Menninger, K. 1992. *Number Words and Number Symbols: A Cultural History of Numbers.* New York: Dover. (Translated by P. Broneer from the revised German edition of 1957/58: *Zahlwort und Ziffer. Eine Kulturgeschichte der Zahl.* Göttingen: Vandenhoeck und Ruprecht.)
Reid, C. 2006. *From Zero to Infinity: What Makes Numbers Interesting.* Natick, MA: A. K. Peters.

II.2 Geometry

Jeremy Gray

1 Introduction

The modern view of geometry was inspired by the novel geometrical theories of HILBERT [VI.63] and Einstein in the early years of the twentieth century, which built in their turn on other radical reformulations of geometry in the nineteenth century. For thousands of years, the geometrical knowledge of the Greeks, as set out most notably in EUCLID's [VI.2] *Elements*, was held up as a paradigm of perfect rigor, and indeed of human knowledge. The new theories amounted to the overthrow of an entire way of thinking. This essay will pursue the history of geometry, starting from the time of Euclid, continuing with the advent of non-Euclidean geometry, and ending with the work of RIEMANN [VI.49], KLEIN [VI.57], and POINCARÉ [VI.61]. Along the way, we shall examine how and why the notions of geometry changed so remarkably. Modern geometry itself will be discussed in later parts of this book.

2 Naive Geometry

Geometry generally, and Euclidean geometry in particular, is informally and rightly taken to be the mathematical description of what you see all around you: a space of three dimensions (left–right, up–down, forwards–backwards) that seems to extend indefinitely far. Objects in it have positions, they sometimes move around and occupy other positions, and all of these positions can be specified by measuring lengths along straight lines: this object is twenty meters from that one, it is two meters tall, and so on. We can also measure angles, and there is a subtle relationship between angles and lengths. Indeed, there is another aspect to geometry, which we do not see but which we reason about. Geometry is a mathematical subject that is full of *theorems*—the isosceles triangle theorem, the Pythagorean theorem, and so on—which collectively summarize what we can say about lengths, angles, shapes, and positions. What distinguishes this aspect of geometry from most other kinds of science is its highly deductive nature. It really seems that by taking the simplest of concepts and thinking hard about them one can build up an impressive, deductive body of knowledge about space without having to gather experimental evidence.

But can we? Is it really as simple as that? Can we have genuine knowledge of space without ever leaving our armchairs? It turns out that we cannot: there are other geometries, also based on the concepts of length and angle, that have every claim to be useful, but that disagree with Euclidean geometry. This is an astonishing discovery of the early nineteenth century, but, before it could be made, a naive understanding of fundamental concepts, such as straightness, length, and angle, had to be replaced by more precise definitions—a process that took many hundreds of years. Once this had been done, first one and then infinitely many new geometries were discovered.

3 The Greek Formulation

Geometry can be thought of as a set of useful facts about the world, or else as an organized body of knowledge. Either way, the origins of the subject are much disputed. It is clear that the civilizations of Egypt and Babylonia had at least some knowledge of geometry—otherwise, they could not have built their large cities, elaborate temples, and pyramids. But not only is it difficult to give a rich and detailed account of what was known before the Greeks, it is difficult even to make sense of the few scattered sources that we have from before the time of Plato and Aristotle. One reason for this is the spectacular success of the later Greek writer, and author of what became the definitive text on geometry, Euclid of Alexandria (ca. 300 B.C.E.). One glance at his famous *Elements* shows that a proper account of the history of geometry will have to be about something much more than the acquisition of geometrical facts. The *Elements* is a highly organized, deductive body of knowledge. It is divided into a number of distinct themes, but each theme has a complex theoretical structure. Thus, whatever the origins of geometry might have been, by the time of Euclid it had become the paradigm of a logical subject, offering a kind of knowledge quite different from, and seemingly higher than, knowledge directly gleaned from ordinary experience.

Rather, therefore, than attempt to elucidate the early history of geometry, this essay will trace the high road of geometry's claim on our attention: the apparent certainty of mathematical knowledge. It is exactly this claim to a superior kind of knowledge that led eventually to the remarkable discovery of *non-Euclidean geometry*: there are geometries other than Euclid's that are every bit as rigorously logical. Even more remarkably, some of these turn out to provide better models of physical space than Euclidean geometry.

