began to make systematic notes: "It has now been several weeks since I began to set out in writing several results of my own thinking on this subject, in part already forty years old but that I never wrote down, as a result of which I had to begin the whole work over again three or four times in my head. I did not want, however, for this to buried along with me" (letter to Schumacher).

But in 1832 he received from Bolyai Farkas a short essay by his son János, *Appendix Scientiam Spatii Absolute* (*The Science of Absolute Space*) (so called because it was published as an appendix to a large book written by the father). "My son values your opinion more than the opinion of all Europe." The content of the book startled Gauss: a complete and systematic construction of non-Euclidean geometry. These were not the fragmentary remarks and conjectures of Schweikart–Taurinus. Gauss himself had intended to produce such an account in the near future. He wrote Gerling: "I found all my own ideas and results, developed with great elegance, although because of the conciseness of the account, in a form that is accessible with difficulty to someone to whom this area is foreign. . . . I believe that this young geometer Bolyai is a genius of the first order." And to the father, he wrote ". . .all of the paper's contents, and the way your son has attacked the matter, coincide almost completely with my own reflections which I partly carried out thirty to thirty-five years ago. In fact I am extremely surprised by it. My intention was to leave my own work, of which at the present time only a small part is written down, unpublished during all of

[27] According to legend, the inhabitants of Boeotia were famous in ancient Greek for their stupidity.

my lifetime. . . . On the other hand, I had intended to write it all down little by little, so that it at least would not disappear with me. I am thus quite surprised that I can spare myself these efforts, and it makes me very happy that it is the son of one of my old friends who has come ahead of me in such a remarkable manner."[28] Bolyai János received no public appreciation or support from Gauss whatsoever. Gauss evidently interrupted his systematic notes on non-Euclidean geometry at that time, although sporadic notes from the 1840 remain.

In 1841 Gauss became acquainted with the German edition of Lobachevsky's works (Lobachevsky's first publications date from 1829). True to form, Gauss was interested in the author's other publications, but commented about him only in letters to his close correspondents. Incidentally, on Gauss' proposal, Lobachevsky was elected in 1842 as a corresponding member of the Göttingen Royal Society, "as one of the most excellent mathematicians of the Russian people." Gauss personally notified Lobachevsky of his election. However, non-Euclidean geometry was not mentioned either in Gauss' presentation or in the diploma given Lobachevsky.

We know of Gauss' work on non-Euclidean geometry only from the posthumous publication of his archives. Gauss thus guaranteed that he could work quietly by refusing to publicize his great discovery, resulting in arguments that continue to this day about the appropriateness of his position.

We should note that Gauss was interested in more than the purely logical question of the provability of the parallel postulate. He was interested in the place of geometry in the natural sciences and the truth of the geometry of our physical world (see his above statement of 1817). He discussed the possibility of an astronomical verification, speaking with interest of Lobachevsky's ideas in this regard. In his work on geodesy, Gauss even measured the sum of the angles of the triangle formed by the German mountains of Hohenhagen, Brocken, and Inselsberg. It differed from 2π by no more than $0.2°$.

Electrodynamics and Terrestrial Magnetism

By the end of the 1820s Gauss, now on the far side of fifty, began to look for areas of science that were new for him. This is seen in two publications of 1829 and 1830. The former set out his thoughts on general principles of mechanics (his "Principle of Least Constraint"), while the latter was devoted to the study of capillary phenomena. Gauss decided to study

[28]This translation is taken from Hall, *Gauss*, p. 114.

physics, but had not yet defined his specific interests. In 1831 he tried to study crystallography. This was a very difficult year in his life: His second wife died, and he began to suffer from terrible insomnia. In the same year, the twenty-seven-year-old physicist Wilhelm Weber (1804–1891) was invited to Göttingen, at Gauss' initiative. Gauss had met him in 1828, at the home of Alexander von Humboldt's (1769–1859). His reserve had become legendary, but all the same he found a scientific partner in Weber as he had never had before.

"The inner difference between the two men was also expressed quite clearly in their outer appearance. Gauss had a powerful, stocky physique, a true Lower Saxon, laconic and not easily accessible. This was in strong contrast to the small, delicate, agile Weber, whose friendly, loquacious nature betrayed at once the true Saxon, though he was actually born in Wittenberg, in the land of the "double Saxons" [Doppelsachsen]. In the Gauss–Weber Monument in Göttingen this contrast has been minimized for artistic reasons, and even their ages appear closer than they were" (Klein).[29]

Gauss' and Weber's interests lay in the area of electrodynamics and terrestrial magnetism. Their work had not only theoretical but also practical results. In 1833 they invented an electromagnetic telegraph (this event is recorded on their common memorial). The first telegraph connected the observatory and the physics institute but, for financial reasons, its creators did not succeed in developing it further.

In the course of his research, Gauss arrived at the conclusion that an absolute system of physical units was needed for work in magnetism. He began with a number of independent quantities and expressed the remaining quantities in terms of them.

The study of terrestrial magnetism relied both on observations at the magnetic observatory at Göttingen, and on materials that had been collected in various countries by the "Society for the Observation of Terrestrial Magnetism," created by Humboldt after his return from South America. At the same time, Gauss created one of the most important chapters in mathematical physics—potential theory.

Gauss' and Weber's joint work was interrupted in 1843, when Weber and six other professors were expelled from Göttingen after signing a letter to the king that cited breaches in the recent constitution (Gauss had not signed it). Weber returned to Göttingen only in 1849, when Gauss was already 72 years old.

We conclude our story of Gauss with Klein's words: "To me Gauss is like the highest peak amidst our Bavarian mountains as it appears to a

[29] *Development*, pp. 17–18.

Carl Friedrich Gauss.

spectator from the north. From the east the gradually ascending foothills culminate in the one gigantic colossus, which falls away steeply into the lowlands of a new formation, into which its spurs project for many miles and in which the water that streams from it begets new life."[30]

Appendix: Construction Problems Leading to Cubic Equations

In *Disquisitiones Arithmeticae*, Gauss states without proof that it is impossible to construct, with straightedge and compass, regular n-gons for primes n that are not Fermat primes, and in particular, a regular 7-gon. This negative result must have surprised his contemporaries no less than the possibility of constructing a regular 17-gon. After all, $n = 7$ is the first value of n for which no construction was found, despite many attempts. The Greek geometers undoubtedly suspected there was something troublesome about this problem and it was not without cause, let us say, that Archimedes proposed a method for constructing a regular 7-gon using conic sections. However, the question of proving that the construction was impossible evidently did not even arise.

One must say that proofs of negative assertions have always played a

[30] *Development*, p. 57.

fundamental role in the history of mathematics. An impossibility proof requires that every conceivable method of solution, construction, or proof somehow be considered, while it suffices to give just one specific method for a positive solution.

Impossibility proofs in mathematics had a famous beginning, when the Pythagoreans (sixth century B.C.), in trying to reduce all of mathematics to integers, buried this idea with their own hands: It turned out that there exists no fraction whose square equals 2. Another way to say this is that the diagonal and sides of a square are incommensurable. Thus integers and their ratios are insufficient to describe a very simple situation. This discovery surprised the greatest thinkers of ancient Greece. Legend states that the gods punished the Pythagorean who revealed this fact (he died in a shipwreck). Plato (429–348B.C.) tells how the existence of irrational quantities astonished him. Plato once ran into a "practical" problem that caused him to rethink the possibilities of geometry.

"In his work entitled *Platonicus*, Eratosthenes relates that, when God announced to the Delians through an oracle that, in order to be liberated from the pest, they would have to make an altar, twice as great as the existing one, the architects were much embarrassed in trying to find out how a solid could be made twice as great as another one. They went to consult Plato, who told them that the god had not given the oracle because he needed a doubled altar, but that it had been declared to censure the Greeks for their indifference to mathematics and their lack of respect for geometry" (Theon of Smyrna).[31] Do not say that Plato did not make use of the right moment to propagandize for science! According to Eutocius, an analogous problem (doubling the volume of Glaucus' tombstone) figured in one version of the legend of Minos.

We are talking about finding the sides of a cube with twice the volume of a given cube, i.e., of constructing the roots of the equation $x^3 = 2$. Plato sent the Delians to Eudoxus and Helicon. Menaechmus, Archytas, and Eudoxus proposed various solutions, but none found a construction with straightedge and compass. Eratosthenes, who built a mechanical device for solving the problem, later called his predecessors' solutions too complicated in a poem carved on a marble slab in Ptolemy's temple in Alexandria: "Do not thou seek to do the difficult business of Archytas' cylinders, or to cut the cone in the triads of Menaechmus, or to compass such a curved form of lines as is described by the god-fearing Eudoxus."[32] Menaechmus noted that the problem is equivalent to finding two mean proportionals,

[31] As translated in Bartel L. van der Waerden, *Science Awakening* I, translated by Arnold Dresden, Oxford University Press, New York, p. 161.

[32] Ibid.

i.e., for given a and b find x and y that satisfy $a : x = x : y = y : b$. The latter problem was solved using conic sections. We know nothing about Eudoxus' "curved lines." As far as a mechanical solution is concerned, Eratosthenes was not the first. According to Plutarch, "Plato himself censured those in the circles of Eudoxus, Archytas, and Menaechmus, who wanted to reduce the duplication of the cube to mechanical constructions, because in this way they undertook to produce two mean proportionals by a non-theoretical method; for in this manner the good in geometry is destroyed and brought to nought, because geometry reverts to observation instead of raising itself above this and adhering to the eternal, immaterial images, in which the immanent God is the eternal God."[33] Incidentally, Eutocius ascribes to Plato himself (evidently erroneously) a mechanical solution to the Delian problem, using a carpenter's square with grooves and adjustable rulers. Plato, with his aversion to "material things, which require extended operations with unworthy handicrafts" (Plutarch),[34] is often contrasted to Archimedes (287–212B.C.), who was glorified for his many inventions and in particular for the machines used in the defense of Syracuse. The same Plutarch also claimed that Archimedes only yielded to the persuasion of King Hiero" to direct his art somewhat away from the abstract... [and to occupy] himself in some tangible manner with the demands of reality," although he believed that the practical was "lowbrow and ignoble, and he only gave his effort to matters which, in their beauty and their excellence, remain entirely outside the realm of necessity."[35]

Along with the Delian problem, Greek geometry left several other problems for which a construction with straightedge and compass was not found: trisecting an angle (dividing it into three equal parts), squaring the circle, and constructing a regular n-gon, in particular a 7-gon and a 9-gon. The Greeks, and even more so the Arab mathematicians, were aware of the connection between these problems and cubic equations.

The problem of the regular 7-gon reduces to the equation $z^6 + z^5 + z^4 + z^3 + z^2 + z + 1 = 0$ (see equation (2), p. 269), or

$$\left(z^3 + \frac{1}{z^3}\right) + \left(z^2 + \frac{1}{z^2}\right) + \left(z + \frac{1}{z}\right) + 1 = 0.$$

Passing to the variable $x = z + \frac{1}{z}$, we obtain the equation

$$x^3 + x^2 - 2x - 1 = 0.$$

[33] Ibid., p. 163.
[34] Ibid.
[35] Ibid., p. 209.

We will show that the roots of the equations for doubling the cube and for the 7-gon cannot be quadratic irrationals, which will imply the impossibility of constructing them with straightedge and compass. We will prove a result that applies to a rather general situation.

Theorem. *If a cubic equation $a_3x^3+a_2x^2+a_1x+a_0=0$ with integer coefficients has a quadratic irrational root, then it also has a rational root.*

Proof. Let x_1 be a quadratic irrational root. Then x_1 can be obtained from the integers by arithmetic operations and extracting square roots; let us analyze this construction. We first take the square roots of certain rational numbers, $\sqrt{A_1}$, $\sqrt{A_2}, \ldots, \sqrt{A_a}$, and then the square roots of certain numbers obtained by arithmetic operations from the rationals and the $\sqrt{A_i}$, say $\sqrt{B_1}$, $\sqrt{B_2}, \ldots, \sqrt{B_b}$, etc. At each step, we take roots of numbers that are expressed arithmetically in terms of all those obtained previously. We thus obtain "stages" of quadratic irrationals. Let $\sqrt{N}$ be one of the numbers obtained at the last step before forming x_1. We fix our attention on how $\sqrt{N}$ enters into x_1. It turns out that we may assume that x_1 has the form $\alpha+\beta\sqrt{N}$, where α, β are quadratic irrationals into which $\sqrt{N}$ does not enter. It suffices to note that arithmetic operations on expressions of the form $\alpha+\beta\sqrt{N}$ lead to the same type of expression. This is obvious for addition and subtraction and can be directly verified for multiplication. For division we must eliminate $\sqrt{N}$ from the denominator:

$$\frac{\alpha+\beta\sqrt{N}}{\gamma+\delta\sqrt{N}}=\frac{(\alpha+\beta\sqrt{N})(\gamma-\delta\sqrt{N})}{\gamma^2-\delta^2N}.$$

If we now substitute $x_1=\alpha+\beta\sqrt{N}$ into the equation and carry out the operations, we obtain a relation of the form $P+Q\sqrt{N}=0$, where P and Q are polynomials in α, β, a_i. If $Q\neq 0$, then $\sqrt{N}=-\frac{P}{Q}$, and by substituting this into the expression for x_1 we can represent x_1 without $\sqrt{N}$. If $Q=0$, then it can be verified that $x_2=\alpha-\beta\sqrt{N}$ is also a root, and since $-\frac{a_2}{a_3}=x_1+x_2+x_3$ is the sum of the roots (Vieta's theorem), we obtain $x_3=\frac{-a_2}{a_3}-2\alpha$. Thus we again have a root that is a quadratic irrational, expressible in terms of $\sqrt{A_i}$, $\sqrt{B_j}, \ldots$, as is x_1, but without $\sqrt{N}$. Continuing this process further, we eliminate all radicals in the expression for a root of the equation by stages, beginning with the last stage. After this we obtain a rational root, and the proof is complete. □

It now remains to verify that the equations we are interested in have no rational roots. Suppose an equation has leading coefficients $a_3=1$. Then each rational root is an integer. It suffices to substitute $x=\frac{p}{q}$ (p, q

relatively prime) into the equation, multiply both sides by q^3, and see that p^3 and thus p are divisible by q, i.e., $q = 1$. Furthermore, if α is a root, then $x^3+a_2x^2+a_1x+a_0 = (x-a)(x^2+mx+n)$, where $a_2 = -\alpha+m$, $a_1 = -\alpha m+n$, $a_0 = -\alpha n$, i.e., $m = a_2+\alpha$, $n = a_1+a_2\alpha+\alpha^2$. This means that if a_i and α are integers, then m and n are integers and α must divide a_0. So with $a_3 = 1$, the search for rational roots of the equation reduces to sorting out a finite number of possibilities, the divisors of the constant term. It is now easy to verify that the equations we are considering have no integer roots, so they have no roots that are quadratic irrationals.

Felix Klein

Felix Klein's fame is based on work that he carried out over a period of one decade. Klein stopped working actively in mathematics at the age of 33, but to the end of his days he remained at the center of life in scientific organizations and devoted himself completely to pedagogical and literary activities.

A Knight's Spurs

Felix Klein was born in 1849 in Düsseldorf. He completed gymnasium there and in 1865 entered the University of Bonn. By the following year Professor Julius Plücker (1801–1868) had enlisted the 17-year-old student as an assistant in physics. Plücker had started his scientific work as a geometer, but had gradually switched over to experimental physics. However, in the last years of his life, after a twenty-year hiatus, Plücker returned to geometry. "This change played a decisive role in my own development," wrote Klein. The posthumous publication of Plücker's last memoir was prepared by Klein. Perhaps this was the reason why his 1868 dissertation in which Klein said he "earned his knight's spurs" and his first publication in 1869 were in geometry.

Lacking a teacher, Klein became a "knight-errant." He visited the fundamental centers of mathematics in Germany (Göttingen, Berlin) and established personal contacts with Alfred Clebsch (1833–1872), Weber, and Weierstrass. They paid attention to him immediately, which raised the hopes of the young scientist who wanted to and was able to learn. No less important were Klein's contacts with his contemporaries. Klein's friendship with the great Norwegian Sophus Lie (1842–1899) was especially fortunate. They had met in Berlin in 1870. Lie was seven years older than

Felix Klein.

Klein, but in 1870 had taken only his first steps in geometry. Soon Klein and Lie went to Paris. There they became acquainted with the methods of the French geometers, who were able to obtain important geometrical results with surprising ease, "out of thin air" (Lie). Especially significant for the future scientific fates of Klein and Lie were meetings with Camille Jordan (1838–1922). In 1870 Jordan had just come out with a extensive work on the theory of *finite groups*, drawing broad attention to the work of Galois (1811–1832). Perhaps Klein's first work, which was devoted to a geometric study of the so-called Kummer surfaces,[1] served as a "ticket of admission" to Jordan, who had earlier undertaken an algebraic study of these surfaces.

Klein left France because of the Franco-Prussian war. At the start of the war Klein came down with typhus and when he recovered he settled in Göttingen. A time of great achievement was beginning for Klein. Bourbaki wrote that Klein concluded a "golden century" of geometry. But before we talk about the brilliant conclusion of this century, we will recall its beginning.

[1] Eduard Kummer (1810–1893).

A "Golden Century" of Geometry

As early as the 17th century Desargues (1593–1662) and Pascal (1623–1662) used central projections to obtain remarkable geometric results, which were almost forgotten within a century and a half. Later, Gaspard Monge (1746–1818) gave new attention to the great possibilities of the projection method. He discussed this in his course on descriptive geometry, which he gave at the École Polytechnique. Bourbaki counts the "golden century" of geometry as starting with Monge's *Descriptive Geometry* (1795).

Among Monge's students was Jean-Victor Poncelet (1788–1867). Klein wrote, "It was a new kind of geometrical intuition, 'projective thinking,' that enabled him to surpass his predecessors."[2] Poncelet created projective geometry over the course of two years that he spent as a captive in Saratov, Russia, after the Franco-Russian War of 1812. He discussed his results with his fellow captives, who had also heard Monge at the École Polytechnique. These results were published in 1822 in his *Treatise on the Projective Properties of Figures*.

As did his predecessors, Poncelet augmented each line by a point at infinity, assuming that all parallel lines have a common point at infinity where they "intersect." All the points at infinity form the line at infinity. On the augmented plane parallelism becomes a special case of intersection and requires no special consideration (for example, the statement that through a point not on a line there is a unique line parallel to the given one becomes the statement that through two distinct points, one of which is ordinary and the other is at infinity, there is a unique line). Under a central projection a finite point can have no image ("goes to infinity"), but on the plane augmented by the points at infinity this map is one-to-one.

A central projection transforms one plane into another; by applying several projections in succession we can return to the original plane, obtaining a transformation of this plane. Displacements, homotheties, and dilatations belong to these transformations, which came to be called projective transformations. Projective transformations are one-to-one (on the augmented plane) and take lines to lines (it was later clarified that every transformation with these properties is projective). Projective transformations that map the line at infinity into itself are called affine; affine transformations are one-to-one on the usual plane. Poncelet studied geometric objects that are preserved under projective transformations. It turns out that under a projective transformation a conic section maps to a conic section (but, for example, a hyperbola can map to a parabola and every conic

[2]Felix Klein, *Development of Mathematics in the 19th Century*, translated by M. Ackerman, Math Sci Press, Brookline, MA, 1979, p. 73.

section can be mapped to a circle with a projective transformation). The following observation turned out to be especially fruitful. Let A, B, C, D be collinear points and $\{A, B, C, D\} = \frac{AC \cdot BD}{AD \cdot BC}$ be the cross-ratio, or anharmonic ratio, of the four points. Suppose some projective transformation transforms the points A, B, C, D into the points A', B', C', D' (they will necessarily be collinear). Then $\{A, B, C, D\} = \{A', B', C', D'\}$, that is under a projective transformation the cross-ratio of the four points is preserved. If one of the points, D, for example, is a point at infinity then $\{A, B, C, D\}$ is equal to $\frac{AC}{BC}$ and we find that ratios of lengths of collinear segments are preserved by affine transformations. (Why is this true?)

Furthermore, Poncelet tried to eliminate exceptional cases of reciprocal arrangements of conic sections. Why, for example, can two ellipses intersect at four points but a pair of circles in no more than two? This question has a surprising answer. Aside from pairs of *real* points of intersection, circles have a *universal pair* of common points (that are the same for all circles in the plane!), unseen because they are... imaginary and points at infinity at the same time. These points are called *cyclic*.

Now a few words about four German mathematicians: Ferdinand Möbius (1790–1868), Jakob Steiner (1796–1863), Christian von Staudt (1798–1867), and Plücker, who was introduced above. A bitter struggle is associated with their names, between the analytic and synthetic directions in geometry.