The *Elements* opens with four books on the study of plane figures: triangles, quadrilaterals, and circles. The famous theorem of Pythagoras is the forty-seventh proposition of the first book. Then come two books on the theory of ratio and proportion and the theory of similar figures (scale copies), treated with a high degree of sophistication. The next three books are about whole numbers, and are presumably a reworking of much older material that would now be classified as elementary number theory. Here, for example, one finds the famous result that there are infinitely many prime numbers. The next book, the tenth, is by far the longest, and deals with the seemingly specialist topic of lengths of the form $\sqrt{a} \pm \sqrt{b}$ (to write them as we would). The final three books, where the curious lengths studied in Book X play a role, are about three-dimensional geometry. They end with the construction of the five regular solids and a proof that there are no more. The discovery of the fifth and last had been one of the topics that excited Plato. Indeed, the five regular solids are crucial to the cosmology of Plato's late work the *Timaeus*.

Most books of the *Elements* open with a number of definitions, and each has an elaborate deductive structure. For example, to understand the Pythagorean theorem, one is driven back to previous results, and thence to even earlier results, until finally one comes to rest on basic definitions. The whole structure is quite compelling: reading it as an adult turned the philosopher Thomas Hobbes from incredulity to lasting belief in a single sitting. What makes the *Elements* so convincing is the nature of the arguments employed. With some exceptions, mostly in the number-theoretic books, these arguments use the axiomatic method. That is to say, they start with some very simple axioms that are intended to be self-evidently true, and proceed by purely logical means to deduce theorems from them.

For this approach to work, three features must be in place. The first is that *circularity* should be carefully avoided. That is, if you are trying to prove a statement P and you deduce it from an earlier statement, and deduce that from a yet earlier statement, and so on, then at no stage should you reach the statement P again. That would not prove P from the axioms, but merely show that all the statements in your chain were equivalent. Euclid did a remarkable job in this respect.

The second necessary feature is that the rules of inference should be clear and acceptable. Some geometrical statements seem so obvious that one can fail to notice that they need to be proved: ideally, one should use no properties of figures other than those that have been clearly stated in their definitions, but this is a difficult requirement to meet. Euclid's success here was still impressive, but mixed. On the one hand, the *Elements* is a remarkable work, far outstripping any contemporary account of any of the topics it covers, and capable of speaking down the millennia. On the other, it has little gaps that from time to time later commentators would fill. For example, it is neither explicitly assumed nor proved in the *Elements* that two circles will meet if their centers lie outside each other and the sum of their radii is greater than the distance between their centers. However, Euclid is surprisingly clear that there are rules of inference that are of general, if not indeed universal, applicability, and others that apply to mathematics because they rely on the meanings of the terms involved.

The third feature, not entirely separable from the second, is adequate definitions. Euclid offered two, or perhaps three, sorts of definition. Book I opens with seven definitions of objects, such as "point" and "line," that one might think were primitive and beyond definition, and it has recently been suggested that these definitions are later additions. Then come, in Book I and again in many later books, definitions of familiar figures designed to make them amenable to mathematical reasoning: "triangle," "quadrilateral," "circle," and so on. The postulates of Book I form the third class of definition and are rather more problematic.

Book I states five "common notions," which are rules of inference of a very general sort. For example, "If equals be added to equals, the wholes are equals." The book also has five "postulates," which are more narrowly mathematical. For example, the first of these asserts that one may draw a straight line from any point to any point. One of these postulates, the fifth, became notorious: the so-called *parallel postulate*. It says that "If a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which are the angles less than two right angles."

Parallel lines, therefore, are straight lines that do not meet. A helpful rephrasing of Euclid's parallel postulate was introduced by the Scottish editor, Robert Simson. It appears in his edition of Euclid's *Elements* from 1806. There he showed that the parallel postulate is equivalent, if one assumes those parts of the *Elements* that do not depend on it, to the following statement: given any line m in a plane, and any point P in that plane that does not lie on the line m, there is exactly one line n in the plane that passes through the point P and does not meet the line m. From this formulation it is clear that the parallel postulate makes two assertions: given a line and a point as described, a parallel line *exists* and it is *unique*.