Here the words "analysis" and "synthesis" are used in a nonstandard sense: analytic geometry uses the method of coordinates, which makes it possible to apply algebra and analysis to geometry; synthetic geometry operates with direct spatial constructions.

The most bitter relationship was a duel between the analyst Plücker and the synthesist Steiner. Möbius (analyst) and von Staudt (synthesist) stayed on the sidelines. Klein was drawn very easily into the battle on the side of the analysts but was able to stay out of the skirmish, perhaps by following the rule of his acquaintance, the physiologist Carl Ludwig (1816–1895): "one goes about 600 kilometers from home and looks at the situation from there."[3]

The analyst's work required above all perfecting the method of coordinates. For the synthesist it was important to give a coordinate-free definition of the objects of projective geometry, e.g., of second-order curves. Steiner, a very colorful figure in the history of mathematics, did this. A Swiss peasant who walked behind a plough until he was 19, Steiner began to study mathematics at a mature age. He was decisively opposed to using

[3]Klein, *Development of Mathematics in the* 19*th Century*, p. 105.

imaginary numbers in geometry, calling them "ghosts" or "in shadowland." Von Staudt showed that one can associate purely real constructions equivalent to the imaginary objects that arise in projective geometry. A second important achievement of von Staudt is that he was able to determine the cross-ratio of four points directly, without using distances (which are not preserved under projective transformations).

Finally, here is one more name, that of the English mathematician Arthur Cayley (1821–1895), who studied mathematicians for a long time without interrupting his legal practice. We will mention just one work of Cayley, his famous *A Sixth Memoir upon Quantics* (1859). Cayley noticed that Euclidean displacements and homotheties (conformal transforms) are distinguished from all projective transformations by the fact that they preserve cyclic points. As a result, using cyclic points all objects of Euclidean geometry (distances, angles, etc.) can be determined through projective concepts (ones that are preserved under projective transformations). Cayley called projective geometry *descriptive* and Euclidean geometry *metric*, and wrote, "Metric geometry is, thus, a part of descriptive geometry and descriptive geometry is all of geometry." We have to keep in mind that earlier things seemed to be exactly the opposite, namely that projective geometry was a comparatively meager part of Euclidean geometry. Cayley further remarked that, starting from projective geometry one can introduce distances that are different from Euclidean ones (Cayley metrics or mensuration): each such distance in the plane is associated with some second-order curve (real or imaginary), where the distance does not change under any projective transformation that preserves the curve.

The Cayley–Klein Model

In 1869 Klein became acquainted with Cayley's theory and at the end of that year rather superficially with hyperbolic geometry. At the same time he had the idea that one of Cayley's metrics leads to hyperbolic geometry. This was a conjecture with barely any reasoning behind it. Cayley's theory and hyperbolic geometry are radically different externally (Cayley's calculations with the cross-ratio versus Lobachevsky's axiomatic presentation), and Cayley's geometry was not yet developed enough to verify the axioms of hyperbolic geometry. In February, 1870, Klein reported on Cayley's theory in Weierstrass' seminar and decided to put forward his hypothesis. In this seminar it was unusual to discuss "fantasy" projects: they explained to the "presumptuous" young man that "these are two systems that are very far apart." Klein was so ill-prepared to defend his hypothesis that he "allowed them to change my mind." Later he complained that Weierstrass

"did not have the inclination to recognize from a distance the outline of yet unscaled heights." But Klein did not stop believing in his hypothesis. In the summer of 1871, with the help of his friend Otto Stolz (1842–1905), he studied non-Euclidean geometry quite thoroughly and was convinced that his conjecture was true. Even though he possessed a proof, it was not easy for Klein to convince those around him that it was true. The most annoying thing for Klein was probably that up to the end of his life Cayley remained among those who did not agree with him. Klein wrote, "An aging spirit is not in a position to draw conclusions from the positions he himself created."

Here are a few words about the Cayley–Klein model itself. The "points" in this model are the points interior to a circle (the circle can be replaced by a domain that is bounded by any second-order curve), and the "lines" are the chords of this circle (without endpoints). The points of intersection of the "lines" are defined in a natural way; it is clear that through a "point" not on a "line" an infinite number of "lines" pass that do not intersect the original "line," which contradicts the parallel postulate of Euclidean geometry. We still have to convince ourselves that all the remaining Euclidean axioms are satisfied for the model: this will mean that the Klein model is a model of hyperbolic geometry. It is comparatively simple to verify the axioms about the relative positions of points and lines. But when it comes to verifying the axiom of congruence then first of all we have to agree when to consider two segments as being equal; we cannot inherit the corresponding ideas from Euclidean geometry. Cayley, following Klein, sets the length of a segment AB equal to $|\ln\{A, B, \alpha, \beta\}|$, where α and β are the points of intersection of the "line" AB with the boundary of the circle being considered (this is called the absolute circle). Projective transformations that preserve the absolute circle preserve the "distance" defined in this way, i.e., are displacements in the Cayley–Klein model of hyperbolic geometry.

As early as 1868, the Italian mathematician Eugenio Beltrami (1835–1900) saw another way to establish hyperbolic geometry. He discovered a surface, the pseudosphere, where the shortest lines (geodesics) act like the lines in hyperbolic geometry. Then Beltrami transformed the pseudosphere into a circle in a certain way and obtained the same formulas that Klein later did in his theory.

Klein studied other non-Euclidean geometries to which the Cayley metrics led, discovering in particular a model for Riemannian geometry (where the sum of the angles of a triangle is greater than π, while in hyperbolic geometry it is always less than π).

Let us now discuss what the Cayley–Klein model gives for hyperbolic geometry. First of all it differs from the axiomatic method of presentation

in that it is more visual, Klein prefaced his publication (1871) by saying that his goal was "to give a more visual presentation of the mathematical results of the work that relate to the theory of parallel lines, and to make them clearly understandable" (Beltrami stated his goal in roughly the same way). However, the construction of the model solved much more than just the problem of method. The Cayley–Klein model is now considered above all as a way to prove the consistency of hyperbolic geometry. In the Cayley–Klein model the objects of hyperbolic geometry are expressed in the language of Euclidean geometry, so that after translation into this language the theorems of hyperbolic geometry become theorems of Euclidean geometry and thus hyperbolic geometry is consistent if Euclidean geometry is consistent.

Klein saw the basic meaning of the model he constructed differently. He considered projective geometry as being of paramount importance, having Euclidean and hyperbolic geometry as independent parts on a par with one another. This underscores the independence of the model of hyperbolic geometry from Euclidean geometry, for which in turn what was important was the possibility of constructing projective geometry without using Euclidean geometry.[4] Exactly this caused Cayley to suspect a vicious circle. Klein wrote, "In place of the idea of constructing models of the non-Euclidean geometries within the usual metric geometry we have put the task of providing a foundation free of all metric concepts for a superordinate geometry that will comprehend all the familiar geometries in a lucid system."[5]

The Erlanger Program

For centuries the word "geometry" was only used in the singular. But then came Lobachevsky's hyperbolic geometry and Riemannian geometry, and mathematicians finally understood that there exist many different geometries. The natural question arose: What is such a geometry? In 1872 Klein expressed his point of view in lectures that he gave in connection with assuming the position of professor at the University of Erlangen, in Germany. This is how the "Erlanger program," Klein's most well-known written work, came about. It contains essentially no new results but focuses on searching for a principle to allow the systematization of the very amorphous subject that geometry had become at the time.

According to Klein, *the fundamental attribute of every geometry is some collection G of one-to-one transformations of a set M*. There should be enough

[4] According to von Staudt (see p. 315).

[5] Klein, *Development of Mathematics in the 19th Century*, p. 139.

transformations so that each point of M can be mapped to another by some other transformation in G (in this case we say that G acts transitively on M). Such a viewpoint was evoked, of course, by projective geometry, in which, from the very beginning, certain transformations (the central projections) were primary, just as in the traditional presentation of Euclidean geometry other objects were primary: lines, intervals, various figures, etc.

Next, the collection of transformations G must be a *group*. This means that any two transformations in G applied successively can be replaced by a single transformation that is also in G. Moreover, together with each transformation $g \in G$ there is an element of G that is *inverse* to it, denoted by g^{-1} (if g maps x to y, then g^{-1} maps y to x). For example, the motions of a plane or the projective transformations of the projective plane form a group.

Thus, some geometry is associated with each *group G of transformations*. Of what does such a geometry consist? Above all, of finding *invariants of the group G*, properties that are preserved under the action of the transformations in G (more precisely, if some object of our geometry has an invariant property, then no matter what transformation in G we apply to it we obtain an object that also has this property). For the group of displacements of Euclidean geometry all known geometrical properties are invariants, since we do not distinguish the positions of figures in the plane. However, in a traditional geometry course there are nontrivial statements about invariants of transformations that are not displacements. Homotheties preserve the equality of angles, the property that a curve is a circle, ratios of lengths of intervals, and ratios of areas. Given a supply of invariance properties we can construct new ones. For homotheties, the properties that a line bisects an angle and that a curve is a semicircle are invariant. For axis dilatations, the property that a curve is a circle will not be invariant but the property that a curve is an ellipse (also a hyperbola or parabola) will be invariant. Ratios of lengths of intervals that lie on the same line (but not on different ones) are preserved, as are ratios of areas. It follows that the properties that a point divides an interval in a given ratio and that a line is the median of a triangle are invariant. One can show that every affine transformation can be represented as a composition of displacements and axis dilatations, and therefore all the above properties are invariant under affine transformations (an example of a projective invariant, the cross-ratio, was given on p. 314).

The introduction of invariants was only the outer layer of geometry. Its fundamental content consisted of *theorems on the relations between invariance properties*. (These relations are called *syzygies*.) For example, the theorem that the medians of a triangle meet at a single point that divides them in

the ratio 1:2 is based on affine invariants: to be a point of intersection of lines, to divide the line segments in a given ratio, and to be the median of a triangle. Precisely because of this, if it is valid for one triangle, then it is valid for its image under an affine transformation and therefore it suffices to verify it for one triangle, e.g., an equilateral triangle. (We can transform any triangle to any other using affine transformations.) The dependence between homothetic invariants appears in theorems on the intersection of bisectors and altitudes.

In 1837 Michel Chasles (1793–1880) turned his attention to using geometric transformations to obtain new theorems: "Now anyone is in a position to take any well-known truth and apply various general principles of transformations to it; thus he obtains other truths.... Genius is no longer needed to play one's part in constructing the majestic temple of science." However, if we understand Chasles' recipe literally—to take any theorem and apply an arbitrary transformation to it—then we obtain a true assertion but one that is stated so clumsily that it has little chance of being in the "temple of science." Think, for example, of what the theorem on the intersection of bisectors would become if we applied an axis dilatation. How can we explain what line will be the result of transforming a bisector? Klein explained that, on the contrary, it was important to understand which transformations do not change an assertion, to select a transformation that simplifies the picture as much as possible, and to prove the assertion in the simpler form. Here is a traditional example. Any triangle can be transformed into an equilateral one by an affine transformation, and since the theorem on the intersection of the medians deals with a relation between affine invariants it suffices to verify this theorem for equilateral triangles, which in turn is very simple to do.

These considerations allow us to make Chasles' recipe more precise. Imagine some relation between affine invariants in an equilateral triangle of unit area. For example, let λ be the area of the hexagon formed by the "tridians," the lines joining the vertices of a triangle with the points that trisect the opposite sides. Then in *any* triangle the ratio of the area of the hexagon formed by the tridians to the area of the entire triangle equals λ. Now you can easily think of other theorems of this kind.

One of the most important moments in Klein's reasoning is the explanation of the interrelation between the geometries associated with groups G_1 and G_2 when $G_1 \subset G_2$. (We say that G_1 is a subgroup of G_2.) The larger group G_2 has fewer invariants than G_1, and all theorems about G_2 are also valid for the geometry associated with the smaller group G_1.

Therefore, in each specific geometry it is important to find those statements that remain valid for geometries with larger transformation groups.

Sometimes the possibility of "carrying over" statements to a geometry with a larger transformation group becomes clear only after reformulating the statements.

In the language of the Erlanger program, Cayley's ideology consists of saying that one can move in the reverse direction by considering a group of transformations that preserve some fixed object. Here we can often construct invariants of a subgroup using invariants of the group (distances in Euclidean and non-Euclidean geometry using the cross-ratio).

It is usually simpler to describe invariants for a larger group and the relations among them. In particular, the problem of finding invariants for the projective group can be completely solved algebraically.

The "Erlanger program" brought the "golden century" of classical geometry to a close. The number of new geometries increased and geometrical language gradually penetrated into a significant part of mathematics. "Classical geometry outgrew itself and changed from a living, self-sufficient science into a universal language of modern mathematics with a significant amount of flexibility and convenience" (Bourbaki).

An External Student in Riemann's School

After the "Erlanger program," Klein turned to the theory of algebraic functions, an area in which Gauss, Legendre, Abel, Jacobi, Weierstrass, and Riemann had worked. The ideas of Bernhard Riemann (1826–1866), whom Klein did not know personally, turned out to be the closest to his own. In Klein's words, he was an "external student in Riemann's school, . . . and, as we know, if external students take on some task then they work especially zealously because only a deep interest motivates them." Klein wrote later that he saw his problem in a combination of Riemann and Galois—that is, in bringing group theory into the geometric theory of functions of a complex variable. In Klein's own opinion, this was the main area of his scientific work.

Unfortunately, we cannot talk much about this aspect of Klein's work because it would require special knowledge of the reader that goes beyond the usual curriculum. But all the same, we will recall one event.

Klein studied the so-called *uniformization problem*. He considered some important special cases and hoped with time to solve the general problem. But in 1881 Klein discovered a series of articles by a French mathematician who was unknown to him, Henri Poincaré (1854–1912), who had essentially solved the uniformization problem.[6] Klein took this dramatic event

[6]The general uniformization problem was one of the problems in Hilbert's 1900 list and

appropriately. He began a correspondence with Poincaré, and they exchanged 26 letters. Klein, who was already a well-known mathematician but was only 5 years older than Poincaré, took on the role of a very tactful teacher. He acquainted Poincaré with Riemann's theory, which Poincaré did not know but mastered instantaneously. Klein decided to compete with Poincaré by improving his proof of the basic result and noting its generalization. This story ended sadly for Klein: "The price that I had to pay for my work was in any case very great, since my health turned out to be completely shattered. . . . The situation improved somewhat only in the autumn of 1884, but I was never able to attain my previous level of creative activity. . . . My own creative work in the area of theoretical mathematics ended in 1882."

The Last Forty Years

After 1886, Klein worked in Göttingen. Thanks to Klein this city turned into the real capital of mathematics. At his initiative talented young mathematicians were invited to Göttingen, Hilbert among them. Klein never stopped being interested in new ideas. His lecture courses, which were in part written up and published, were devoted to the most varied fields of mathematics, mechanics, and physics. Klein's work in and public activities organizations was multifaceted. For 50 years he directed the publication of one of the key mathematics journals, *Mathematische Annalen*. His *Lectures on the Development of Mathematics in the Nineteenth Century*, delivered during 1914–1919 and published posthumously (he died in 1925) by his students Richard Courant (1888–1972) and Otto Neugebauer (1899–1990), were particularly graceful. Here is an extract from the preface: "These lectures are the ripe fruit of a rich life, lived at the center of scientific events, an expression of penetrating wisdom and deep historical understanding, of great humanitarian culture and of a masterful gift of exposition."

Klein spent a significant part of his time and energy on the problem of teaching mathematics in the schools and on preparing teachers, which probably no mathematician of his stature had done before. "There is hardly a subject," wrote Klein, "where such routine in education is the rule as in mathematics education. A course in elementary mathematics takes the form of defining a framework and precisely measuring its limits once and for all. From time to time in some way or other a few problems are replaced by others, some paragraphs are removed and others are introduced; but essentially this is almost not reflected at all in the material of school mathematics. New algebra textbooks carry the imprint of Euler's algebra, as

was completely solved in 1907, independently by Poincaré and Paul Koebe (1882–1945).

new geometry texts carry that of Legendre. One might think that mathematics is a dead science, that nothing changes in it, and that there are no new ideas in this field of knowledge, at least none that could be adopted by nonspecialists or be the subject of general education."

In teaching, Klein strove to take into account the state of modern science and the connections between mathematics and physics. He recommended using transformations in geometry systematically and rejected the traditional division of school mathematics into subjects. The school curriculum should be imbued with the concept of function, and ways of teaching pupils "functional thinking" should be thought through carefully. The presentation of geometry, in Klein's opinion, should not begin axiomatically, and the axiomatic method should make its appearance only when pupils are prepared to understand it.

With the greatest tact Klein maintained contact with those who studied school mathematics, staying clearly within the bounds of his competence and never entering into questions that required the direct experience of working in schools. Klein gave lectures for teachers, which were published in part. The most well known is his *Elementary Mathematics from an Advanced Standpoint*. This is not a lecture on mathematical methods and is not a comprehensive course of school mathematics. In the preface to this book Klein writes, "I prefer to close with the wish that the present lithographed volume may prove useful by inducing many of the teachers of our higher schools to renewed use of independent thought in determining the best way of presenting the material of instruction. This book is designed solely as such a mental spur, not as a detailed handbook. The preparation of the latter I leave to those actively engaged in the schools. It is an error to assume, as some appear to have done, that my activity has ever had any other purpose."[7]

[7] Felix Klein, *Elementary Mathematics from an Advanced Standpoint: Arithmetics, Algebra, Analysis*, translated by E. R. Hedrick and C. A. Noble, MacMillan, New York, 1932, p. IV.

The Magic World of Henri Poincaré

> I described a imaginary world whose inhabitants inevitably had to come to create the geometry of Lobachevsky. H. Poincaré

When the history of hyperbolic geometry is discussed today, one can get the impression that had the creators of non-Euclidean geometry proved its consistency that would have been favorably received. But it was not the lack of proof that disturbed the critics above all. People were used to having geometry deal with our real space and having this space described by Euclidean geometry. It was characteristic that Gauss distinguished geometry from the other branches of mathematics, considering it similar to mechanics in experimental science. But here Gauss, as well as Lobachevsky and Bolyai, understood that first of all, logical, orderly geometric constructions that have no physical reality are possible—"imaginary" geometries—and second, that it is not so unquestionable that on an astronomical scale Euclidean geometry governs our world. However, what was understood by only a few mathematicians was absolutely inaccessible to nonprofessionals. They measured the claims of hyperbolic geometry with the Euclidean ruler of their own geometric intuition—and came up with an inexhaustible source for their wit. Nikolai Chernyshevsky wrote to his sons from exile that the entire city of Kazan was laughing at Lobachevsky: "What is the 'curvature of a ray' or 'curved space'? What is geometry without the axiom of parallel lines?" He compared this to "squaring a boot" and "extracting the roots of a boot-top," and said that it was as ridiculous as "writing Russian without verbs." The same thing happened to the poet Afanasy Fet ("Whispers, timid respiration, trills

of nightingale"), at whom they also "laughed until their sides hurt."[1]

A new stage in the development of non-Euclidean geometry began when its first models appeared. Now we accept these models as a means of proving that hyperbolic geometry is consistent, but they were remarkable not only for that. Even at a friendly glance hyperbolic geometry seemed much too refined, not connected to the rest of mathematics, but the Cayley–Klein model showed that it arose naturally along the road to projective geometry, which was very popular at the time. On the other hand, considering models whose basic ideas are constructed from the objects of the Euclidean geometry to which we are accustomed made it possible to replace a formal axiomatic presentation of non-Euclidean geometry with something more visual.

Henri Poincaré constructed another model when studying purely analytic questions in the theory of functions of a complex variable. He unexpectedly discovered that the transformations he found could be interpreted as *displacements* in the hyperbolic plane. This discovery made such a strong impression on him that many years later he recalled how it came into his head, "without any, it seemed, previous thoughts," when he was stepping onto a bus during a trip to Coutance. Within ten years Poincaré had made a remarkable addition to his model by giving it a "physical" basis. This chapter is devoted to discussing Poincaré's model.

A Detour into Physics

Our geometric representations have physical premises. For example, we take light rays to be lines. A ray of light that comes to us continues to appear straight even if it is refracted along the way (for example, in going from air to water). In order to clear up this illusion we have to conduct an experiment or examine where it comes from.

Suppose we have an optically heterogeneous medium in the upper half-plane ($y > 0$), in which the speed of light changes according to the law $c(x, y) = y$ (independent of the direction of the light). It follows from Fermat's principle that the path along which light propagates between two points is the path along which it requires the least time to travel. In our medium (where $c(x, y) = y$) light will propagate along the curves L (see

[1]This incident is described by Vladimir Nabokov in his novel *The Gift*. Chernyshevsky was a 19th century socialist reformer and radical journalist who was exiled to Siberia. Lobachevsky's professorial position was at the University of Kazan, where he was Rector. Fet wrote a poem without verbs that Chernyshevsky quoted here.—*Transl.*

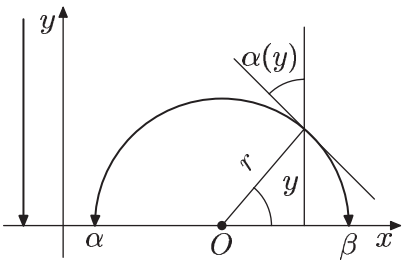


Figure 1.

Figure 1) for which

$$\frac{\sin\alpha(y)}{y} = k, \tag{1}$$

where $\alpha(y)$ is the angle that the tangent to L at the point with ordinate y makes with the vertical and k is a number that is fixed for all points of L. It is clear that all circles centered on the x-axis (i.e., that meet this axis in a right angle) satisfy equation (1); for each such circle $k = \frac{1}{r}$, where r is the radius. For $k = 0$ we obtain vertical lines. One can show that there are no other curves that satisfy (1) and this has a physical explanation (for example, light propagates from a given point in a given direction along a unique path).

Circles orthogonal to the x-axis and vertical lines (more correctly, their intersections with the upper half-plane) will play a major role in our story.

"Poincaria" and Its Geometry

Poincaré's world, which we call *Poincaria* in his honor, is the upper half-plane $\{(x, y), y > 0\}$ without its boundary $\{y = 0\}$ (this is important!).[2] The *Poincarians* who inhabit Poincaria take as their "lines" the upper semicircles centered on the x-axis (without endpoints!) and vertical rays (see Figure 2). We will call these lines P-lines. P-lines appear to the Poincarians as infinite (light propagates along them infinitely long) and their endpoints are invisible, as is the entire x-axis. Thus the Poincarians think that their Poincaria is unbounded in all directions. We will call the invisible points of a P-line its *points at infinity*; for a ray we will take one of its points at infinity as ∞. A P-line is uniquely determined by its pair of points at infinity (why?).

[2]We could have considered a "three-dimensional" world but it is easier draw pictures in the plane and because of this we will work with objects there.

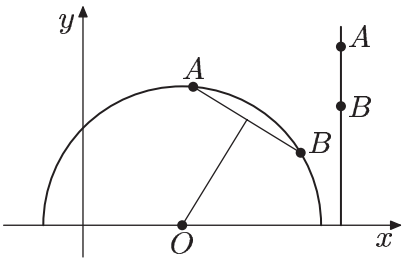


Figure 2.

Thus we will denote a P-line by $L(\alpha, \beta)$, where α, β are real numbers that are the coordinates of its points at infinity on the x-axis (one may be ∞).

Let us try along with the Poincarians to construct the geometry of their space. Just as for us in the Euclidean space where we live, some statements (axioms) seem obvious to the Poincarians and they take them without proof and derive more complicated assertions (theorems) from them. To us, looking at Poincaria from the outside, all these statements appear differently than they do to the Poincarians (for example, P-lines for us are semicircles or rays), and so we will "translate" the Poincarians' statements into our own "prosaic" Euclidean language and prove them ourselves.

For example, the Poincarians know that through two distinct points there is a P-line that is in fact unique. For us this means that through two distinct points in the upper half-plane there is a unique semicircle that is orthogonal to the x-axis or there is a vertical ray (prove it!); see Figure 2. We note that the physical explanation of this statement, that light propagates between two points along a unique path, is the same for Poincarians as for us (in geometry this explanation does not constitute a proof). It is not hard to see that all the axioms of Euclidean geometry that deal with the relative positions of points and lines and the order of points along a line are valid in Poincaria. (In order to get used to Poincaria examine the P-line segments and P-half-planes into which a P-line divides it. For a bounded P-line, the P-segments that connect points in one P-half-plane do not intersect the P-line and those that join points in different P-half-planes do intersect it. Draw P-triangles and P-polygons, and think about P-convex sets if you know about "ordinary" ones. Figure 3 will help.)

The difference between the geometry of Poincaria and Euclidean geometry becomes apparent when we consider the relative positions of two P-lines. We already know that two distinct P-lines can intersect in no more than one point. If they do not intersect at all, then they have a common (invisible!) point at infinity or they do not even have common points on

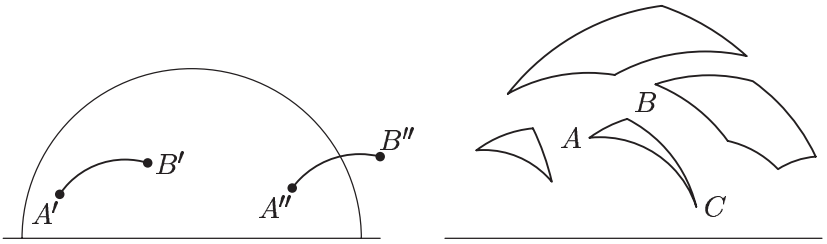


Figure 3.

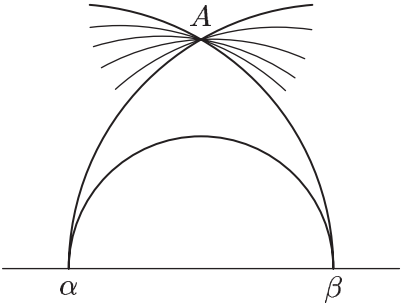


Figure 4.

the invisible boundary. In the first case we will say the P-lines are parallel and the second case that they are superparallel. If we have a P-line $L(\alpha, \beta)$, then through a point outside it there can pass only two P-lines parallel to $L(\alpha, \beta)$ (corresponding to the points at infinity α and β, respectively; see Figure 4) and an uncountable set of superparallel P-lines that lie between the parallel ones. Thus the parallel axiom is not valid in Poincaria. (This does not surprise us, the observers, since the Poincarians do not know that their "lines" are not "real"!) This allows us to hope that the geometry of Poincaria will turn out to be hyperbolic.

The main thing we need to do now is to define *distance* and *displacement* in Poincaria.

Distance and Displacement

From the point of view of optics the most natural quantity to take as the distance between two points A and B in Poincaria is the time it takes for light to travel from A to B: then P-lines will be the shortest lines between two points that lie on them. It follows from physical considerations that

the distance $\rho(A, B)$ defined this way has the usual properties of Euclidean distance:

1. $\rho(A, B) = \rho(B, A)$.
2. If A, B, C lie on the same P-line and $B \in [AC]$, then $\rho(A, B)+\rho(B, C) = \rho(A, C)$ (light propagates from A to C along the P-line and passes through B).
3. For any points A, B, C, we have $\rho(A, B) + \rho(B, C) \geq \rho(A, C)$, the *triangle inequality*, and equality holds only when $B \in [AC]$. (If this inequality were not valid, then light would need less time to travel along the polygonal P-line ABC than along the P-line AC, which cannot be since AC is the quickest path.)

For Poincarians the distance ρ we have introduced is primary (note that relative to this distance light propagates with unit speed), and there is no reason for them to express ρ in any other way; for us it is natural to express ρ in terms of our Euclidean distance. This is not simple; we must deal with nonuniform movement of light, and to compute the time that light takes we must calculate integrals. Therefore, we will only give the final answer:

$$\rho(A, B) = \ln \frac{r' + r}{r' - r}, \tag{2}$$

where r is the Euclidean distance between the points A and B and r' is the Euclidean distance between A and the point B' symmetric to B about the x-axis. The logarithm is taken to base e (for any other base we obtain ρ up to a constant multiplier). The Euclidean distance is notable for being preserved by many transformations of the plane, and such transformations are called *displacements*. Let us see what a displacement looks like in Poincaria, i.e., a P-displacement, which is a transformation that preserves ρ and therefore maps P-lines to P-lines.

We begin with transformations that do not leave any point in place. First of all, there are the usual *parallel translations* along the x axis, $T_a(x, y) = (x + a, y)$. Parallel translations preserve the Euclidean distance, preserve the speed of light $c(x, y) = y$ and thus also the time taken by light on the path between two points A and B, and finally they map P-lines to P-lines. On the other hand, *homotheties* $F_b(x, y) = (bx, by), b > 0$, which change the Euclidean distance and therefore the magnitude of the speed of light $c(x, y)$ proportionately, do not change the time spent by light, i.e., the P-distance $\rho(A, B)$. Thus what seems to us to be a homothety (centered on the x-axis) appears to Poincarians as a displacement. Using the P-displacements we

have indicated we can map any point to any point. For example, the point (x_0, y_0) maps to $(0, 1)$ under the P-displacement $(\frac{x-x_0}{y_0}, \frac{y}{y_0})$. Relative to the P-displacements we have introduced, which we will call *P-translations*, P-lines fall into two classes: we can map semicircles into one another and, separately, we can map rays into one another (why?).

Let us now explain how we can use P-translations and the properties of the P-distance ρ to find a simple derivation of equation (2), which expresses ρ in terms of Euclidean distance, in the special case where both points A and B are on the y-axis: $A = (0, y_1), B = (0, y_2)$. We set $\rho(A, B) = \varphi(y_1, y_2)$ and find the form of the function φ. Since ρ is preserved by Euclidean homotheties with center O,

$$\varphi(by_1, by_2) = \varphi(y_1, y_2). \tag{3}$$

Moreover, if $C = (0, y_3)$ is a third point on the y-axis, then by what we said above,

$$\varphi(y_1, y_2) + \varphi(y_2, y_3) = \varphi(y_1, y_3). \tag{4}$$

Let $\psi(z) = \varphi(z, 1)$. By (3),

$$\varphi(y_1, y_2) = \psi\left(\frac{y_1}{y_2}\right) = \psi(z_1),$$
$$\varphi(y_2, y_3) = \psi\left(\frac{y_2}{y_3}\right) = \psi(z_2),$$
$$\varphi(y_1, y_3) = \psi\left(\frac{y_1}{y_3}\right) = \psi(z_3).$$

Taking (4) and the last three equations into account, we obtain

$$\psi(z_1 \cdot z_2) = \psi(z_1) + \psi(z_2),$$

from which, assuming that ψ is a sufficiently "good" function with positive values, we obtain that $\psi(z) = k \cdot \ln|z|$, where k is a constant factor that can be calculated directly.

The P-displacements we have found are not enough: we have no transformations with which we can map P-lines of one kind (semicircles) into P-lines of the other kind (rays). For this we add *P-symmetries* with respect to P-lines. For rays this is the usual *axial symmetry* and for semicircles it is an *inversion*. (For example, the P-symmetry with respect to the P-line $L(-1, 1)$ is the inversion with respect to the circle of radius 1 centered at $O = (0, 0)$; it maps a point A other than the center O to the point A' lying on the ray OA for which $|OA| \cdot |OA'| = 1$.) We know that under an inversion circles and lines map to circles and lines, and that the magnitudes

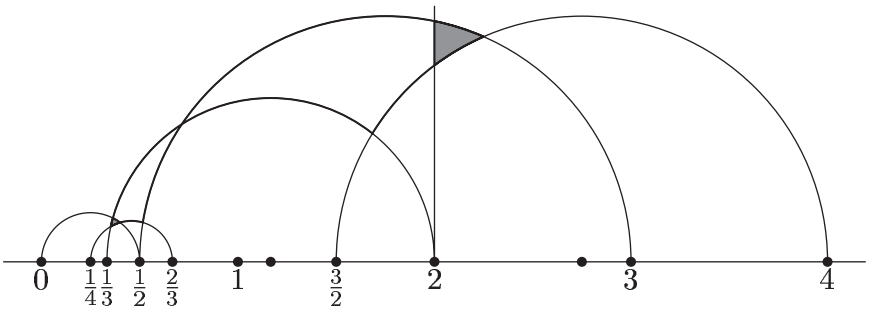
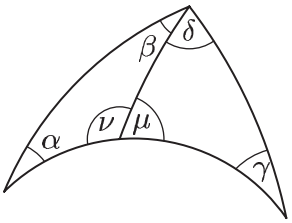


Figure 5.

of angles are preserved. In the language of Poincaria this means that, for example, under a P-symmetry with respect to the P-line $L(-1, 1)$ the P-line $L(\alpha, \beta)$ maps to the P-line $L(\frac{1}{\alpha}, \frac{1}{\beta})$. In particular, the P-lines $L(\alpha, 0)$, which are semicircles for $\alpha \neq \infty$, map to the P-lines $L(\frac{1}{\alpha}, \infty)$, which are rays. Thus P-symmetries map Poincaria to itself and P-lines map to P-lines. It is verified separately (we will omit the verification) that P-symmetries do not change the P-distance ρ. Incidentally, in Poincaria every transformation that maps P-lines to P-lines preserves ρ (here there are no homotheties). This is the most importance difference between hyperbolic and Euclidean geometry.

The P-displacements that we can obtain by combining P-symmetries with P-translations are enough to map any P-line into any P-line; moreover, any given point of the first P-line can be combined with any given point of the second, and any P-ray with another P-ray (prove it!). This means that using P-displacements we can combine any P-segments of equal P-length, and we find that such segments are P-congruent. One can show that all P-displacements reduce to the ones we have described.

Under P-displacements an angle maps to an angle that is equal to it in the Euclidean sense (since this is so for parallel shifts, homotheties, axial symmetries, and inversions). Therefore, the concept of equality of angles in Poincaria does not differ from the Euclidean concept. Taking this situation into account the Poincarians prove, just as we do, the validity of these two tests, that triangles are congruent: using two sides and the angle between them and using a side and the two angles adjacent to it. The proof of a third test, using three sides, is more complicated: after all, our proof of this uses the fact that circles intersect in no more than two points. Fortunately, it turns out that P-circles are the same as Euclidean ones (that lie completely in the upper half-plane), but their P-centers do not coincide with the usual



$$\pi - (\alpha + \beta + \gamma) =$$
$$[\pi - (\alpha + \beta + \nu)] + \pi - (\mu + \delta + \gamma)$$

Figure 6.

ones (this is not so simple), and therefore as far as the third test goes everything is fine in Poincaria. However, in Poincaria there is still another test for congruent triangles: *triangles with pairwise equal angles are congruent*! (Translate this statement into the language of Euclidean geometry and try to prove it.) This means that the area of a triangle in Poincaria (as well as the triangle itself) is determined by the magnitudes of its angles α, β, and γ. In hyperbolic geometry the sum of the angles of a triangle is less than π. The quantity $\pi - (\alpha + \beta + \gamma)$ is called the defect of the triangle. We can note that the defect of a triangle behaves the same way as the area. More precisely, if a given triangle is cut by a line through its vertex, then its area will be equal to the sum of the areas of the triangles we obtain. The same is true for the defect of every triangle; it equals the sum of the defects of the triangles that are formed this way (see Figure 6). We can conclude from this that the area of a triangle in hyperbolic geometry is proportional to the defect $\pi - (\alpha + \beta + \gamma)$.

Some Problems

1. (a) Convince yourself that all P-lines orthogonal to a fixed P-line are superparallel (see Figure 7).
   (b) Show that every pair of superparallels has a unique P-perpendicular in common (see Figure 8(a)–(b)).

2. Verify that the P-bisectors of a P-triangle intersect at a single point, namely the center of the inscribed P-circle. Think about what you can say about the circumscribed circle—does it always exist? (See Figure 9. In this figure, the P-triangles A_iBC_i are isosceles, with axis of symmetry $L(0, \infty)$, $i = 1, 2, 3$. The perpendiculars to the P-midpoints of the sides of the triangles are shown in the figure.)

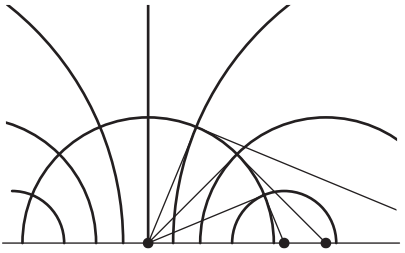


Figure 7.

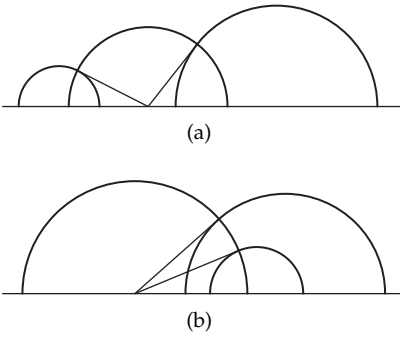


Figure 8.

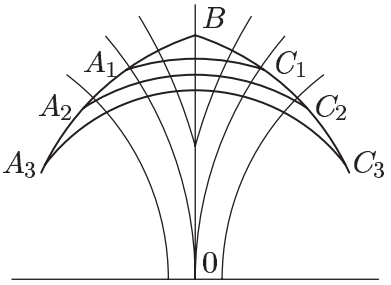


Figure 9.

3. Persuade yourself that the altitudes of an obtuse (but not acute) *P*-triangle can be superparallel (Figure 10). What can you say about the medians?

4. Show that in an isosceles *P*-triangle the base angles are equal and the bisector of the vertex angle is the median and altitude. Prove a fourth test for *P*-congruence for this case.

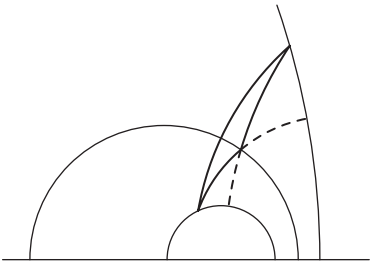


Figure 10.

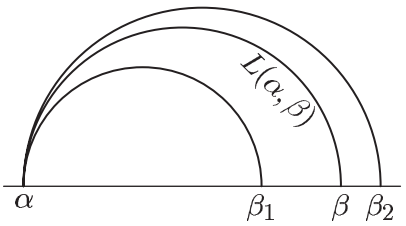


Figure 11.

5. Let $L(\alpha, \beta)$, $L(\alpha, \beta_1)$, $L(\alpha, \beta_2)$ be three parallel P-lines (see Figure 11). Prove there exists a P-displacement that maps $L(\alpha, \beta)$ to itself and $L(\alpha, \beta_1)$ to $L(\alpha, \beta_2)$.

   From this it follows that in hyperbolic geometry we cannot define the distance between parallels.

6. If a P-line L_1 intersects a P-line L_0 or is superparallel to it, then it projects onto L_0 in the form of a finite P-interval. If L_1 is parallel to L_0, then the projection is a P-ray.

7. Let L_0 be a P-line perpendicular to L_1 and let A be a point of L_0 at a distance x from L_1 (see Figure 12). Pass a P-line M_x through A, parallel to L_1, and let $\varphi(x)$ denote the magnitude of the angle formed by M_x with L_0. Find $\varphi(x)$ and show that $\varphi(x) \to \frac{\pi}{2}$ as $x \to 0$ and $\varphi(x) \to 0$ as $x \to \infty$.

The function $\varphi(x)$ is called the *Lobachevsky function*. It relates magnitudes of angles to lengths, and since there is an absolute unit of measurement for angles (the full angle of 2π) there is one in hyperbolic geometry for lengths. (It carries over from angles using the function φ.) In Euclidean

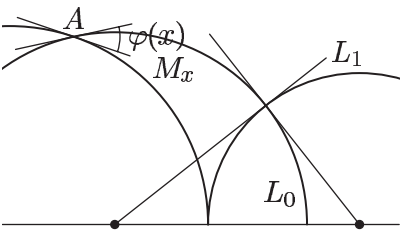


Figure 12.

geometry $\varphi(x) \equiv \frac{\pi}{2}$, and thus there is no analogous absolute unit of measurement for length.

Rigid Bodies in Poincaria

So far we have been guided only by optical premises in our geometric considerations. Here we must stress that Poincaré's geometry is non-Euclidean not because the Poincarians have different optics laws than we do: we construct (model) Poincaria in our own world and the laws of physics do not change! The optical illusions of the Poincarians are explained by the optical inhomogeneity of their world.

Although the most striking realization of a straight line is no doubt a light ray, we still cannot measure length using the time it takes for light to propagate—for this we have no ruler. It is probably worth giving the Poincarians a ruler. Of course, the Poincarians make a ruler from a "P-line" but if they take this ruler from one place to another, it will no longer look "straight." From the viewpoint of a Poincarian a rigid body can change its shape when it moves. How should a Poincarian react to this? It is clear that we must link the concept of a rigid body to the geometry of Poincaria, or else the Poincarians will have to believe in the existence of supernatural forces. Henri Poincaré thought of a clever way out of this seemingly hopeless situation: *he used the phenomenon that bodies expand with heat*. Suppose all bodies in Poincaria have the same coefficient of thermal expansion and zero thermal conductivity, and that the measurements of bodies are proportional to the absolute temperature T. We note that under these conditions Poincarians cannot measure temperature with an ordinary thermometer, since such a measurement presupposes comparing the expansion of bodies with different coefficients of expansion. A rigid body is characterized by the fact that when it moves in a medium of constant temperature, the distance $r(A, B)$ (Euclidean) between any two of its points A and B is pre-

served. But if a body is displaced from a domain with temperature T_1 to a domain with temperature T_2, then the distance between the points is multiplied by T_2/T_1 (in other words, the ratio $r(A, B)/T$ remains valid). And what will happen if the body suddenly finds itself in a domain with different temperatures?