It is worth noting that Euclid himself was probably well aware that the parallel postulate was awkward. It asserts a property of straight lines that seems to have made Greek mathematicians and philosophers uncomfortable, and this may be why its appearance in the *Elements* is delayed until proposition 29 of Book I. The commentator Proclus (fifth century C.E.), in his extensive discussion of Book I of the *Elements*, observed that the hyperbola and asymptote get closer and closer as they move outwards, but they never meet. If a line and a curve can do this, why not two lines? The matter needs further analysis. Unfortunately, not much of the *Elements* would be left if mathematicians dropped the parallel postulate and retreated to the consequences of the remaining definitions: a significant body of knowledge depends on it. Most notably, the parallel postulate is needed to prove that the angles in a triangle add up to two right angles—a crucial result in establishing many other theorems about angles in figures, including the Pythagorean theorem.

Whatever claims educators may have made about Euclid's *Elements* down the ages, a significant number of experts knew that it was an unsatisfactory compromise: a useful and remarkably rigorous theory could be had, but only at the price of accepting the parallel postulate. But the parallel postulate was difficult to accept on trust: it did not have the same intuitively obvious feel of the other axioms and there was no obvious way of verifying it. The higher one's standards, the more painful this compromise was. What, the experts asked, was to be done?

One Greek discussion must suffice here. In Proclus's view, if the truth of the parallel postulate was not obvious, and yet geometry was bare without it, then the only possibility was that it was true because it was a theorem. And so he gave it a proof. He argued as follows. Let two lines m and n cross a third line k at P and Q, respectively, and make angles with it that add up to two right angles. Now draw a line l that crosses m at P and enters the space between the lines m and n. The distance between l and m as one moves away from the point P continually increases, said Proclus, and therefore line l must eventually cross line n.

Proclus's argument is flawed. The flaw is subtle, and sets us up for what is to come. He was correct that the distance between the lines l and m increases indefinitely. But his argument assumes that the distance between lines m and n does not *also* increase indefinitely, and is instead bounded. Now Proclus knew very well that *if* the parallel postulate is granted, *then* it can be shown that the lines m and n are parallel and that the distance between them is a constant. But until the parallel postulate is proved, nothing prevents one saying that the lines m and n diverge. Proclus's proof does not therefore work unless one can show that lines that do not meet also do not diverge.

Proclus's attempt was not the only one, but it is typical of such arguments, which all have a standard form. They start by detaching the parallel postulate from Euclid's *Elements*, together with all the arguments and theorems that depend on it. Let us call what remains the "core" of the *Elements*. Using this core, an attempt is then made to derive the parallel postulate as a theorem. The correct conclusion to be derived from Proclus's attempt is not that the parallel postulate is a theorem, but rather that, given the core of the *Elements*, the parallel postulate is equivalent to the statement that lines that do not meet also do not diverge. Aganis, a writer of the sixth century C.E. about whom almost nothing is known, assumed, in a later attempt, that parallel lines are everywhere equidistant, and his argument showed only that, given the core, the Euclidean definition of parallel lines is equivalent to defining them to be equidistant.

Notice that one cannot even enter this debate unless one is clear which properties of straight lines belong to them by definition, and which are to be derived as theorems. If one is willing to add to the store of "commonsense" assumptions about geometry as one goes along, the whole careful deductive structure of the *Elements* collapses into a pile of facts.

This deductive character of the *Elements* is clearly something that Euclid regarded as important, but one can also ask what he thought geometry was *about*. Was it meant, for example, as a mathematical description of space? No surviving text tells us what he thought about this question, but it is worth noting that the most celebrated Greek theory of the universe, developed by Aristotle and many later commentators, assumed that space was finite, bounded by the sphere of the fixed stars. The mathematical space of the *Elements* is infinite, and so one has at least to consider the possibility that, for all these writers, mathematical space was not intended as a simple idealization of the physical world.

4 Arab and Islamic Commentators

What we think of today as Greek geometry was the work of a handful of mathematicians, mostly concentrated in a period of less than two centuries. They were eventually succeeded by a somewhat larger number of Arabic and Islamic writers, spread out over a much greater area and a longer time. These writers tend to be remembered as commentators on Greek mathematics and science, and for transmitting them to later Western authors, but they should also be remembered as creative, innovative mathematicians and scientists in their own right. A number of them took up the study of Euclid's *Elements*, and with it the problem of the parallel postulate. They too took the view that it was not a proper postulate, but one that could be proved as a theorem using the core alone.