What quantity will be preserved under these conditions? Suppose, for example, that a sufficiently large rigid body is displaced in a medium where on one side of some line m the temperature is T_1 and on the other side it is T_2. Let A and B be two points of the body where the temperatures are T_1 and T_2, respectively. We take a polygonal line with endpoints at A and B and a vertex C on the line m. Let $|AC| = r_1, |CB| = r_2$ and consider the quantity $\frac{r_1}{T_1} + \frac{r_2}{T_2}$. It turns out that for motion in a medium with these temperatures the minimum value of $\frac{r_1}{T_1} + \frac{r_2}{T_2}$ is preserved, taken over all polygonal lines with vertices on m and with endpoints at the two given points A and B. Furthermore, we can repeat exactly the same argument in applying Fermat's principle, for example, to deduce Snell's law of refraction and obtain that the smallest value we are seeking will correspond to the polygonal line for which $\frac{\sin\alpha_1}{T_1} = \frac{\sin\alpha_2}{T_2}$, where α_i is the angle between the corresponding section of the polygonal line and the normal to the line m.

Now suppose that at a point (x, y) in Poincaria an absolute temperature of $T(x, y) = y$ is constantly maintained. Because of the temperature regime chosen, when rigid bodies (in our sense) move, the P-distance rather than the Euclidean distance will be preserved, and from the point of view of a Poincarian (they do not sense temperature differences) the size of a body moving in such a medium will be preserved, i.e., it is P-rigid. It remains only to check that all objects have small heat capacity and are displaced slowly enough to be in thermal equilibrium, and that temperature changes are not noticeable to a Poincarian. In the end the Poincarians not only do not see the boundary of their world but they can never reach it; as they approach the boundary the temperature approaches absolute zero, and therefore the size of every object also approaches zero without changing the proportions between objects. Henri Poincaré tried to eliminate any chance that the Poincarians would know that their non-Euclidean world was completely constructed within our Euclidean one. But did he foresee everything?

The Enigma of Ramanujan

> Ramanujan used to say that the goddess of Namakkal inspired him with the formulae in dreams. It is a remarkable fact that frequently, on rising from bed, he would note down results and rapidly verify them, though he was not always able to supply a rigorous proof. P. V. Seshu Ayar and R. Ramachandra Rao[1]

A Letter to Cambridge

At the very beginning of 1913, Professor Godfrey H. Hardy of Cambridge University received a letter from far-away Madras, in India. At the age of 36, Hardy (1877–1947) was already one of the leading specialists in analysis and number theory and had written a series of excellent mathematical works. The sender of the letter, Srinivasa Ramanujan, worked as a clerk in the accounts section of the Madras Port Trust at the paltry salary of 20 pounds a year. He wrote about himself that he had no university education and had studied mathematics on his own after finishing school, not according to the accepted system but "striking out a new path for myself."[2] The mathematical content of the letter looked awkward enough—the author could certainly be taken to be a self-confident amateur.

On its own such a letter could not have made a big impression on Hardy. But the letter included some formulas that the author proposed to publish, if they were interesting, something he himself could not do because of his poverty. A look at the formulas made Hardy sit up and

[1] Ayar and Rao, "Srinivasa Ramanujan (1887–1920)," from *Collected Papers of Srinivasa Ramanujan*, Cambridge University Press, Cambridge, UK, 1927.

[2] Hardy includes Ramanujan's letter in his obituary, "Srinivasa Ramanujan (1887–1920)," *Proc. London Math. Soc.* (2), **19** (1921), pp. xl–lviii.—*Transl.*

Insert 1. *An example of an infinite sum that Ramanujan computed.*

$$1-5\left(\frac{1}{2}\right)^3+9\left(\frac{1\cdot 3}{2\cdot 4}\right)^3-13\left(\frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\right)^3+\cdots=\frac{2}{\pi}.$$

This is a surprising formula and was one of those Ramanujan attached to his first letter to Hardy. For a long time Hardy could not understand how the sum of the alternating series $a_0+a_1+a_2+\cdots$ with the general term

$$a_n=(-1)^n(4n+1)\left(\frac{1\cdot 3\cdot 5\cdot\cdots\cdot(2n-1)}{2\cdot 4\cdot 6\cdot\cdots\cdot 2n}\right)^3$$

suddenly turns out to equal $\frac{2}{\pi}$. The reader can use a calculator to see that the formula is valid as an approximation, but the proof that it is exactly true is not elementary.

take notice: he understood that this was something out of the ordinary. He answered Ramanujan with interest and an intense correspondence sprang up between them. Gradually Hardy collected about 120 diverse formulas.

Ramanujan's formulas basically involved relations between infinite radicals (Insert 2), infinite series, products, and continued fractions (Inserts 1, 3, and 4), and identities for integrals. Above all it was clear that they went far beyond the limits of elementary mathematics. Also, they raised a series of questions. Were they known? If so, did the author of the letter obtain them on his own; if not, were they true? Hardy soon understood that the situation was paradoxical. He, who was without a doubt the leading specialist in modern analysis, was dealing with a gold-mine of formulas he did not know!

The formulas with infinite series greatly impressed Hardy (see Insert 1). After studying them, he concluded: "...Ramanujan must possess much more general theorems and was keeping a great deal up his sleeve."[3]

The relations with infinite continued fractions especially surprised Hardy (one of the later relations of this kind is shown in Insert 3): "...[these relations] defeated me completely; I had never seen anything in the least like them before. A single look at them is enough to show that they could only be written down by a mathematician of the highest class."[4]

[3]G. H. Hardy, The Indian Mathematician Ramanujan, in *Ramanujan:* 12 *Lectures on Subjects Suggested by His Life and Work*, Cambridge University Press, Cambridge, UK, 1940, p. 9.
[4]Ibid.

Insert 2. *Infinitely repeating radicals.*

$$\sqrt{1+2\sqrt{1+3\sqrt{1+4\sqrt{1+\cdots}}}}=3.$$

Ramanujan obtained this beautiful formula while he was still of school age, in the following way: he wrote down a sequence of obvious equations,

$$n(n+2)=n\sqrt{1+(n+1)(n+3)}=n\sqrt{1+(n+1)\sqrt{1+(n+2)(n+4)}}=\cdots,$$

and set $n=1$. Ramanujan was not interested in the legality of passing to the limit. Operating the same way, the reader can try to derive the similar formula

$$\sqrt{6+2\sqrt{7+3\sqrt{8+4\sqrt{9+\cdots}}}}=4.$$

Insert 3. *A numerical identity with an infinite sum and a continued fraction.*

$$1+\frac{1}{1\cdot 3}+\frac{1}{1\cdot 3\cdot 5}+\frac{1}{1\cdot 3\cdot 5\cdot 7}+\frac{1}{1\cdot 3\cdot 5\cdot 7\cdot 9}+\cdots + \frac{1}{1+\cfrac{1}{1+\cfrac{2}{1+\cfrac{3}{1+\cfrac{4}{1+\cdots}}}}}=\sqrt{\frac{\pi e}{2}}.$$

This is perhaps Ramanujan's most beautiful formula, a true product of the mathematical art. It unexpectedly relates an infinite series and an infinite continued fraction. It is surprising that neither the series nor the continued fraction is expressed in terms of the constants π and e, but their sum turns out to equal $\sqrt{\frac{\pi e}{2}}$ in some incomprehensible way!

The Wonder from Kumbakonam

How was this mathematician formed, who so surprised Hardy? Srinivasa Ramanujan Iyengar was born on December 22, 1887 in southern India, in the village of Erode. His childhood was mostly spent in the small town of Kumbakonam, 260 kilometers from Madras, where his father worked as an accountant in a small textile shop. Ramanujan belonged to the Brahmin caste, but it had been a long time since his family had been wealthy. His parents, especially his mother, were deeply religious. Ramanujan was

Insert 4. *The Rogers–Ramanujan identity.*

$$1+\frac{x}{1-x}+\frac{x^4}{(1-x)(1-x^2)}+\frac{x^9}{(1-x)(1-x^2)(1-x^3)}+\cdots$$
$$=\frac{1}{(1-x)(1-x^6)(1-x^{11})\cdot\cdots\cdot(1-x^4)(1-x^9)(1-x^{14})\cdots}.$$

Ramanujan found this identity in 1911 but was not able to prove it. Hardy could not prove it either. In 1917, while looking at the journal literature (which he rarely did), Ramanujan came across an 1894 article by the English mathematician Leonard Rogers (1862–1933) that had remained unnoticed and in which this formula was proved. Moreover, it turned out that this identity was closely related to the number $p(n)$ of partitions of an integer n into summands (see Insert 5). And it had recently appeared in research in statistical physics.

Insert 5. *The Hardy–Ramanujan theorem.*

This theorem gives an estimate of the number $p(n)$ of partitions of a natural number n into integer summands. For example, $p(5)=7$, since $5=4+1=3+2=3+1+1=2+2+1=2+1+1+1=1+1+1+1+1$. The approximation is

$$p(n)\approx A_n e^{\pi\sqrt{\frac{2}{3}\left(n-\frac{1}{24}\right)}},$$

where $A_n=\frac{1}{2\pi\sqrt{2}}\left(\frac{\pi}{\sqrt{6}(n-\frac{1}{24})}-\frac{1}{2(n-\frac{1}{24})^{3/2}}\right)$ is a function of n. The most mysterious part of the formula for $p(n)$ is the small "correction," $-\frac{1}{24}$, that Ramanujan thought of. No one, not Hardy and not even Ramanujan himself, could explain where it came from. Was this another intervention of the goddess Namakkal? Whether by that way or another, exactly this mysterious correction guaranteed that the estimate was accurate. However, Hardy and Ramanujan did not limit themselves to an approximation formula and later obtained a precise *equation* for $p(n)$. For example, for $n=200$ the above formula gives the very good approximation $p(200)\approx 3,972,998,993,185$ while the exact value is $p(200)=3{,}972{,}999{,}029{,}388$, a relative error of 9.1×10^{-9}.

raised in the traditions of his caste. His childhood, spent in a town where every stone was associated with the ancient religion, surrounded by people who always felt that they belonged to the highest caste, played a large role in his formative years.

At the age of 5, Ramanujan went to school, and he finished primary school when he was 10. He started to show outstanding talent and received a stipend that guaranteed him a high school education at half the cost. At age 14, a student from Madras gave him a two-volume text on trigonometry by Sidney Loney (1860–1939). Ramanujan quickly learned trigonometry and the student was able to use his help in solving problems. The first stories and legends about Ramanujan belong to this period in his life. It is said that he himself discovered "Euler's formula on the sine and cosine" and was very upset to find it in the second volume of Loney's book.

The "young Brahmin" believed that, as in the other sciences, one had to search for the "higher truth" inherent in mathematics and to ask one's teachers. But his elders made unconvincing references to the Pythagorean theorem and to calculations with percents.

"A Synopsis of Elementary Results in Pure and Applied Mathematics"

This two-volume text by the English mathematician George Carr (1837–?), written in 1880–1886, fell into Ramanujan's hands in 1903 when he was 16. This book played an enormous role in Ramanujan's development. It contained a collection of 6165 theorems and formulas, almost without proofs and with minimal explanations. The book was mainly devoted to algebra, trigonometry, analysis, and analytic geometry.

Carr's book motivated the boy to derive the formulas on his own. This is attested to by those who knew Ramanujan in those years. The scope of his main interests gradually changed: magic squares, then quadrature of the circle (he found π with enough precision to calculate the length of the equator with an error of no more than 1–2 meters, according to legend. This was the start of a genuine mathematical life!

Carr's book turned out to be rather successful in shaping Ramanujan's mathematical world. But his orientation towards this book had other consequences. Since the book contained no proofs and at most hints of the arguments, Ramanujan formed a distinctive method of establishing mathematical truths. And he lacked appropriate guidance in India for carrying out strict proofs.

"His ideas of what constituted a mathematical proof were of the most shadowy description. All his results, new or old, right or wrong, had been arrived at by a process of mingled argument, intuition and induction. . . ."[5]

Ramanujan's mathematical destiny was really completely determined

[5]Hardy, "Srinivasa Ramanujan (1887–1920)," p. xxx.

during these years—he never changed the direction of his scientific research or his method of thinking. Here we can express regret that Ramanujan's formative years were spent under difficult conditions. Normally he would doubtless have become a mathematician with the best professional preparation, but can we be certain that he would have been so unique? Would Ramanujan have seen so much if he had been taught from childhood how to act in mathematics and brought his results to publication with strict proofs? Would he have built his mathematical world on the basis of all of human achievement and not on a comparatively small number of facts?

From Numbers to Formulas

In the development of Ramanujan's mathematical world, it was important to combine an initial stock of mathematical facts (basically drawn from Carr's book) with his huge stock of observations about specific numbers. He had collected such facts since childhood. His school friends recalled that Ramanujan knew an enormous number of terms in the decimal expansions of e, π, and other numbers. He possessed an amazing ability to notice arithmetical patterns and patiently examined a huge amount of numerical material—an art in which Euler and Gauss were virtuosi but that had significantly declined by the 20th century. He discovered a lot in the numerical "stockroom" by chance. Hardy later recalled how he visited Ramanujan in the hospital and said that he had taken a taxi with the "dull" number 1729. Ramanujan became excited and exclaimed, "No, it is a very interesting number; it is the smallest number expressible as a sum of two cubes in two different ways."[6] ($1729 = 1^3 + 12^3 = 9^3 + 10^3$. In Hardy's book on Ramanujan's work he says neatly that "every positive integer was one of Ramanujan's personal friends."[7]

Ramanujan quickly filled out the stock of facts drawn from Carr. In so doing he rediscovered results of Euler, Gauss, and Jacobi with surprising speed. This is how the young Gauss in Braunschweig, with no access to the literature, reconstructed in short order what it took his great predecessors decades to do. One can only be surprised that reconstructing mathematics can take place so quickly.

Ramanujan's collecting observations about specific numbers gradually took second place to the world of formulas. Formulas for him were not helpful methods for proofs or calculations but were a goal in themselves.

[6]Hardy, "Srinivasa Ramanujan (1887–1920)," p. xxxv.

[7]This also appears in the obituary just cited. Hardy attributes the comment to John H. Littlewood (1885–1977).—*Transl.*

He infinitely appreciated the internal beauty of formulas. His formulas can be thought of as beautiful paintings.

Choosing a Profession

In 1904 Ramanujan entered the University of Madras and had his first success not only in mathematics but also in the English language. However, he began to focus completely on mathematics and this told on him right away. He did not even complete the first course, wandered through another, made an attempt to return to the university, and finally ended as an external student in 1907 but failed the examination. In 1909 he married; his wife was nine years old. She lived until 1994, touchingly preserving the memory of her great husband. Ramanujan had to think about making a living, but he could not find suitable work. In 1910 he showed his mathematical results to Ramaswari Aiyar, the founder of the Indian Mathematical Society, and then to Seshu Ayar, a teacher at the college in Kumbakonam, and Ramachandra Rao, a prominent bureaucrat who was educated in mathematics and later became Ramanujan's biographer.

Rao helped Ramanujan from his own funds and then arranged for him to become a clerk in the postal service. In 1911 a note about Ramanujan's results by Seshu Ayar appeared in print, and later his own article was published. Influential English officials began to play a role in Ramanujan's destiny. Starting May 1, 1913, he was guaranteed a special stipend of 75 rupees (5 pounds) a month for two years. This was enough for a modest life, and Ramanujan left his career as a clerk. He became a "professional mathematician."

Thus Ramanujan found definite recognition among those around him, but not understanding. Recall that he wrote to Hardy at the beginning of 1913. What did he expect from Hardy? To find, finally, someone who was able to understand and appreciate his results, to help and direct his future research? His reasons were probably more prosaic: he did not need glory and recognition from the outside world, but rather a guarantee that he would be able to exist.

We have to say that Ramanujan chose his addressee exceptionally successfully for his scientific plan. It would have been hard to find another mathematician in the world who could have oriented himself towards Ramanujan's results so quickly and effectively. Hardy understood very soon that what was required of him was not an evaluation of the results of an obscure amateur or junior colleague, but the saving of an enormous talent. At the same time, the thought did not escape him that Ramanujan was telling him only a little of what he knew, that he possessed very general re-

sults and was showing him only particular illustrations. But the important thing was that he could not reconstruct Ramanujan's method and he was eager to learn the path that his surprising correspondent had taken. Unexpectedly Ramanujan steadfastly refused to describe his method. From a letter of February 27, 1913: "...you ask me to communicate the methods of proof.... What I tell you is this. Verify the results I give and if I agree with your results, got by treading the groove in which the present day mathematicians move, you should at least grant me that there may be some truth in my fundamental basis."[8]

Hardy suspected that Ramanujan was afraid that his methods could be appropriated by others and tried to allay his apprehensions, but on April 17 received the answer: "...I am a little pained to see what you have written.... I am not in the least apprehensive of my method being utilized by others. On the contrary my method has been in my possession for the last eight years and I have not found anyone to appreciate the method. As I wrote in my last letter I have found a sympathetic friend in you and I am willing to place unreservedly in your hands what little I have. It was on account of the novelty of the method I have used that I am a little diffident even now to communicate my own way of arriving at the expressions I have already given...."[9]

For Hardy there was no doubt: Ramanujan needed to be in contact with real mathematicians. This could not be guaranteed in India and he had to move to England right away. Negotiations for a stipend in Cambridge were successful. However he had to convince Ramanujan himself, whose current situation was completely settled, that the trip was necessary. Also his mother, whose consent was required for her son, was categorically opposed. Friends tried to mobilize public opinion and the Cambridge mathematician Eric H. Neville (1889–1961) who had visited Madras in early 1914, was actively involved. He turned to the rector of the university for support, but was unsuccessful.

What the scientist could not manage was easily done by the goddess Namakkal (according to legend, Ramanujan learned new formulas from her lips in his dreams). His mother saw her son in a dream, sitting in a large hall in the company of Europeans, and the goddess commanded her not to oppose the trip. On March 17, 1914, Ramanujan left for England. He would receive 250 pounds sterling a year for two years. Of this his mother would receive 50 pounds. Soon after he arrived the stipend was increased to 60 pounds.

[8]Hardy, "Srinivasa Ramanujan (1887–1920)," p. xxvii.
[9]Hardy, "Srinivasa Ramanujan (1887–1920)," pp. xxix–xxx.

In Cambridge

Ramanujan was 27 years old. He had spent his best years doing mathematics in India, without contact with serious scientists and without access to the mathematical literature. In various countries and at various times a person feels complete at various ages. In India at the turn of the 20th century, with a very low life expectancy, 27 was the age of a mature individual. Ramanujan's widow recalled that he liked to compose horoscopes, and that his own horoscope predicted his death before reaching the age of 35.

Hardy was faced with a very crucial decision: should he interrupt Ramanujan's research so that he could master modern mathematics? Hardy's choice was, evidently, the only one possible: not to change the style and direction of Ramanujan's research but to adjust it from the standpoint of modern mathematics as far as possible, trying to explain new things and turning his attention to the appropriate literature. Hardy wrote, "His mind had hardened to some extent, and he never became at all an 'orthodox' mathematician, but he could still learn to do new things, and do them extremely well. It was impossible to teach him systematically, but he gradually absorbed new points of view. In particular he learnt what was meant by proof, and his later papers, while in some ways as odd and individual as ever, read like the works of a well-informed mathematician. His methods and his weapons, however, remained essentially the same."[10]

Ramanujan worked very intensely and productively. He and Hardy had many interests in common. His fantastic intuition, combined with Hardy's refined technique, bore remarkable fruit. Recognition came to Ramanujan: in 1918 he became a professor at Cambridge University and was elected to the Royal Society (the British equivalent of the Academy of Sciences). Never before had an Indian achieved such honors.

Life was not simple for Ramanujan. He strictly followed all his religious restrictions, as he had promised his parents. In particular, he was a vegetarian and had to cook for himself. He refused to break the rules, even when he fell seriously ill in 1917. His irregular diet probably hastened his illness. (Ramanujan himself thought this, as his widow recalled.) Ramanujan spent his remaining two years in England going to hospitals and sanatoria, and was forced to reduce the intensity of his mathematical research.