Among the first to attempt a proof was Thābit ibn Qurra. He was a pagan from near Aleppo who lived and worked in Baghdad, where he died in 901. Here there is room to describe only his first approach. He argued that if two lines m and n are crossed by a third, k, and if they approach each other on one side of the line k, then they diverge indefinitely on the other side of k. He deduced that two lines that make equal alternate angles with a transversal (the marked angles in figure 1) cannot approach each other on one side of a transversal: the symmetry of the situation would imply that they approached on the other side as well, but he had shown that they would have to diverge on the other side. From this he deduced the Euclidean theory of parallels, but his argument was also flawed, since he had not considered the possibility that two lines could *diverge* in both directions.

The distinguished Islamic mathematician and scientist ibn al-Haytham was born in Basra in 965 and died in Egypt in 1041. He took a quadrilateral with two equal sides perpendicular to the base and dropped a perpendicular from one side to the other. He now attempted to prove that this perpendicular is equal to the base, and to do so he argued that as one of two original perpendiculars is moved toward the other, its tip sweeps out a straight line, which will coincide with the perpendicular just dropped (see figure 2). This amounts to the assumption that the curve everywhere equidistant from a straight line is itself straight, from which the parallel postulate easily follows, and so his attempt fails. His proof was later heavily criticized by Omar Khayyam for its use of motion, which he found fundamentally unclear and alien to Euclid's *Elements*. It is indeed quite distinct from any use Euclid had for motion in geometry, because in this case the nature of the curve obtained is not clear: it is precisely what needs to be analyzed.

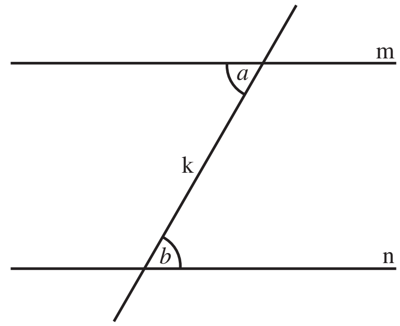

Figure 1 The lines m and n make equal alternate angles *a* and *b* with the transversal k.

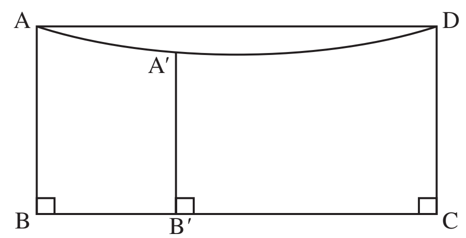

Figure 2 AB and CD are equal, the angle ADC is a right angle, A′B′ is an intermediate position of AB as it moves toward CD.

The last of the Islamic attempts on the parallel postulate is due to Naṣır al-Dīn al-Ṭūsī. He was born in Iran in 1201 and died in Baghdad in 1274. His extensive commentary is also one of our sources of knowledge of earlier Islamic mathematical work on this subject. Al-Ṭūsī focused on showing that if two lines begin to converge, then they must continue to do so until they eventually meet. To this end he set out to show that

(∗) if l and m are two lines that make an angle of less than a right angle, then every line perpendicular to l meets the line m.

He showed that if ($*$) is true, then the parallel postulate follows. However, his argument for ($*$) is flawed.

It is genuinely difficult to see what is wrong with some of these arguments if one uses only the techniques available to mathematicians of the time. Islamic mathematicians showed a degree of sophistication that was not to be surpassed by their Western successors until the eighteenth century. Unfortunately, however, their writings did not come to the attention of the West until much later, with the exception of a single work in the Vatican Library, published in 1594, which was for many years erroneously attributed to al-Ṭūsī (and which may have been the work of his son).

5 The Western Revival of Interest

The Western revival of interest in the parallel postulate came with the second wave of translations of Greek mathematics, led by Commandino and Maurolico in the sixteenth century and spread by the advent of printing. Important texts were discovered in a number of older libraries, and ultimately this led to the production of new texts of Euclid's *Elements*. Many of these had something to say about the problem of parallels, pithily referred to by Henry Savile as "a blot on Euclid." For example, the powerful Jesuit Christopher Clavius, who edited and reworked the *Elements* in 1574, tried to argue that parallel lines could be defined as equidistant lines.

The ready identification of physical space with the space of Euclidean geometry came about gradually during the sixteenth and seventeenth centuries, after the acceptance of Copernican astronomy and the abolition of the so-called sphere of fixed stars. It was canonized by NEWTON [VI.14] in his *Principia Mathematica*, which proposed a theory of gravitation that was firmly situated in Euclidean space. Although Newtonian physics had to fight for its acceptance, Newtonian cosmology had a smooth path and became the unchallenged orthodoxy of the eighteenth century. It can be argued that this identification raised the stakes, because any unexpected or counterintuitive conclusion drawn solely from the core of the *Elements* was now, possibly, a counterintuitive fact about space.