It was not easy for Ramanujan to enter into life in Cambridge, which was full of conventions and traditions that were alien to him. His natural politeness and inclination not to be a source of discomfort for those around him, characteristic of Indian culture, helped him adapt to university life at

[10]Hardy, *The Indian Mathematician Ramanujan*, p. 10.

least outwardly.

Hardy did a lot for Ramanujan: he followed his research, tried to fill the gaps in his education, and was concerned about his position in society and in life. Up to the last moment Ramanujan was filled with touching gratitude and love for him.

Return and Death

Ill, Ramanujan began to think of returning home. It was only at the beginning of 1919 that his health improved enough for him to take the long trip by sea. A position had been prepared for him at the University of Madras—his fame had reached India. Ramanujan wrote a grateful letter to the rector, apologizing for the fact that recently his illness had not given him the opportunity to work intensively enough. But he could not even start work at the university. Only a little time was left for him to live in his country (and to live at all). After three months in Madras, Ramanujan moved to Kumbakonam. In January, 1920 he sent his last letter to Hardy, telling of his work on a new class of theta functions. Neither the doctor nor his relatives could convince the ill and dying scientist to interrupt his work. Ramanujan died on April 26, 1920. He was not yet 33.

Remembrance

The news of Ramanujan's death shocked his friends in India and England. They felt obliged to understand the astounding phenomenon that he had been. Hardy wrote, "It is possible that the great days of formulae are finished, and that Ramanujan ought to have been born 100 years ago; but he was by far the greatest formalist of his time."[11]

Friends and colleagues tried to evaluate Ramanujan's place in modern mathematics. They had no doubt of his amazing abilities and the fantastic beauty of his formulas, but agreed that the very subjects that Ramanujan persisted in choosing did not allow him to take his rightful place in the history of mathematics.

More than three-quarters of a century has passed, and today we see distinctly what Hardy and his contemporaries could not foresee. Ramanujan's genius turned out to be in harmony not only with the past but also with future mathematics. Ramanujan's arithmetic formulas not infrequently turned out to be the keys to new steps in algebraic number theory, and one can only be amazed at how he could have seen them without knowing

[11] Hardy, *The Indian Mathematician Ramanujan*, p. 14.

what was needed to see anything at all. Also, there is renewed interest in concrete, explicit formulas both within mathematics and in the area of its applications. Modern mathematical and theoretical physics sometimes turn to rather abstract areas of mathematics and very refined and explicit formulas play an important role. Here are two recent examples associated with Ramanujan.

Rodney J. Baxter, who became famous for constructing exactly solved models of statistical mechanics, unexpectedly discovered that the Rogers–Ramanujan identity (Insert 4 on p. 340) constantly came up in the "hard hexagon" model.

Nobel laureate Steven Weinberg recently recalled that in the early 1970s when he was studying string theory, which is very popular now, he ran into the problem of estimating the partition function $p(n)$ for large n. It turned out that Hardy and Ramanujan had obtained the needed formulas in 1918 (Insert 5 on p. 340).

The beauty of Ramanujan's formulas gave him the ability to come to life again under the most unusual circumstances.

On the Advantages of Coordinates and the Art of Chaining Hyperboloids

> In this way we have briefly and clearly set out everything that the ancients left unexplained about planar and solid loci. Pierre Fermat
>
> I maintain that now I have omitted nothing from the basics that are needed to understand curves. René Descartes

The pioneering ideas of the great mathematicians underwent many changes before ending up in the pages of textbooks. In their refined form it is easier to master them and the areas in which they can be applied are clearer, but something that is hard to perceive has been lost. Perhaps this is the logic of their discovery, a feeling for the material, and simply an excitement in face of the possibilities that are opening up. How different is the enthusiasm of the creators of analytic geometry from the feeling of the student who studies it today! We will recall here only a few episodes from the history of the creation of analytic geometry without trying to recreate this history completely, and we will finish the story with an étude in the style of 19th century analytic projective geometry, but with more up-to-date material. It is very tempting to try to argue the way they could a hundred years ago! Namely, we will prove the theorem of five hyperboloids in five-dimensional space. Two two-dimensional hyperboloids of one sheet are said to be chained if they have a common generator that is the line of intersection of the three-dimensional planes they generate. Correspondingly the minimal dimension in which chained hyperboloids exist

is equal to 5. Several hyperboloids are called chained if they are pairwise chained and if the generators of the chainings for each pair belong to the same family. If the generators chaining different pairs of hyperboloids are distinct, then the chaining is called nondegenerate.

Theorem. *If there is a nondegenerate chaining of four two-dimensional hyperboloids of one sheet in five-dimensional space, then every fifth hyperboloid that is chained to three of them is chained to the fourth.*

An important component of a mathematician's professionalism is the ability to evaluate the difficulty of a problem a priori. In some sense, mathematicians believe that there is a law of conservation of "nontriviality," and so they have a bias against easily-solved problems that the experts say are difficult. One of the manifestations of this tradition is the belief that amateurs are unable to solve long-standing problems. The history of mathematics shows that, although there are counterexamples to this, on average this rule is satisfied, at least during periods of time comparable to a human lifespan. Those cases where the the difficulty of problems has been sharply overestimated correspond to revolutionary changes in mathematics. When these changes "mature" over a noticeable period of time (as with algebraic notation or infinitesimal calculus) we are able to get used to them. Another situation arises when some new possibility is unexpectedly discovered that leads to a decided reappraisal of the value of a problem. Not infrequently, we even see a desire to declare the new methods illegal. A striking illustration is Paul Gordan's (1837–1912) reaction to David Hilbert's (1862–1943) solution his problem on the finiteness of the number of invariants: "This is theology, not mathematics." Instead of constructing invariants directly, which was then the usual way to proceed and which resulted in separate and difficult cases, Hilbert proved their existence for the general case in one stroke.

However, it is probable that no revolution in mathematics has taken place so sharply as the penetration of analytic methods into geometry. Ideas about the difficulty of geometric problems collapsed and the role of geometric intuition—the pride of mathematicians—was devalued. What had taken refined argument was now obtained through standard calculations. Along with this came a conservative trend that fought for the genuine geometry that was being replaced by boring algebra. For comparison we note that the creation of analytical mechanics was much more painless, when Euler and Lagrange, rejecting Newton's geometric methods, turned mechanics into an area of mathematical analysis using the method of coordinates. The situation in geometry recalls somewhat the transition to ma-

chine production, when the art of the craftsman was lost to the monotonous assembly line of automatic work. Today we see clearly that the analytical methods did not destroy geometric intuition but, to the contrary, allow us to "save" it for comparatively simple situations and create intuition at a much higher level. However, we cannot deny that much of what was done in synthetic geometry (i.e., in traditional geometry without using coordinates) has been lost and there is no turning back.

Thus in the 1630s the two leading mathematicians of the time, Fermat and Descartes, discovered that coordinates could be used to make an equation in two unknowns correspond to a curve in the plane. This was an unexpected change in view, in particular because it was thought that since an equation with two unknowns has an infinite number of solutions there is no sense considering it (this is sometimes said in schools even now). Thanks to the geometric approach, this infinite set unexpectedly is granted the right of citizenship. No less fruitful was the inverse possibility of making curves correspond to their equations. This is where analytic geometry began.

A decisive discovery was that straight lines correspond to first-degree equations in the plane and conic sections to second-degree equations. So the two most basic objects of Greek geometry turned out to be the simplest ones from the analytic point of view. Geometers of the time dreamt of mastering and surpassing Apollonius' theory of conic sections. Fermat and Descartes were convinced that most of the statements would be surprisingly simple to prove in analytic language. Descartes succeeded in solving analytically several inaccessible problems of the Greeks on the locus of points. As statements made as epigraphs show, the creators of analytic geometry saw no limits to what they could do (they called the line and circle the planar loci of the Greeks, and the conic sections the solid loci). Even though a lot was not cleared up (they did not consider negative coordinates, there is no clear-cut theorem on expressing a second-order equation in canonical form, etc.), all the grounds for optimism were there.

Above all, new horizons opened up to geometry that would completely change its structure. There was no doubt that the next chapter to be created in geometry would be the theory of third-order curves. Descartes had considered some curves in this class, but it was natural to try to construct a general theory as detailed as for second-order curves. First it was necessary to classify such curves. This problem turned out not to be simple, and Newton solved it in the 1660s. (His manuscript was published much later, in 1704.) The solution speaks eloquently to the complexity of this problem: there are 72 different forms of third-degree curves. Nevertheless, the forms of these curves can be made visible since all these forms are divided into

four types:

$$xy^2 + cy = P(x), \qquad xy = P(x), \qquad y^2 = P(x), \qquad y = P(x),$$

where P is a third-degree polynomial in x.

The question of what the classification signifies automatically arises. For a curve given by an equation in some coordinate system, we seek a coordinate system in which its equation looks especially simple. We can often fix such a system almost uniquely, and then it is natural to think of the corresponding equation as canonical. More generally, properties of the equation that do not depend on the coordinate system—invariants—comprise the geometry of the curve. We see that the definition of the subject in geometry began to appear very early in analytical language. In synthetic language, instead of changing coordinate systems, the figures themselves are transformed and, as became clear only at the end of the 19th century (in Klein's Erlagen program), one studies their properties that do not change under transformation. Thus, different areas of analytic geometry correspond to different classes of coordinate systems, and in synthetic geometry to transformation groups.

Newton, studying the third-degree curve, explained many general things that were necessary for separating out the geometric component from the algebraic facts about the equations.

1. First of all it was necessary to specify the geometric meaning of the original analytic characteristic of the curve—its order (the degree of its given equation). Newton remarked that the order coincides with the largest number of points at which a straight line can intersect the curve. (In the case of a curve of order n, an nth-degree equation in one variable defines these points.) Here there are difficulties handling imaginary intersection points, for which there are still no methods.

2. We recall the most important generalization of this statement: if curves of orders k and l intersect in more than kl points, then they have an infinite number of intersection points. In other words, in the latter case they have a common component (an algebraic curve may fall into several components, e.g., the curve $Q_1Q_2 = 0$ decomposes into $Q_1 = 0$ and $Q_2 = 0$). If in this case one curve is irreducible in a natural way (does not fall into components), then it is completely contained in another. Maclaurin, Newton's younger contemporary, stated this theorem and Bezout proved it almost a hundred years later. The theorem bears Bezout's name today. A more precise statement includes complex intersections and intersections at infinity.

3. Newton had many opportunities to satisfy himself that the method of coordinates was an effective one. He showed this while extending various

facts about conic sections to algebraic curves of higher order. Here, for example, is what happens with the theory of diameters. Recall that if we take chords of a conic section, say an ellipse, that are parallel to some direction, then their midpoints lie on a straight line, called a diameter. If we take a set of parallel lines and each line intersects a curve of order n in n points $x_1, x_2, \ldots, x_n$, then we consider their centers of gravity $\frac{x_1+x_2+\cdots+x_n}{n}$. Note that the centers of gravity do not depend on the choice of coordinates for the line. Following Newton, we will not discuss the case of imaginary intersection points. Newton's theorem states that all the centers of gravity lie along one line.

Indeed, we choose a coordinate system so that the parallel lines are given by equations $y = \text{const}$. Let $F(x, y) = 0$ be the equation of a curve in this system,

$$F(x, y) = ax^n + bx^{n-1}y + cx^{n-1} + \cdots .$$

Then the points of intersection $x_1, x_2, \ldots, x_n$ are the roots of $F(x, y) = 0$ with respect to x for fixed y. By Vieta's theorem, $\frac{x_1+x_2+\cdots+x_n}{n} = -\frac{by+c}{an}$, i.e., all the centers of gravity lie on the line $anx + by + c = 0$, Newton's diameter.

4. Almost simultaneously with the creation of analytic geometry, Desargues and later Pascal set out the foundations of projective geometry. The original observation was that applying a central projection makes it possible to simplify geometric considerations (e.g., each conic section can be obtained from every other one by a projection). This gives rise to a method that competes with analytic geometry for simplifying and extending Apollonius' theory. Pascal prepared a comprehensive treatise which was lost, and works on projective geometry were forgotten about for a hundred years. Newton probably did not know about them. However, it did not escape him that using a projection (considering the "shadow of an illuminated point") greatly simplifies the theory of not only conic sections but of general algebraic curves. Newton's most important observation was that the order of a curve is preserved under projection. Applying this idea, he established that by using a projection any third-order curve can be reduced to a curve of the form $y^2 = P(x)$, where P is a cubic polynomial. Clairaut later proved this result.

5. The geometry of third-order curves is essentially richer than the geometry of second-order curves. First of all, some special points appear: double points (points of self-intersection, such as $(0, 0)$ for the curve $y^2 = x^2(x + 1)$) and cusps (such as $(0, 0)$ for $y^2 = x^3$). Furthermore, in the general case a tangent has a double point of intersection at the point of contact (the corresponding polynomial of one variable has a double root), but there can be a triple intersection point (for equations of higher degree

there can be even greater multiplicities). Triple intersections are called points of inflection ($(0, 0)$ for $y = x^3$). Third-degree curves can have up to three inflection points (the points $(0, 0)$, $(1, 0)$, $(2, 0)$ of $y^3 = x(x-1)(x-2)$). Maclaurin noted that in this case all these points must lie on the same line ($y = 0$ in the example). In the 18th century algebraic curves, above all third- and fourth-degree curves, were at the center of mathematicians' attention. These questions were put forth in the first analytic geometry texts. However, much remained unexplained. It became clear that to construct a harmonious theory it was necessary to add points at infinity as well as imaginary points, since algebraic equations of higher degree had to be solved all the time (as early as 1717, James Stirling (1692–1770) referred to a curve with a double imaginary point at infinity).

We can take these two situations into account within the realm of complex projective geometry, which was first introduced by Poncelet. His first surprising observation was that all circles intersect in two imaginary points at infinity. These points, called cyclic points, govern the entire conformal geometry of the plane. Another great discovery of Poncelet (shared with Joseph Gergonne (1771–1859)) was the principle of duality, by which every statement in plane geometry has a dual statement with lines replaced by points and vice versa. In connection with this it is natural to associate with a curve not only the set of its points but also the set of its tangents. This leads to an invariant that is dual to the order, called the class p. This is the largest number of tangents passing through a point. For a nonsingular third-degree curve, $p = 6$. The general formula for a nonsingular curve has the form

$$p = n(n-1). \tag{1}$$

It is surprising that Poncelet's remarkable discovery, the most famous creation of a new kind of geometric intuition, was made using synthetic language, since projective geometry had not yet been successfully combined with analytic geometry. Coordinates that would have served for all points of the projective plane, including points at infinity, had not yet been conceived.

Such coordinates appeared during the years 1827–1828 in the work of August Möbius (1790–1868) and Plücker. Plücker's construction of homogeneous coordinates was especially simple and convenient. It placed a point of the projective plane in correspondence with a triple of numbers $x = (x_0, x_1, x_2) \neq (0, 0, 0)$ up to a constant multiplier: $(x_0, x_1, x_2) = (\lambda x_0, \lambda x_1, \lambda x_2)$. Each line in the projective plane $\mathbb{P}^2$ is given by an equation $(\xi, x) := \xi_0 x_0 + \xi_1 x_1 + \xi_2 x_2 = 0$, where $\xi \neq (0, 0, 0)$. Also, ξ and $\lambda\xi$ correspond to the same line. Therefore, the lines in $\mathbb{P}^2$ naturally form another projective plane $\mathbb{P}^2_\xi$ with homogeneous coordinates ξ. The points of

$\mathbb{P}^2 = \mathbb{P}^2_x$ correspond to the lines in $\mathbb{P}^2_\xi$. In the end the principle of duality, which is completely nontrivial in synthetic language, becomes almost obvious in analytic language.

A general projective transformation of coordinates has the form $\bar{x} = \sum a_{ji}x_i$, where $\det(a_{ji}) \neq 0$. They are characterized by being one-to-one on $\mathbb{P}^2$ and taking lines to lines.

In order to fix an affine structure on $\mathbb{P}^2$ we need to specify that some line, for example (but not necessarily) $x_0 = 0$, is a line at infinity. Away from $x_0 = 0$ we can always choose coordinates of the form $(1, \bar{x}_1, \bar{x}_2)$. Then $\bar{x}_1 = \frac{x_1}{x_0}$, $\bar{x}_2 = \frac{x_2}{x_0}$ will be the corresponding Cartesian coordinates. When taking the analytic approach from the very beginning, we do not single out a line at infinity.

Starting off from Plücker's interpretation of duality, it is natural to consider, along with a curve $\Gamma = \Gamma_x$ on $\mathbb{P}^2_x$, the dual curve Γ_ξ on the dual plane $\mathbb{P}^2_\xi$. The tangents to Γ_x correspond to the points of Γ_ξ. The class of Γ_x coincides with the order of Γ_ξ. Plücker resolved a conjecture that Poncelet could not unravel (Poncelet's paradox). The question was that, as it is easy to see, equation (1) is not dual to itself. Plücker discovered that this equation is only valid for curves without singularities. For example, if Γ is a nonsingular third-degree curve, then its dual curve must have singularities. Plücker found a formula for the class of curves with singularities:

$$p = n(n-1) - 2d - 3r, \tag{2}$$

where d is the number of double points and r is the number of cusps. This equation is self-dual. Note that the double points of the dual curve Γ_ξ (let there be δ such points) correspond to the double tangents to the original curve, i.e., to the lines that are tangent at two points of Γ. The cusps of Γ_ξ correspond to the tangents at the inflection points of Γ (let there be ρ such points). Then $n = p(p-1) - 2\delta - 3\rho$. At the same time we obtain a formula for the number of inflection points,

$$\delta = 3n(n-2) - 6d - 8r. \tag{3}$$

In the case where there are no singularities, it is not hard to obtain this formula directly, by writing down the condition that a point is an inflection point in the form of a system of algebraic equations and then applying Bezout's theorem. In particular, if Γ is a nonsingular third-order curve, then $p = 6$; it has no double tangents and has nine inflection points, some of which can turn out to be complex.

Strictly speaking, equation (2) is valid if the singularities of the curve are no more complicated than the simplest self-intersections and cusps, and (3)

is valid if all inflection points are the simplest (the multiplicity of tangency equals three) and no line can be tangent to the curve at more than two points. The complete statements of these results are more cumbersome.

Plücker thought deeply about the question of complex singular points of real curves. Properly speaking, the formulas we have presented are valid for complex singularities and inflection points. The question of which of these points can be real is rather nontrivial. It is not at all necessary that all the inflection points counted by (3) be real. For example, among the nine points of inflection of a third-order curve without singularities no more than three can be real. Plücker wrote, "A new flight of spatial intuition is needed in order to include what is imaginary and will remain imaginary in all cases."

One of the most remarkable periods in the history of analytic geometry is associated with Plücker. His student Klein wrote, "Plücker's goal in geometry and his achievement are a new construction of analytic geometry. He followed a method that arose out of the tradition of Monge: completely joining together the construction with an analytic formula. . . . In Plücker's geometry a simple combination of formulas is translated into the language of geometric correspondences and, conversely, the latter are sent to analytic operations. Calculations are omitted by Plücker when possible, but in return the acuteness of internal perception and geometric interpretation that analytic equations have is developed and broadly applied to the point of virtuosity."

Plücker turned out to be the object of an attack by Steiner, a famous geometer but an aggressive opponent of analytical methods in geometry, i.e., using equations and working with imaginary objects. Steiner's attack was so energetic that Plücker interrupted his research in geometry for 20 years and returned to it only a short time before he died.

We will give a few examples of Plücker's geometrical constructions. However, first we recall that an nth-degree equation in two variables has $\frac{(n+3)n}{2}+1$ coefficients, which are determined up to a constant multiplier (this is not hard to prove by induction). Therefore, a curve of order n is determined by giving $\frac{(n+3)n}{2}$ points (we then get the necessary number of linear equations to determine the coefficients in the equation). In particular, to determine a line we must specify two points, for a conic section we need five, and for a third-order curve we require nine. However, these points must be in general position, and as n grows this condition becomes more delicate. We note that by Bezout's theorem two curves of order n usually intersect in n^2 points. But when $n > 3$ we have $n^2 > \frac{n(n+3)}{2}$, and through these n^2 points pass two (in fact an infinite number of) curves of order n.

This situation, called Cramer's paradox after Gabriel Cramer (1704–1752), greatly disturbed geometers from Maclaurin to Euler, and probably only Plücker finally understood it.