In 1663 the English mathematician John Wallis took a much more subtle view of the parallel postulate than any of his predecessors. He had been instructed by Halley, who could read Arabic, in the contents of the apocryphal edition of al-Ṭūsī's work in the Vatican Library, and he too gave an attempted proof. Unusually, Wallis also had the insight to see where his own argument was flawed, and commented that what it really showed was that, in the presence of the core, the parallel postulate was equivalent to the assertion that there exist similar figures that are not congruent.

Half a century later, Wallis was followed by the most persistent and thoroughgoing of all the defenders of the parallel postulate, Gerolamo Saccheri, an Italian Jesuit who published in 1733, the year of his death, a short book called *Euclid Freed of Every Flaw*. This little masterpiece of classical reasoning opens with a trichotomy. Unless the parallel postulate is known, the angle sum of a triangle may be either less than, equal to, or greater than two right angles. Saccheri showed that whatever happens in one triangle happens for them all, so there are apparently three geometries compatible with the core. In the first, every triangle has an angle sum less than two right angles (call this case L). In the second, every triangle has an angle sum equal to two right angles (call this case E). In the third, every triangle has an angle sum greater than two right angles (call this case G). Case E is, of course, Euclidean geometry, which Saccheri wished to show was the only case possible. He therefore set to work to show that each of the other cases independently self-destructed. He was successful with case G, and then turned to case L "which alone obstructs the truth of the [parallel] axiom," as he put it.

Case L proved to be difficult, and during the course of his investigations Saccheri established a number of interesting propositions. For example, if case L is true, then two lines that do not meet have just one common perpendicular, and they diverge on either side of it. In the end, Saccheri tried to deal with his difficulties by relying on foolish statements about the behavior of lines at infinity: it was here that his attempted proof failed.

Saccheri's work sank slowly, though not completely, into obscurity. It did, however, come to the attention of the Swiss mathematician Johann Lambert, who pursued the trichotomy but, unlike Saccheri, stopped short of claiming success in proving the parallel postulate. Instead the work was abandoned, and was published only in 1786, after his death. Lambert distinguished carefully between unpalatable results and impossibilities. He had a sketch of an argument to show that in case L the area of a triangle is proportional to the difference between two right angles and the angle sum of the triangle. He knew that in case L similar triangles had to be congruent, which would imply that the tables of trigonometric functions used in astronomy were not in fact valid and that different tables would have to be produced for every size of triangle. In particular, for every angle less than 60° there would be precisely one equilateral triangle with that given angle at each vertex. This would lead to what philosophers called an "absolute" measure of length (one could take, for instance, the length of the side of an equilateral triangle with angles equal to 30°), which LEIBNIZ's [VI.15] follower Wolff had said was impossible. And indeed it is counterintuitive: lengths are generally defined in relative terms, as, for instance, a certain proportion of the length of a meter rod in Paris, or of the circumference of Earth, or of something similar. But such arguments, said Lambert, "were drawn from love and hate, with which a mathematician can have nothing to do."

6 The Shift of Focus around 1800

The phase of Western interest in the parallel postulate that began with the publication of modern editions of Euclid's *Elements* started to decline with a further turn in that enterprise. After the French revolution, LEGENDRE [VI.24] set about writing textbooks, largely for the use of students hoping to enter the École Polytechnique, that would restore the study of elementary geometry to something like the rigorous form in which it appeared in the *Elements*. However, it was one thing to seek to replace books of a heavily intuitive kind, but quite another to deliver the requisite degree of rigor. Legendre, as he came to realize, ultimately failed in his attempt. Specifically, like everyone before him, he was unable to give an adequate defense of the parallel postulate. Legendre's *Éléments de Géométrie* ran to numerous editions, and from time to time a different attempt on the postulate was made. Some of these attempts would be hard to describe favorably, but the best can be extremely persuasive.