We will look at how Plücker proved Pascal's Theorem. Recall that six points $A_1, A_2, \ldots, A_6$ on a conic section $Q = 0$ are consecutively connected by a polygonal line called Pascal's hexagon (this polygonal line can have self-intersections). Let p_i be the side A_iA_{i+1} and $L_i = 0$ be the equation of the line passing through p_i. Let B_1, B_2, B_3 be the points of intersection of the opposing sides (p_1, p_4), (p_2, p_5), (p_3, p_6), respectively. We claim that B_1, B_2, B_3 lie on one line. Pascal reduced the general case to the case of a circle, but even for a circle the proof is not so simple.

Here is how Plücker reasoned. He passed third-order curves through the nine points $\{A_i, B_j\}$. There is a single curve that passes through nine points in general position, but these points are not in general position (see above about Cramer's paradox). All curves in the set

$$L_1L_3L_5 + \mu L_2L_4L_6 = 0, \tag{4}$$

which depends on an arbitrary parameter μ, will pass through $\{A_i, B_j\}$. Note each A_i, B_j is annihilated by a factor in each of the two terms. Let C be any point of $Q = 0$ other than A_j. We choose μ so that the coordinates of C satisfy (4). We fixed a third-order curve but when intersecting it with a second-order curve $Q = 0$ there can either be $2 \times 3 = 6$ or an infinite number points of intersection. Since we have at least the seven intersection points $A_1, \ldots, A_6, C$ the number of intersections is infinite, and since $Q = 0$ is irreducible it must be completely contained in (4), i.e., the left-hand side of (4) must be divisible by Q. After dividing by Q we find a linear expression M, and the points B_1, B_2, B_3 must lie on the line $M = 0$ since they cannot lie on the conic section $Q = 0$ (otherwise $Q = 0$ would have to intersect one of the lines $L_j = 0$ in three points).

This method, which uses sets of curves and the choice of a suitable curve from the set, which by Bezout's theorem should decompose (into a conic section and a line in the example), is quite characteristic of Plücker. The undetermined multiplier μ is always present in his arguments and is often called "Plücker's μ."

Remark. We will mention another analytic proof of Pascal's Theorem that is built on a completely different idea and clarifies the exceptional role of this theorem in projective geometry. We give the proof for the case of a circle and a hexagon with parallel sides (the general case reduces to this special one).

We define the addition of points on a circle corresponding to adding

their arc-coordinates modulo 2π. It is possible to realize this addition geometrically. We fix a point O which will play the role of zero. If a, b are two points on the circle, then the line passing through O parallel to the chord ab intersects the circle exactly in the point $a + b$ (why?). Since this addition corresponds to arithmetic addition of the coordinates modulo 2π, it is commutative and associative. Commutativity is clear geometrically, but what does associativity mean geometrically? Let us consider a hexagon $A_1A_2 \cdots A_6$ inscribed in a circle with A_3A_4 parallel to A_1A_6 and A_4A_5 parallel to A_1A_2. Then Pascal's theorem asserts that A_2A_3 and A_5A_6 are also parallel.

Let us take A_1 as the zero point and denote A_3 by x, A_4 by y, and A_5 by z. From the assumptions of parallelism it follows that A_6 corresponds to $x+y$ and A_2 to $y+z$. Hence $(x+y)+z$ (respectively, $x+(y+z)$) will be the intersection point of the line through A_1 parallel to A_5A_6 (respectively, to A_2A_3). Because of associativity, $(x+y)+z = x+(y+z)$, these must be the same line. So A_2A_3 and A_5A_6 are parallel and Pascal's theorem is proven.

This proof is connected with Hilbert's fundamental idea that numbers can be defined intrinsically in projective geometry. Desargues' theorem is responsible for commutativity and Pascal's theorem is responsible for associativity. The proof is also connected with the arithmetic of cubic curves.

Plücker again raised the question of reducing the equation of a curve to a canonical form, with the aim of having it look simpler and of having its algebraic structure directly reflect some geometric properties of the curve. For example, Plücker showed that the equation of a third-degree curve can always be written in the form

$$L_1L_2L_3 - M^3 = 0, \tag{5}$$

where $\{L_j, M\}$ are linear forms of the coordinates. To prove this representation is possible, we count the number of independent parameters in (5). There are three coefficients in a linear form and 12 in the four forms $\{L_j, M\}$, but (5) is preserved if we multiply L_1, L_2, L_3 by $\alpha_1, \alpha_2, \alpha_3$ and M by $\sqrt[3]{\alpha_1\alpha_2\alpha_3}$. Therefore, there are $12-3 = 9$ independent parameters, and since a general third-degree equation, as we saw, contains nine independent parameters (one less than the number of coefficients), Plücker concluded that a general equation can always be transformed into (5). To this argument we must also add some points to make it a rigorous proof. This was probably one of the first examples where counting the number of parameters was used as a heuristic method and also as a method of proof.

What sort of geometry lies behind (5)? Let A_j be the intersection of the lines $L_j = 0$ and $M = 0$. These points lie on the curve (5) and A_j is a

triple point of intersection of the curve and the line $L_j = 0$. This means that A_1, A_2, A_3 are inflection points and the lines $L_j = 0$ are tangent at these points. Moreover, $M = 0$ is the line containing the three inflection points A_1, A_2, A_3. Such lines are called inflection lines. If we consider complex inflection points, then, as we have noted, a nonsingular curve has nine such points. It turns out that a (complex) line through any two inflection points must contain a third. There are twelve inflection lines. This configuration of nine points and twelve inflection lines is very interesting and is the object of special study in projective geometry.

Plücker also proposed a special structure for the fourth-degree equation

$$L_1L_2L_3L_4 - \Omega^2 = 0, \tag{6}$$

where $\{L_i\}$ are linear forms and Ω is a quadratic polynomial. Recall that there are 14 independent parameters in the general fourth-degree equation. There are six coefficients in a quadratic polynomial so in (6) there are $3\cdot 4+6 = 18$ coefficients. Here we can multiply $\{L_i\}$ by numbers $\{\alpha_i\}$ and simultaneously multiply Ω by $\sqrt[4]{\alpha_1\alpha_2\alpha_3\alpha_4}$. Then the number of independent parameters equals $18 - 4 = 14$. Plücker concluded that the representation (6) is a general one.

Furthermore, the intersection points of each line $L_i = 0$ with the conic section $\Omega = 0$ are double intersection points with the curve given by (6). Thus, each line $L_i = 0$ has two points of tangency with the curve (instead of the four intersection points for general lines). Such tangents are called bitangents. So four bitangents are involved in (6) and eight of their points of tangency lie on the same conic section. There is a curiosity associated with this discovery of Plücker. It is not hard to count that a fourth-order curve has 28 bitangents (including complex ones). Plücker erroneously proposed that the points of tangency of any four of them lie on a conic section. In fact each pair of bitangents is contained in only five foursomes with this property. Plücker's error was discovered by none other than Steiner. Algebraic curves actually arose in synthetic geometry, and the example we have presented shows that the representatives of this school had enough geometric intuition to compete on a par with those of the analytic school. The ostentatious refusal to use analytic methods only gradually revealed the weak position of Steiner's school.

According to Plücker, the points of three-dimensional projective space $\mathbb{P}^3$ are given by 4-tuples of homogeneous coordinates $x = (x_0, x_1, x_2, x_3) \neq (0, 0, 0, 0)$, $x \sim \lambda x$, $\lambda \neq 0$. The planes in $\mathbb{P}^3$ comprise the dual projective space $\mathbb{P}^3_\xi$, and $\xi \in \mathbb{P}^3$ corresponds to the plane $\langle \xi, x\rangle = 0$. At the same time the manifold of lines G in $\mathbb{P}^3$ is a completely new geometric object. Its study was one of Plücker's main achievements. Lines in $\mathbb{P}^3$ are given by four

parameters, e.g., we can fix two different planes and then almost all lines are given by points of intersection with these planes. Thus the manifold G is four-dimensional. It is natural to introduce coordinates (Stiefel[1]) that can be taken as a generalization of homogeneous coordinates. Take the line through a pair of its distinct points x, y. Put x, y into a matrix $X = \binom{x}{y}$ with two rows and four columns. The line consists of points of the form $z = \lambda_1 x + \lambda_2 y$, $\lambda = (\lambda_1, \lambda_2) \neq (0, 0)$, i.e., λ contains homogeneous coordinates for the line. The matrix X is determined by the line up to left multiplication by a nondegenerate 2×2 matrix: $X \to gX$ (corresponding to passing to another pair of points on the line). Let $X = (X_1, X_2)$, where X_1, X_2 are 2×2 matrices. Then if $\det X_1 \neq 0$, the coordinates can be chosen so that $X_1 = E$ is the identity matrix (take points x with $x_0 = 1, x_1 = 0$ and y with $y_0 = 1, y_1 = 0$). There is an affine coordinate chart on G with coordinates $X_2 = (u_{ij})$ and whose closure coincides with G. These coordinates exist for almost all points of G. Plücker passed from the matrix X to its minors. These are the famous Plücker coordinates, which are very convenient for constructing the geometry of lines. We will consider two problems associated with the geometry of lines.

1. We represent the space $\mathbb{P}^3$ as the union of pairwise disjoint lines. Note that if we fix an affine structure for $\mathbb{P}^3$, then lines that intersect projectively either go to intersecting lines or to parallel lines (that intersect "at infinity"). At the same time skew lines in affine space correspond to disjoint lines. Therefore, in affine language we speak of representing space in the form of a union of pairwise skew lines.

We indicate a specific partition, namely, we consider lines that join x and $\sigma(x) = (-x_1, x_0, -x_3, x_2)$, i.e., $X = \binom{x}{\sigma(x)}$. We only have to verify that $\sigma(x) \neq \lambda x$ and that if y is a point on such a line, then $\sigma(y)$ will also lie on that line. This follows from the relation $\sigma(\lambda_0(x) + \lambda_1\sigma(x)) = -\lambda_1 x + \lambda_0\sigma(x)$, which can be verified directly. The property we have indicated means that lines either do not intersect or coincide. In G there is a two-dimensional submanifold of lines.

2. We study surfaces in $\mathbb{P}^3$ with two families of generating lines. It is taught in analytic geometry that among the second-order surfaces in $\mathbb{R}^3$ there are two types that have this property: hyperboloids of one sheet (their equations reduce to the form $x^2 + y^2 - z^2 = 1$) and hyperbolic paraboloids (their canonical equation is $z = x^2 - y^2$). We will show that even in the class of all surfaces (not just second-order) there are no other surfaces with this property. It is clear that there is a large number of surfaces with one system of generators (the developable surfaces), however, the condition that there

[1]Eduard Stiefel (1909–1978).

exist two systems is very restrictive and leaves only the hyperboloids of one sheet and the hyperbolic paraboloids. Note that in projective space hyperboloids of one sheet and hyperbolic paraboloids are (projectively) equivalent: in suitable homogeneous coordinates their equation has the form $x_0^2 + x_1^2 - x_2^2 - x_3^2 = 0$. Thus, from the projective point of view we can speak only about hyperboloids of one sheet.

Let us make our terminology more precise. We will say that there are two families of generating lines on a surface S if through each point there are at least two different lines that completely lie on S. Such a surface S is called irreducible if there is no smaller surface $S_0 \subset S$ with two families of generating lines. We want to avoid a discussion of analytic subtleties relating to a precise definition of the notion of a surface, and so we will only appeal to intuitive ideas about surfaces. This part of our discussion will not be rigorous, but those who have mastered the corresponding techniques will easily see how to make our arguments rigorous, say, for analytic surfaces.

Theorem. *Every nonplanar irreducible (analytic) surface in $\mathbb{P}^3$ with two families of generating lines is a hyperboloid of one sheet.*

Proof.

1. If the given surface has a flat part, then it coincides with a plane. Roughly speaking, a generator passing through a point of a flat part generates a plane.

2. Fix a generator l on the surface. Then all generators intersecting l generate the surface. It is obvious that the points of their union depend on two parameters and we have agreed not to make this more precise analytically.

3. If there are two nonintersecting generators l_1, l_2, then the generators intersecting both l_1, l_2 also generate the surface.

In fact, since the generators intersecting l_1 generate the surface, through each point of l_2 there is a generator intersecting l_1. The union of these generators must also coincide with the surface.

4. Analogously, the union of the generators that intersect any number of given pairwise disjoint generators coincides with the surface.

5. Given a pair of intersecting generators l_1, l_2, almost all generators intersecting one of them do not intersect the other. Indeed, let m be a segment on l_2. Through each point of m there must pass a generator different from l_2. If all of them intersect l_1, then they generate a part of a plane passing through l_1, l_2.

Thus, there is an infinite set of pairwise disjoint generators on the surface. It is enough for us that there are three such generators.

6. We will show that the given surface is uniquely determined by three of its pairwise disjoint generators. It coincides with the union of all lines in $\mathbb{P}^3$ that intersect all three fixed lines.

Using Plücker's favorite method of counting the number of parameters, we can see that this assertion is plausible. The set of all lines depends on four parameters. The intersection with each line gives one condition on the parameters (in other words, through each point in space there is a two-parameter family so through each point of a line there is a three-parameter family of lines). Requiring an intersection with three lines we impose three conditions, so there remains a one-parameter family of lines that generates the surface.

For correctness we should convince ourselves that these conditions are independent. Therefore, we will choose our lines more effectively. Let l_1, l_2, l_3 be pairwise disjoint lines. We will find the lines that intersect all of them. Let $A \in l_1$; then $A \notin l_2$. Pass a plane through A and l_2. This plane intersects l_3 in some point B. In the projective space $\mathbb{P}^3$ a line that does not lie in a plane intersects it, and the line l_3 cannot lie in the plane since l_2 and l_3 do not intersect. The line AB will be the unique line intersecting all three lines l_1, l_2, l_3 and passing through A. Thus, the set of lines intersecting three pairwise disjoint lines can be parameterized by the points of intersection with one of them.

7. We will prove that the union of these lines is a hyperboloid of one sheet. We will carry out the proof analytically. At the same time we prove that a hyperboloid of one sheet, in fact a unique one, passes through three pairwise disjoint lines in $\mathbb{P}^3$.

Let l_1, l_2, l_3 be three pairwise disjoint lines in $\mathbb{P}^3$. We will consider the lines as intersections of planes, i.e., as systems of two linear equations $\langle \xi, y \rangle = 0$ and $\langle \eta, y \rangle = 0$. Choose homogeneous coordinates so that l_1 is given by the system $y_0 = 0$, $y_1 = 0$ and l_2 by $y_2 = 0$, $y_3 = 0$. This can be done because the lines do not intersect (and the left-hand sides of the equations can be taken as coordinates). Then let l_3 be given by the system

$$\begin{cases} \xi_0 y_0 + \xi_1 y_1 + \xi_2 y_2 + \xi_3 y_3 = 0, \\ \eta_0 y_0 + \eta_1 y_1 + \eta_2 y_2 + \eta_3 y_3 = 0. \end{cases}$$

The vectors (ξ_0, ξ_1), (η_0, η_1) cannot both be zero, since then $ł_2$ and $ł_3$ would coincide. Moreover, these vectors cannot be proportional and, in particular, neither one can be zero. Indeed, if $(\eta_0, \eta_1) = (\lambda\xi_0, \lambda\xi_1)$, then the point $(-\xi_1, \xi_0, 0, 0)$ will be common to $ł_2$ and $ł_3$. We can similarly show that (ξ_2, ξ_2) and (η_3, η_3) cannot be proportional. Therefore, we can make the

following change of coordinates:

$$x_0 = \xi_0 y_0 + \xi_1 y_1, \qquad x_1 = \eta_0 y_0 + \eta_1 y_1,$$
$$x_2 = \xi_2 y_2 + \xi_3 y_3, \qquad x_3 = \eta_2 y_2 + \eta_3 y_3.$$

By what we said above about the proportionality of the vectors, such a substitution is admissible. In these coordinates l_1 is given by the equations $x_0 = x_1 = 0$, l_2 by $x_2 = x_3 = 0$, and l_3 by $x_0 + x_2 = x_1 + x_3 = 0$. However, all these lines lie on a hyperboloid of one sheet,

$$x_0 x_3 - x_1 x_2 = 0. \tag{7}$$

The lines we are considering belong to a one-parameter family of generators

$$\begin{aligned} &\lambda_0 x_0 + \lambda_1 x_1 = 0, \\ &\lambda_0 x_2 + \lambda_1 x_3 = 0, \quad (\lambda_0, \lambda_1) \neq (0, 0). \end{aligned} \tag{8}$$

A second family of generators consists of the lines

$$\begin{aligned} &\mu_0 x_0 + \mu_1 x_2 = 0, \\ &\mu_0 x_1 + \mu_1 x_3 = 0, \quad (\mu_0, \mu_1) \neq (0, 0). \end{aligned} \tag{9}$$

Recall that the generators in any one family are pairwise disjoint and that any generators in different families intersect, so that generators of each family pass through each point.

In each family the lines are parameterized by the points of the projective line $\mathbb{P}^1_\lambda$, $\mathbb{P}^1_\mu$, where λ, μ are corresponding homogeneous coordinates. Two curves on the manifold G of lines are associated with each hyperboloid. The curves, which are in 1–1 correspondence with the projective line, are called rational (admit a rational parameterization). Rational curves are not only lines in projective space but are second-order curves. Thus, with each hyperboloid of one sheet in $\mathbb{P}^3$ we associate two distinguished rational curves on G. In the spirit of Plücker's geometric approach the same geometric objects appear either in the form of hyperboloids of one sheet in the point geometry of $\mathbb{P}^3$ or in the form of the simplest rational curves on the manifold G of lines. (The class of these curves can be described directly.)

In passing to affine language we can encounter two possibilities: either no generator lies in a plane at infinity and then we obtain a hyperboloid of one sheet in the affine sense, or such a generator exists and then we obtain a hyperbolic paraboloid. In the latter case a pair of intersecting generators (one from each family) lies in a plane at infinity. This is connected to the fact that in any plane passing through one generator there is a second

generator that intersects the first. This can be rephrased as follows: given three pairwise disjoint lines in three-dimensional affine space $\mathbb{R}^3$, if there is no plane parallel to all three lines, then we can draw a hyperboloid of one sheet through them, and if there is such a plane, then we can draw a hyperbolic paraboloid. A line at infinity of this plane belongs to a second family of generators. In other words, for each family of generators of a hyperbolic paraboloid there is a plane to which they are all parallel.

The statement we have proved about hyperboloids of one sheet can be interpreted as a statement about the "rigidity" of surfaces formed by two families of linear generators. Such a construction cannot "budge" if three lines of one family are fixed. This situation is used in real constructions using linear rods in the shape of a hyperboloid, e.g., in the famous Shukhov radio tower in Moscow.

We now move to the final section devoted to chaining hyperboloids. The main objects of classical projective geometry are various kinds of configurations of points, lines, and planes. Included in these are cases where some geometric relations (e.g., three lines pass through one point, three points lie on one line, six points lie on one conic section, etc.) lead to others: the configurations of Desargues, Pascal, etc. The result we have given allows us to study configurations in multidimensional projective space, including two-dimensional hyperboloids of one sheet. Since we will only consider two-dimensional hyperboloids (i.e., hyperboloids in $\mathbb{P}^3$), we will often omit "two-dimensional" below.

We will give the points of n-dimensional projective space $\mathbb{P}^n$ homogeneous coordinates $(x_0, x_1, \ldots, x_n)$. Fix a geometric relation for lines in $\mathbb{P}^n$. Through any two disjoint lines in $\mathbb{P}^n$ we can pass a unique three-dimensional plane. For three pairwise disjoint lines there is a geometric relation: "lie in one three-dimensional plane." For four pairwise disjoint lines lying in one three-dimensional plane there is the relation that they "belong to one (two-dimensional) hyperboloid of one sheet." By contrast, three pairwise disjoint lines lying in one three-dimensional plane always generate a hyperboloid of one sheet.

As we said at the start, two two-dimensional hyperboloids of one sheet in $\mathbb{P}^n$ are called chained if they have a common generator that is the line of intersection of the three-dimensional planes they generate. Correspondingly the minimal dimension in which chained hyperboloids exist is equal to 5. Several two-dimensional hyperboloids of one sheet in $\mathbb{P}^n$ are called chained if they are pairwise chained and the lines for each pairwise chaining belong to a single family of generators, which we call a distinguished family. A chaining is called nondegenerate if each of these lines belongs to only two of the hyperboloids.

Suppose we have four nondegenerate chained two-dimensional hyperboloids of one sheet in five-dimensional projective space $\mathbb{P}^5$. Giving such a foursome is equivalent to giving four three-dimensional planes in $\mathbb{P}^5$ in general position. Let us figure out what this means. Recall that, as a rule, k-dimensional and l-dimensional planes in $\mathbb{P}^n$ intersect in a $(k+l-n)$-dimensional plane (and in a higher-dimensional plane in the degenerate case). Then two three-dimensional planes in $\mathbb{P}^5$ usually intersect in a line $(3+3-5=1)$ and three planes generally do not intersect $(1+3<5)$. Thus, three three-dimensional planes in $\mathbb{P}^5$ in general position have no points in common. Consequently, any two of these planes intersect in a line (if two were to meet in a two-dimensional plane, then their intersection with a third plane would be a point: $2+3-5=0$).

Correspondingly, four planes are in general position if any three of them have no points in common and thus any two meet in a line. Then in each such three-dimensional plane there is a threesome of lines in which it intersects the other planes. These lines are pairwise disjoint, and we can draw a (two-dimensional) hyperboloid of one sheet through them. This leads to four nondegenerately chained hyperboloids. On the other hand, if we have four nondegenerately chained hyperboloids, then the three-dimensional planes they generate will obviously be in general position. We can now make a fundamental statement which is somewhat stronger than the theorem on five hyperboloids that we stated at the beginning of this chapter.

Theorem. *Suppose we are given four nondegenerately chained two-dimensional hyperboloids of one sheet in five-dimensional projective space $\mathbb{P}^5$. Then every three-dimensional plane intersecting two of them in generators of the distinguished families also intersects the two remaining hyperboloids in generators of the distinguished families. The four lines where the plane intersects the hyperboloids lie on a single hyperboloid of one sheet.*

Corollary (on five hyperboloids). *Given four nondegenerately chained two-dimensional hyperboloids of one sheet in $\mathbb{P}^5$, every hyperboloid chained with three of them is chained with the fourth.*

The corollary holds because the plane for the fifth hyperboloid satisfies the hypotheses of the theorem. The statement of the theorem is stronger than that of the corollary. We can arbitrarily choose one generator from the distinguished families for two hyperboloids. They will not intersect, and a unique three-dimensional plane passes through them. This plane intersects the planes for the two other hyperboloids in lines. We claim that these lines lie on the hyperboloids. This is a very strong assertion, since

on a three-dimensional plane there is a four-parameter family of lines and we claim that the line of intersection falls in a one-parameter subfamily of generators of the hyperboloid. We add to this that the four lines of intersection lie on a single hyperboloid.

We have a two-parameter family of planes that satisfy the hypotheses of the theorem: they can be given, arbitrarily choosing one generator from the distinguished families, on two hyperboloids. One hyperboloid arises from each plane and the hyperboloids (a two-parameter family) are pairwise chained by the theorem.

We turn to the proof of the theorem. We will proceed analytically, analogous to the proof of the preceding theorem.

We are given a three-dimensional plane in $\mathbb{P}^5$ as the intersection of hyperplanes, i.e., as a system of two linear equations $\langle \xi, y\rangle = \langle \eta, y\rangle = 0$ in homogeneous coordinates. Suppose we have four three-dimensional planes l_1, l_2, l_3, l_4, where no three have any points in common. We choose homogeneous coordinates so that l_1 is given by the system $y_0 = y_1 = 0$, l_2 by $y_2 = y_3 = 0$, and l_3 by $y_4 = y_5 = 0$. This can be done since l_1, l_2, l_3 have no points in common and so the left-hand sides of the equations that determine these planes are independent (the homogeneous equations have only the trivial solution). Let l_4 be given in this coordinate system by the equations

$$\sum_{i=0}^{5} \xi_i y_i = 0, \qquad \sum_{i=0}^{5} \eta_i y_i = 0.$$

The vectors (ξ_0, ξ_1), (η_0, η_1) cannot be zero simultaneously, since this would mean that the planes l_2, l_3, l_4 have the line $y_2 = y_3 = y_4 = y_5 = 0$ in common, and so l_2, l_3 would intersect. We will show that they cannot be proportional either and, in particular, that neither one can be zero. Suppose $(\eta_0, \eta_1) = (c\xi_0, c\xi_1)$. Then the point $(-\xi_1, \xi_0, 0, 0, 0, 0)$ would belong to the three planes l_2, l_3, l_4, and we assumed they had no points in common. We can show analogously that the pairs of vectors (ξ_2, ξ_3), (η_2, η_3) and (ξ_4, ξ_5), (η_4, η_5) are not proportional. As a result we can make the change of coordinates

$$\begin{aligned} &x_0 = \xi_0 y_0 + \xi_1 y_1, \qquad x_1 = \eta_0 y_0 + \eta_1 y_1, \qquad x_2 = \xi_2 y_2 + \xi_3 y_3,\\ &x_3 = \eta_2 y_2 + \eta_3 y_3, \qquad x_4 = \xi_4 y_4 + \xi_5 y_5, \qquad x_5 = \eta_4 y_4 + \eta_5 y_5. \end{aligned}$$

In these coordinates the planes l_1, l_2, l_3, l_4 are given by the corresponding system of equations

$$\begin{gathered} x_0 = x_1 = 0, \qquad x_2 = x_3 = 0, \qquad x_4 = x_5 = 0,\\ x_0 + x_2 + x_4 = x_1 + x_3 + x_5 = 0. \end{gathered}$$

All four of these planes are contained in the family of three-dimensional planes

$$\lambda_0 x_0 + \lambda_1 x_2 + \lambda_2 x_4 = 0, \qquad \lambda_0 x_1 + \lambda_1 x_3 + \lambda_2 x_5 = 0. \tag{10}$$

Let Σ_λ denote the plane given by (10). It is natural to consider the parameters as homogeneous coordinates on the two-dimensional projective plane $\mathbb{P}^2$. Here $(\lambda_0, \lambda_1, \lambda_2) \neq (0, 0, 0)$ and $\lambda, c\lambda$ give the same plane. The planes l_1, l_2, l_3, l_4 correspond to the parameters $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$, $(1, 1, 1)$.

We now consider the pairwise intersections $\Sigma_\lambda \cap \Sigma_\mu$, $\lambda \neq c\mu$. This is the line in $\mathbb{P}^5$ whose points satisfy the four equations comprising the union of (10) for λ and μ. It is obvious that $\Sigma_\lambda \cap \Sigma_\mu$ depends only on the line $\{\lambda, \mu\}$ in the parametric plane $\mathbb{P}^2$ joining the points λ and μ. Therefore, if we fix λ, the lines $\Sigma_\lambda \cap \Sigma_\mu$ in $\mathbb{P}^2$ correspond to the lines that pass through λ. It is natural to introduce the structure of a projective line on the set of lines in the projective plane $\mathbb{P}^2$ that pass through a fixed point. We consider the lines $\Sigma_\lambda \cap \Sigma_\mu$ for fixed λ more concretely. For specificity, let $\lambda_0 \neq 0$ (there is no loss of generality because we can replace λ_0 by any λ_j). As homogeneous coordinates on Σ_λ we take (x_2, x_3, x_4, x_5). In these coordinates the lines $\Sigma_\lambda \cap \Sigma_\mu$ are given by the system

$$\begin{aligned} \bar{x}_0 x_2 + \bar{x}_1 x_4 = 0, &\qquad \bar{x}_0 x_3 + \bar{x}_1 x_5 = 0, \\ \bar{x}_0 = \mu_0\lambda_1 - \lambda_0\mu_1, &\qquad \bar{x}_2 = \mu_0\lambda_2 - \lambda_0\mu_2. \end{aligned} \tag{11}$$

If λ, μ are not proportional, then $\bar{x} = (\bar{x}_0, \bar{x}_1) \neq (0, 0)$. Thus, $\Sigma_\lambda \cap \Sigma_\mu$ depends only on the line $\{\lambda, \mu\}$. From (11) it is clear that the lines $\Sigma_\lambda \cap \Sigma_\mu$ are, for fixed λ, a family of generators for some hyperboloid H_λ of one sheet in the three-dimensional plane Σ_λ. Thus, we have obtained a family of chained hyperboloids H_λ that depend on two parameters $\lambda \in \mathbb{P}^2$. This chaining is nondegenerate since all hyperboloids H_ν, where ν belongs to the line $\{\lambda, \mu\}$, pass through the generator $\Sigma_\lambda \cap \Sigma_\mu$.

It remains to prove that every three-dimensional plane intersecting two different hyperboloids H_λ, H_μ in generators from the distinguished families is one of the planes in the set Σ_λ. Each plane in the set intersects each hyperboloid H_λ in a generator from the distinguished family, including the original foursome. Now let a three-dimensional plane be generated by different lines $\Sigma_\lambda \cap \Sigma_{\lambda'}$ and $\Sigma_\mu \cap \Sigma_{\mu'}$, and let ν be a point on the parametric plane $\mathbb{P}^2$ that is the intersection of the lines $\{\lambda, \lambda'\}$, $\{\mu, \mu'\}$ (they cannot coincide). Then Σ_ν contains $\Sigma_\lambda \cap \Sigma_{\lambda'}$ and $\Sigma_\mu \cap \Sigma_{\mu'}$, and thus coincides with the plane under consideration. (We have essentially proved that a certain system of linear equations in ν has a solution, which we could have done directly.) Thus, the proof of the theorem on chaining hyperboloids is complete.

The Complex World of Roger Penrose

> We cannot think of anything so strange and unlikely that was not already said by some philosopher. René Descartes

At the International Mathematical Congress in Helsinki in the summer of 1978, Roger Penrose[1] (1931–) gave a plenary address entitled, "The Complex Geometry of the Real World." Penrose's fundamental idea was that it is natural to interpret the points of the four-dimensional space-time of Minkowski or Euclid (in Euclidean field theory) as complex lines in a three-dimensional complex space. This idea was developed by Penrose over the years into his "twistor program." (He called the points of the auxiliary three-dimensional complex space twistors.) Not long before the congress, the first results appeared that could not be considered purely interpretive (instanton solutions of the Yang–Mills equations and complex self-dual solutions of the Einstein equations).

Penrose's approach was in essence not new: a complex realization of Minkowski space was contained in Élie Cartan's (1869–1951) theory of homogeneous manifolds. However, the significance lies not in the geometric observation itself but rather in the idea of making it a systematic source of analytical constructions, namely integral representations for the solutions of certain important linear and nonlinear equations of mathematical physics. By a happy coincidence, at just this time in mathematics (algebraic geometry and the theory of functions of several complex variables) some quite nonelementary mathematical machinery appeared that was needed

[1]The translation of this chapter includes material from an English version that appeared in *The Mathematical Intelligencer*, **5** (1983), pp. 27–35.—*Transl.*

to realize this plan (bundles on projective spaces, Cauchy–Riemann cohomology, etc.).

Returning to Penrose's geometric idea, we should probably not be surprised at how complex objects appear in the study of a purely real object such as space-time. It would not have seemed surprising to geometers in the second half of the 19th century. Penrose's construction is connected to mathematical ideas that are a little more than a hundred years old and that in recent decades have undeservedly been forgotten, perhaps because they are so concrete. We are talking about the idea of Julius Plücker (1801–1868) of considering the space whose elements (points!) are lines in ordinary three-dimensional space. Plücker developed this idea over the course of many years and it appears in final form in his memoir *A New Geometry of Space Based on Considering a Line as a Space Element*, published posthumously in 1868–1869 and edited by Felix Klein and Alfred Clebsch. The dimension of the space of lines is four and it is probably the first four-dimensional space that appeared in science. Strangely enough, at a time when four dimensions appeared in relativity theory and when there was general enthusiasm for four-dimensional structures, nobody compared Minkowski's four dimensions with Plücker's, which had appeared 50 years earlier. In a sense this is just what Penrose did 50 years later. We will try to trace a possible path from Plücker to Hermann Minkowski (1864–1909), but for this we must recall still earlier events.

"The Golden Age of Geometry"

This is how Bourbaki described the 19th century, the century of the growth of projective geometry with its fantastic flight of geometric intuition and powerful analytical methods. The leading role of projective geometry in 19th century geometry is indisputable. It is characteristic that for many mathematicians the acceptance of non-Euclidian geometry was connected to its realization as part of projective geometry (the Klein interpretation). But projective geometry (also called the "new geometry") began much earlier. Gerard Desargues (1593–1662), an architect from Lyon, published a book in 1639 entitled *The First Draft of an Attempt to Understand What Becomes of the Meeting of a Cone with a Plane*. Desargues was developing the theory of perspective and studied a central projection of one plane onto another. He noted that the first plane had points that are not mapped anywhere and the other plane had points that were not in the image. He decided to improve the matter by introducing ideal points at infinity. In modern form, his idea was that all parallel lines "intersect" at one common "infinite" point, and all infinite points of a plane constitute an infinite line, which must be

added to the plane. On the extended (projective) plane all statements about parallelism turn into a special case of the usual statements about intersections of lines with no restrictions added (any two distinct lines intersect in a unique point, perhaps at infinity). The ideas of projective geometry were digested with great difficulty, and Desargues could not put them in a form that was easy to understand. Among the members of the Marin Mersenne's (1588–1648) group, an embryo of the Academy of Sciences of Paris, he had found only one disciple. It was the 16-year-old Blaise Pascal (1623–1662), who had proved a famous theorem on a hexagon inscribed in a conic section. The methods of projective geometry enabled Pascal to reduce the general case to that of a circle, since by definition any conic section is obtained from a circle by a central projection. Desargues and Pascal planned to use projective geometry to shed light on Apollonius's theory of conic sections, the apex of Greek geometry. European mathematicians, already unquestionable masters of algebra and analysis, had been trying for a long time to do battle with the great Greeks on their own territory, geometry. Desargues and Pascal were successful, but nobody could understand Desargues' work and Pascal never finished his comprehensive treatise on projective geometry, leaving to his heirs only a small placard with his hexagon theorem. Their work was forgotten for 200 years and when it was remembered, thanks to Michel Chasles (1793–1880), most of their results were rediscovered.

A new life for projective geometry began with the works of Gaspard Monge (1746–1818) and his pupils, among them Jean-Victor Poncelet (1788–1867). As Felix Klein (1849–1925) said, a new type of geometric thinking appears in the works of Poncelet, "projective thinking." While he was in captivity in Saratov, Russia after Napoleon's campaign of 1812, Poncelet fell into turbulent geometric fantasies and shared his discoveries with former fellow students of Monge at the École Polytechnique. He collected his results in his *Treatise on the Projective Properties of Figures*, which was only published 10 years later. He never returned to the systematic study of geometry: government and military service, teaching, studies of fortification and of theory of mechanisms (the "Poncelet water wheel") diverted him. At the end of his life he returned to geometry but mostly regretted that he had not been able to give regular attention to mathematics, that others had not explored projective geometry as he thought they should have, and that Chasles had inopportunely remembered Desargues (i.e., that projective geometry did not originate with Poncelet).

Poncelet began with the observation that just as on a projective plane there are no exceptions to how lines are mutually related, likewise there must be no exceptions for second-order curves. But why, then, did ellipses

usually intersect in four points while their special case, circles, always intersect in only two? Poncelet found the answer: all circles pass through two fixed points, called cyclic points. However, we do not notice these points since they are both infinite and imaginary. This is how complex numbers first appeared in real geometry (mathematicians had only just begun to get used to them in algebra). Cyclic points became one of the main objects of geometry. With their help one could explain all real metric relationships in the plane.

Another astounding discovery of Poncelet is the duality principle, a new method for obtaining geometric assertions for which he shares the honors with Joseph Gergonne (1771–1859). Roughly speaking, it states that in a theorem on the mutual position of points and lines on a projective plane the words "line" and "point" may be interchanged and, after any necessary editing so that the text makes sense (replacing "intersect" by "pass through," etc.), we obtain a new theorem. For example, "two distinct lines intersect" turns into "a unique line passes through two distinct points."

From that time on, projectivity became the reigning method in geometry. However, for a long time projective ideas were considered to be a black box-like device for proving Euclid's theorems. Infinite elements were looked on as ideal, alien elements that simplified considerations (similar to the way complex numbers were first viewed). The consistent projective approach, however, required one to consider both finite and infinite points on an equal footing and, for example, the behavior of curves at infinity (e.g., asymptotes) deserved no special interest. The minds of geometers were occupied for a long time with discussions about the ideas of projective geometry. These discussions are especially noteworthy when we turn to German geometry of the mid-19th century, i.e., the time of such remarkable geometers as Ferdinand Möbius (1790–1868), Julius Plücker, Jakob Steiner (1796–1863), and Christian von Staudt (1798–1867). Their activities were carried out against the background of a bitter struggle between the "analytics" and "synthetics," although nowadays their differences might be no more cause for argument than those of Jonathan Swift's characters, who discussed which end of the egg it is better to begin eating. Analytics mainly used the coordinate representation of geometric figures, as it allowed them to apply the methods of algebra and analysis. Synthetics believed that these methods deprived geometry of its genuine spirit, of true geometric intuition.

The most active synthetic was Steiner, the son of a peasant, who walked behind a plough until he was 19 years old when he became the pupil and colleague of the famous Swiss teacher Johann Pestalozzi (1746–1827), and only turned to mathematics when he was older. Steiner had wonderful

geometric intuition and the flights of his spatial imagination were impossible to depict even with pictures. He refrained from using them at his lectures, which were held in darkened classrooms to help the students to concentrate. Steiner remonstrated emphatically against complex numbers, those "ghosts" and "shadow-land in geometry" which the analytics used so much. Klein believed that perhaps it was Steiner's intolerance that made Plücker, a typical analytic, cut short his studies of geometry and resume them only after Steiner's death.

Projective Coordinates

Above all, the analytics raised the problem of introducing coordinates on a projective plane so as to include both finite and infinite points. The crucial construction here (homogeneous coordinates) is due to Plücker. He proposed to characterize the points of a projective plane not by two but by three numbers $(x_0, x_1, x_2) \neq (0, 0, 0)$, but he assumed that 3-tuples that differ by a common multiplier, (x_0, x_1, x_2) and $(\lambda x_0, \lambda x_1, \lambda x_2)$, correspond to the same point of the plane. Then we may assume, for example, that points with $x_0 \neq 0$ are "finite," and we may always take these points with $x_0 = 1$, i.e., $(1, X_1, X_2)$, where $X_1 = \frac{x_1}{x_0}$, $X_2 = \frac{x_2}{x_0}$ are inhomogeneous (Cartesian) coordinates. Points with $x_0 = 0$ constitute a line at infinity. However, this line may be fixed arbitrarily. Projective transformations of the plane, which transform lines into lines, correspond to linear transformations of homogeneous coordinates. Lines on the projective plane are defined by equations of the form $\xi_0 x_0 + \xi_1 x_1 + \xi_2 x_2 = 0$, where $(\xi_1, \xi_1, \xi_2) \neq (0, 0, 0)$ and is defined up to a scalar multiplier. This led Plücker to consider (ξ_1, ξ_1, ξ_2) as homogeneous coordinates of lines, and then we obtain that lines constitute another (dual) copy of the projective plane. This interpretation makes the Poncelet-Gergonne duality principle completely transparent.

By using homogeneous coordinates it is easy to understand Poncelet's theorem on intersecting circles, which requires a high level of geometric intuition in its synthetic form. In inhomogeneous coordinates equations of circles have the form $X_1^2 + X_2^2 + aX_1 + bX_2 + c = 0$, or in homogeneous coordinates,

$$x_1^2 + x_2^2 + ax_1x_0 + bx_2x_0 + cx_0^2 = 0.$$

It is clear that all these curves contain the pairs of points $(0, 1, i)$, $(0, 1, -i)$, i.e., these points are in fact infinite and imaginary ($\{x_0 = 0\}$ is the line at infinity).

In three-dimensional projective space, points are characterized by four numbers $(x_0, x_1, x_2, x_3) \neq (0, 0, 0, 0)$, defined up to proportionality. We may assume that $\{x_0 = 0\}$ is a plane at infinity. Planes are defined by

equations $x_0\xi_0 + \cdots + x_3\xi_3 = 0$, i.e., there is a duality between the projective space of points and the projective space of planes.

The Manifold of Lines (Plücker Coordinates)

The next natural question that excited Plücker's curiosity concerned constructing the collection of lines in the projective space $\mathbb{P}^3$. It turned out that unlike the case of planes (and of lines on the plane) we arrive at a completely new geometric formation. The set of lines in $\mathbb{P}^3$ depends on four parameters. In Cartesian coordinates X_1, X_2, X_3 almost all lines may be written as $X_1 = \alpha_1 X_3 + \beta_1$, $X_2 = \alpha_2 X_3 + \beta_2$. This parameterization does not include lines that are parallel to the plane X_1OX_2, and there are still lines at infinity.

Plücker proposed introducing coordinates on the whole collection of lines. He reasoned as follows. A line is defined by a distinct pair of points, i.e., in homogeneous coordinates in $\mathbb{P}^3$ by $x = (x_0, x_1, x_2, x_3)$, $\tilde{x} = (\tilde{x}_0, \tilde{x}_1, \tilde{x}_2, \tilde{x}_3)$, where x and $\tilde{x}$ are not proportional. This pair of points, however, may be chosen in many ways. To get rid of this arbitrariness we have to consider expressions

$$p_{ij} = x_i\tilde{x}_j - x_j\tilde{x}_i \tag{1}$$

that no longer depend on the choice of points (up to proportionality). Here we have $p_{ii} = 0$, $p_{ij} = -p_{ji}$. We will call the set of six numbers $p_{01}, p_{02}, p_{03}, p_{12}, p_{13}, p_{23}$ the *Plücker coordinates of the line*. Since the points were given in homogeneous coordinates, the sets $\{p_{ij}\}$ and $\{\lambda p_{ij}\}$ correspond to the same line. If all the p_{ij} vanish, then x and $\tilde{x}$ are proportional, which we have excluded. As a result, it is natural to consider sets of six nonzero numbers $\{p_{ij}\}$ up to proportionality as giving the homogeneous coordinates of a point in the five-dimensional projective space $\mathbb{P}^5$.

Thus, the set of lines turns out to be naturally embedded in $\mathbb{P}^5$. Since it depends on four parameters, the numbers p_{ij} must satisfy one more relation. Indeed, one can verify that this identity is always satisfied:

$$p_{01}p_{23} - p_{02}p_{13} + p_{03}p_{12} = 0. \tag{2}$$

It is also not difficult to see that there are no other relations, namely, for any nonempty set of numbers $\{p_{ij}\}$ satisfying (2) we can find $x, \tilde{x}$ satisfying (1).

From the geometrical point of view (2) defines a second-order surface in $\mathbb{P}^5$. If we pass to coordinates

$$\begin{aligned} &p_{01} = u_0 - u_3, \qquad p_{23} = u_0 + u_3, \qquad p_{02} = u_4 - u_1, \\ &p_{13} = u_4 + u_1, \qquad p_{03} = u_2 - u_5, \qquad p_{12} = u_2 + u_5, \end{aligned}$$

then equation (2) takes the form

$$u_0^2 + u_1^2 + u_2^2 - u_3^2 - u_4^2 - u_5^2 = 0. \tag{3}$$

Thus, the set of lines in three-dimensional projective space $\mathbb{P}^3$ is embedded as the second-order ("quadric") surface (2) or (3) in five-dimensional projective space $\mathbb{P}^5$. This discovery of Plücker played a main role in the molding of contemporary mathematical ideas, establishing an isomorphism of two completely different geometric structures: a manifold of lines in $\mathbb{P}^3$ and a quadric surface in $\mathbb{P}^5$. After this the best geometers, Sophus Lie (1842–1899), Felix Klein, and Élie Cartan lovingly collected similar isomorphisms. But later on interests shifted toward taking a general view of manifolds, working only with coordinates regardless of the geometrical nature of the points.

In the first place, Plücker's successors were interested in the following question. Suppose we consider a quadric in $\mathbb{P}^5$ not with three $+$ signs and three $-$ signs as in (3), which we call signature (3.3), but with signature (4.2) or (5.1). Do these quadrics admit a similar geometric interpretation? Lie discovered that one can just as naturally introduce homogeneous coordinates in the set of spheres in three-dimensional space and obtain a quadric with signature (4.2) in $\mathbb{P}^5$ (Lie sphere geometry). Klein introduced rather delicate "hexaspherical" coordinates in four-dimensional space, and points with these coordinates made up a quadric of signature (5.1) in $\mathbb{P}^5$.

We will also be interested in this problem, but we will consider another way to solve it. Passing to complex space erases the difference between quadrics of different signatures, since multiplication by i enables us to change to coordinates where the equation takes the form $z_0^2 + \cdots + z_5^2 = 0$ (all real quadrics are real forms of a single complex one). If we wish to pass from one real form to another the usual logic of projective geometry makes us "complexify" the problem and move into complex space.

The Complex Picture

Let $\mathbb{CP}^3$ be complex projective space, with complex homogeneous coordinates $z = (z_0, z_1, z_2, z_3)$. The complex line that joins $z, \tilde{z}$ consists of points of the form $\lambda z + \mu\tilde{z}$. In the set of complex lines we introduce complex Plücker coordinates p_{ij} satisfying (2), which reduces to (3) with u_j complex.

Let us consider real subsurfaces of the complex quadric $Q \subset \mathbb{CP}^5$ defined by (3). If we assume all u_j are real, then we obtain the case considered above. However, we can assume that u_0, u_1, u_2, u_5 are real and $u_3 = iv_3$, $u_4 = iv_4$ are purely imaginary, or that only $u_3 = iv_3$ is purely imaginary

and the other coordinates are real. Then we obtain real subsurfaces (u and v are real!),

$$u_0^2 + u_1^2 + u_2^2 + v_3^2 + v_4^2 - u_5^2 = 0, \quad \text{(S)}$$
$$u_0^2 + u_1^2 + u_2^2 + v_3^2 - u_4^2 - u_5^2 = 0. \quad \text{(H)}$$

These are a sphere and a hyperboloid of one sheet in homogeneous coordinates, respectively. Since these real surfaces lie on the complex quadric Q and complex lines correspond to the points of Q, it is natural to try to clarify which complex lines correspond to points of the surfaces (S) and (H).

Interpreting Real quadrics in the Language of Complex Lines (the Case of a Sphere)

In case (S) we have

$$p_{01} = u_0 - iv_3, \qquad p_{23} = u_0 + iv_3, \qquad p_{02} = iv_4 - u_1,$$
$$p_{13} = iv_4 + u_1, \qquad p_{03} = u_2 - u_5, \qquad p_{12} = u_2 + u_5.$$

Thus, the Plücker coordinates corresponding to the points of (S) satisfy the conditions

$$p_{23} = \bar{p}_{01}, \qquad p_{13} = -\bar{p}_{02}, \qquad \operatorname{Im} p_{03} = \operatorname{Im} p_{12} = 0. \quad (4)$$

These conditions completely characterize the points of (S). Then if the line with these Plücker coordinates passes through $z = (z_0, z_1, z_2, z_3)$, we may assume that the other point is $\tilde{z} = (-\bar{z}_3, \bar{z}_2, -\bar{z}_1, \bar{z}_0)$. Thus the complex lines in $\mathbb{CP}^3$ passing through (z_0, z_1, z_2, z_3) and $(-\bar{z}_3, \bar{z}_2, -\bar{z}_1, \bar{z}_0)$ correspond to the points of the real quadric (S).

What is remarkable about these lines? Through each point $z \in \mathbb{CP}^3$ there passes a unique line of this kind. As a result, the whole space $\mathbb{CP}^3$ is the union of nonintersecting lines. This partition (fibration) plays an important role in mathematics and appeared not very long ago independently of Plücker's considerations. If we intersect this fibration of $\mathbb{CP}^3$ with the real projective space $\mathbb{P}^3$, we will obtain a fibration of $\mathbb{P}^3$ into lines that join (x_0, x_1, x_2, x_3) and $(-x_3, x_2, -x_1, x_0)$. In simple terms we obtain a partition of the usual three-dimensional space into pairwise skew lines. (This is the way the problem was stated in the 1979 Moscow Mathematical Olympiad.) The realization of (S) as a fibration of $\mathbb{CP}^3$ is the first of the fundamental constructions of twistor theory.

Realizing a Hyperboloid as a Family of Lines

In case (H) we have

$$p_{23} = \bar{p}_{01}, \qquad \operatorname{Im} p_{13} = \operatorname{Im} p_{02} = \operatorname{Im} p_{03} = \operatorname{Im} p_{12} = 0. \tag{5}$$

First, suppose for simplicity that $p_{03} \neq 0$. Because the coordinates are homogeneous we may also assume $p_{03} = 1$ and choose points on the corresponding line with coordinates $z_0 = \tilde{z}_3 = 1$, $z_3 = \tilde{z}_0 = 0$. Then they are uniquely defined. From (5) it follows that $z = (1, a, c, 0)$, $\tilde{z} = (0, c, b, 1)$, where a, b are real. What is remarkable about the lines that join such pairs of points? It is straightforward to check that all points on these lines (i.e., points of the form $w = \lambda z + \mu \tilde{z}$ with λ, μ complex) must satisfy

$$\operatorname{Im}(w_1 \bar{w}_0 + w_2 \bar{w}_3) = 0. \tag{6}$$

If we remove the restriction $p_{03} \neq 0$, we find that there are no other lines all of whose points satisfy (6). Therefore, condition (6) defines a surface N of real dimension five in $\mathbb{CP}^5$ such that all complex lines in N are lines whose Plücker coordinates satisfy (5), hence they are lines that correspond to points of the real surface (H). Note that the surface N completely contains the real projective space $\mathbb{P}^3$ and that, generally speaking, a family of complex lines depending on four real parameters fills out a domain in $\mathbb{CP}^3$, which follows from considering dimensions. Therefore, we expect N to have some important properties. Indeed N is the unique surface, up to projective transformations, on which a four-parameter family of complex lines can fit. This result has a real analogue. There are plenty of ruled surfaces in three-dimensional space that contain a one-parameter family of lines, but only the hyperboloid of one sheet contains two different families (among all nonflat surfaces the hyperbolic paraboloid $X_3 = X_1^2 - X_2^2$ also has this property, but from the projective point of view it is equivalent to a hyperboloid of one sheet).

Let us sum up. We started from the quadric of real lines in $\mathbb{P}^3$ and then passed to the quadric of complex lines in $\mathbb{CP}^3$. Among the real second-order surfaces that lie on this complex surface there are the surface of real lines and two other types of surfaces: some surfaces correspond to fibrations of $\mathbb{CP}^3$ into complex lines, while others correspond to five-dimensional real surfaces in $\mathbb{CP}^3$ containing a family of complex lines that depend on four real parameters. This example distinctly shows a phenomenon through which the 19th century geometers "suffered." First, real objects often admit a complex interpretation. Second, when we complexify a real problem and then, conversely, see which real problems lead to the same complex ones, we often obtain new meaningful geometrical problems.

A Metric on the Manifold of Lines

Plücker and his successors were also concerned with the geometry of the manifold of lines $Q \subset \mathbb{CP}^5$. They studied how to express various geometric facts about the initial projective space $\mathbb{CP}^3$ in terms of Q. To points in $\mathbb{CP}^3$ there correspond two-dimensional surfaces of lines in Q that pass through these points; to planes in $\mathbb{CP}^3$ there correspond two-dimensional surfaces of lines lying in these planes (two families of flat generators of the quadric Q). The reverse direction was also fruitful, considering families of lines in $\mathbb{CP}^3$ whose Plücker coordinates satisfy one relation (complexes) or two (congruences). Here is an example of such a fact.

Lines in three-dimensional space sometimes intersect. How do we express that fact in Plücker coordinates? It turns out that if $\{p_{ij}\}$ and $\{p'_{ij}\}$ are Plücker coordinates of two lines, then the lines intersect if and only if

$$p_{01}p'_{23} - p_{02}p'_{13} + p_{03}p'_{12} + p_{23}p'_{01} - p_{13}p'_{+02} + p_{12}p'_{03} = 0. \tag{7}$$

We will deduce (7) under the simplifying assumption (which was already made) that $p_{03} \neq 0, p'_{03} \neq 0$. Then $p_{03} = 1, p'_{03} = 1$ and the lines join $(1, \alpha_1, \alpha_2, 0)$, $(0, \beta_1, \beta_2, 1)$ and $(1, \alpha'_1, \alpha'_2, 0)$, $(0, \beta'_1, \beta'_2, 1)$, respectively (actually we have passed from homogeneous to inhomogeneous coordinates). In this case the points of the line p are given by the equations

$$z_1 = \alpha_1 z_0 + \beta_1 z_3, \qquad z_2 = \alpha_2 z_0 + \beta_2 z_3,$$

and those of p' analogously by

$$z_1 = \alpha'_1 z_0 + \beta'_1 z_3, \qquad z_2 = \alpha'_2 z_0 + \beta'_2 z_3.$$

The lines intersect if there is a common solution (z_0, z_1, z_2, z_3) to this system of four equations, or to the system of two equations in two unknowns (z_0, z_3):

$$\begin{cases} z_0(\alpha_1 - \alpha'_1) + z_3(\alpha_2 - \alpha'_2) = 0, \\ z_0(\beta_1 - \beta'_1) + z_3(\beta_2 - \beta'_2) = 0. \end{cases} \tag{8}$$

Thus, lines intersect if and only if this expression vanishes:

$$\rho(\alpha, \beta, \alpha', \beta') = (\alpha_1 - \alpha'_1)(\beta_2 - \beta'_2) - (\alpha_2 - \alpha'_2)(\beta_1 - \beta'_1). \tag{9}$$

A modern mathematician would call expression (9) a "distance." True, (9) may vanish when $p \neq p'$ and is in general a complex number. But this was not taboo even for geometers in the 19th century. Klein recalls how they liked to use lines along which the distance was zero (isotropic lines). Lie called these lines "crazy" and used to say that French geometers knew how

to use them to obtain proofs "out of thin air." We too will call the quantity ρ the distance between the lines $p = (\alpha, \beta)$ and $p' = (\alpha', \beta')$.

Thus, the distance ρ vanishes if and only if the lines intersect. This condition defines the distance almost uniquely. More exactly, the distance is defined up to a conformal change (homothety). This means that angles and ratios of distances in a neighborhood of any fixed point are uniquely defined up to values which are small compared to the distance to this point.

To each point $p \in Q$ we assign the set of points $V_p \subset Q$ that are at zero distance from p, i.e., $\rho(p, p') = 0$ and the lines p, p' intersect. The set V_p is called an isotropy cone and coincides with the intersection of the quadric Q with the tangent plane to Q at the point p.

Distances on (S) and (H)

Let us follow the trail of the distance ρ on the surface (S). We again restrict ourselves to points where $p_{03} \neq 1$. Then condition (4) implies that $\beta_1 = \bar{\alpha}_2$, $\beta_2 = -\alpha_1$, and we can take only the pair of complex numbers $(\alpha_1\alpha_2)$ as coordinates on (S). Thus

$$\rho_{(S)}(\alpha; \alpha') = |\alpha_1 - \alpha_1'|^2 + |\alpha_2 - \alpha_2'|^2. \tag{10}$$

This distance is missing all the shortcomings of a general ρ: It is nonnegative and vanishes only when $\alpha = \alpha'$. This agrees with the fact that lines corresponding to points of (S) do not intersect. We have obtained the usual Euclidian distance on the four-dimensional real sphere in five-dimensional Euclidian space.

Now let us restrict ρ to the hyperboloid (H) and again take points with $p_{03} = 1$. Let $M \subset$ (H) be the set of such points on (H). Then by (5), α_1, β_2 are real and $\beta_1 = \bar{\alpha}_2$. Make the substitution $\alpha_1 = t - x_1$, $\beta_2 = t + x_1$, $\beta_1 = x_2 + ix_3$, where all (t, x) are real. As a result expression (9) takes the form:

$$\rho_{(H)}(t, x; t', x') = (t - t')^2 - (x_1 - x_1')^2 - (x_2 - x_2')^2 - (x_3 - x_3')^2. \tag{11}$$

This is exactly the Minkowski metric. (It is real but not positive definite.) If we take $p \in M$ and intersect the cone V_p with the surface M, we obtain the light cone with vertex p. Thus, the distance on the quadric Q that arises naturally from the geometry of lines induces the Euclidean distance on the sphere (S) and the Minkowski distance on the hyperboloid (H).

The points of $M \subset$ (H) correspond to the lines on the surface N that do not intersect the line $z_0 = z_3 = 0$. The manifold (H) plays an important role in physical theories, being a conformal extension of the Minkowski space M. It is obtained from M by adding a light cone at "infinity" (similar to how

Euclidean space may be extended by adding a point at infinity rather than the whole plane at infinity as for a projective extension). Let us consider projective transformations of $\mathbb{CP}^3$ that preserve N. They transform lines on N into lines on N and intersecting lines into intersecting lines. By the same token, transformations are induced on (H) that map light cones V_p to each other. This is how we obtain all conformal transformations of Minkowski space (shifts, homotheties, inversions), with respect to which (massless) physical theories are often invariant. To obtain the group of proper motions (the Poincaré group) we must restrict ourselves to transformations that also preserve the line $z_0 = z_3 = 0$. Therefore, the geometry of Minkowski space arises completely within the realm of Plücker's geometry of the space of lines.

Is there a natural way to go backwards? When studying Minkowski space, how do we find the auxiliary three-dimensional space (the space of Penrose twistors) whose lines correspond to the points of Minkowski space? We can do this using the light cones V_p. Recall that lines intersecting the line p correspond to points of V_p. All lines corresponding to points on the same generator of V_p (to the light line) intersect p in the same point. Therefore, there is a correspondence between points of N and light lines, so that N may be considered as the set of light lines on (H). In the complex picture, points of $\mathbb{CP}^3$ are identified with complex "light" lines on Q (with half the two-dimensional generators of V_p).

A Remark on Analytical Applications

The geometric picture we have presented is no doubt instructive. But as we have already noted, it is only a means for new analytical constructions within the Penrose theory. Sad to say, we can only touch on this superficially. Penrose's idea is that to analytic objects on the four-dimensional manifold M (on (S) in the Euclidean theory) there must correspond some equivalent objects on N or $\mathbb{CP}^3$. The latter must be simpler than their counterparts on M and (S) and a significant portion of the equations of mathematical physics on M and (S) are just corollaries of the fact that objects, initially defined on a three-dimensional manifold, are somehow translated onto a four-dimensional manifold. We must note that many differential equations arise as relations when we pass (by means of an integral transform) to manifolds of higher dimension. This is an important and as yet inadequately studied source for obtaining and solving equations. In the simplest example, due to Fritz John (1910–1994), we integrate a function along lines in three-dimensional (real) space and obtain solutions of a (ultrahyperbolic) second-order differential equation in the four-dimensional

space of lines. Penrose and his followers encountered similar effects in the more complicated complex situation. Therefore, they had to deal not with functions but with a much more complicated object, namely, cohomology. It turned out that in passing from M and (S) to $\mathbb{CP}^3$ they actually obtained simpler and more classical equations, a variant of the Cauchy–Riemann equations in the theory of analytic functions. This approach embraces not only linear equations of mathematical physics (Dirac–Weyl, Maxwell, the linearized Einstein equation) but also some nonlinear ones (Yang–Mills).

Self-Dual Metrics

In concluding, we will dwell further on one direction of Penrose's work. So far we have dealt with flat Minkowski space-time. In the general theory of relativity we are interested in distortions of four-dimensional manifolds that must satisfy strong nonlinear restrictions (e.g., Einstein's vacuum equation). Constructing solutions to Einstein's equation is a difficult problem. Penrose, starting from the realization of Minkowski space as a family of lines in $\mathbb{CP}^3$, looked for manifolds that would satisfy Einstein's equation as families of lines on some three-dimensional manifolds. A metric here would have to be obtained from the condition on intersecting curves (the distance between intersecting curves is zero). From the start he limited himself to the complex case. There are additional reasons for this in the nonflat case, namely that on the manifold of curves we do not have the nonflat Einstein metric with signature (3.1), like the Minkowski metric, but rather the Riemann metric (4.0).

Passing to the complex case significantly simplifies the situation and makes it more geometric. A number of invariants of the components of the curvature of the manifold that are introduced analytically in the real case acquire a clear geometric meaning in the complex case (Ricci and Weyl tensors). Penrose showed that a certain class of complex solutions of Einstein's equation (self-dual) are obtained if we perturb the complex structure in a particular way in a neighborhood of some line in $\mathbb{CP}^3$ and consider a certain family of curves that are "close to" lines. Unfortunately, there is an aspect of this approach that is especially ineffective for finding the family of curves. However, in certain cases the calculations can lead to a specific expression for the metric.

Some other geometric ideas appeared later for constructing specific solutions to nonlinear equations, including the Einstein equation, using the language of twistors. One such idea consists of the fact that in an eight-parameter family of second-order curves in $\mathbb{CP}^3$ there are four-parameter subfamilies on which the intersection conditions induce the Einstein met-

ric. In this way we obtain certain well-known solutions and many new ones as well. This idea is completely in line with Plücker's ideology: The intersection condition for lines gives a flat metric, and by using conics we construct nonflat metrics.

The ideas of the twistor program have undergone considerable development in recent years, although perhaps the initial hopes for the role of twistors in theoretical physics have turned out to be overly optimistic. Within mathematics twistors have found notable application to multidimensional complex analysis, but most of all to geometry and topology, where they led to a revolution in the theory of four-dimensional manifolds.