Legendre's work was classical in spirit, and he still took it for granted that the parallel postulate had to be true. But by around 1800 this attitude was no longer universally held. Not everybody thought that the postulate must, somehow, be defended, and some were prepared to contemplate with equanimity the idea that it might be false. No clearer illustration of this shift can be found than a brief note sent to GAUSS [VI.26] by F. K. Schweikart, a Professor of Law at the University of Marburg, in 1818. Schweikart described in a page the main results he had been led to in what he called "astral geometry," in which the angle sum of a triangle was less than two right angles: squares had a particular form, and the altitude of a right-angled isosceles triangle was bounded by an amount Schweikart called "the constant." Schweikart went so far as to claim that the new geometry might even be the true geometry of space. Gauss replied positively. He accepted the results, and he claimed that he could do all of elementary geometry once a value for the constant was given. One could argue, somewhat ungenerously, that Schweikart had done little more than read Lambert's posthumous book—although the theorem about isosceles triangles is new. However, what is notable is the attitude of mind: the idea that this new geometry might be true, and not just a mathematical curiosity. Euclid's *Elements* shackled him no more.

Unfortunately, it is much less clear what Gauss himself thought. Some historians, bearing Gauss's remarkable mathematical originality in mind, have been inclined to interpret the evidence in such a way that Gauss emerges as the first person to discover non-Euclidean geometry. However, the evidence is slight, and it is difficult to draw firm conclusions from it. There are traces of some early investigations by Gauss of Euclidean geometry that include a study of a new definition of parallel lines; there are claims made by Gauss late in life that he had known this or that fact for many years; and there are letters he wrote to his friends. But there is no material in the surviving papers that allows us to reconstruct what Gauss knew or that supports the claim that Gauss discovered non-Euclidean geometry.

Rather, the picture would seem to be that Gauss came to realize during the 1810s that all previous attempts to derive the parallel postulate from the core of Euclidean geometry had failed and that all future attempts would probably fail as well. He became more and more convinced that there was another possible geometry of space. Geometry ceased, in his mind, to have the status of arithmetic, which was a matter of logic, and became associated with mechanics, an empirical science. The simplest accurate statement of Gauss's position through the 1820s is that he did not doubt that space might be described by a non-Euclidean geometry, and of course there was only one possibility: that of case L described above. It was an empirical matter, but one that could not be resolved by land-based measurements because any departure from Euclidean geometry was, evidently, very small. In this view he was supported by his friends, such as Bessel and Olbers, both professional astronomers. Gauss the scientist was convinced, but Gauss the mathematician may have retained a small degree of doubt, and certainly never developed the mathematical theory required to describe non-Euclidean geometry adequately.

One theory available to Gauss from the early 1820s was that of differential geometry. Gauss eventually published one of his masterworks on this subject, his *Disquisitiones Generales circa Superficies Curvas* (1827). In it he showed how to describe geometry on any surface in space, and how to regard certain features of the geometry of a surface as intrinsic to the surface and independent of how the surface was embedded into three-dimensional space. It would have been possible for Gauss to consider a surface of constant negative CURVATURE [III.78], and to show that triangles on such a surface are described by hyperbolic trigonometric formulas, but he did not do this until the 1840s. Had he done so, he would have had a surface on which the formulas of a geometry satisfying case L apply.

A surface, however, is not enough. We accept the validity of two-dimensional Euclidean geometry because it is a simplification of three-dimensional Euclidean geometry. Before a two-dimensional geometry satisfying the hypotheses of case L can be accepted, it is necessary to show that there is a plausible three-dimensional geometry analogous to case L. Such a geometry has to be described in detail and shown to be as plausible as Euclidean three-dimensional geometry. This Gauss simply never did.

7 Bolyai and Lobachevskii

The fame for discovering non-Euclidean geometry goes to two men, BOLYAI [VI.34] in Hungary and LOBACHEVSKII [VI.31] in Russia, who independently gave very similar accounts of it. In particular, both men described a system of geometry in two and three dimensions that differed from Euclid's but had an equally good claim to be the geometry of space. Lobachevskii published first, in 1829, but only in an obscure Russian journal, and then in French in 1837, in German in 1840, and again in French in 1855. Bolyai published his account in 1831, in an appendix to a two-volume work on geometry by his father.

It is easiest to describe their achievements together. Both men defined parallels in a novel way, as follows. Given a point P and a line m there will be some lines through P that meet m and others that do not. Separating these two sets will be two lines through P that do not quite meet m but which might come arbitrarily close, one to the right of P and one to the left. This situation is illustrated in figure 3: the two lines in question are n′ and n″. Notice that lines on the diagram appear curved. This is because, in order to represent them on a flat, Euclidean page, it is necessary to distort them, unless the geometry is itself Euclidean, in which case one can put n′ and n″ together and make a single line that is infinite in both directions.

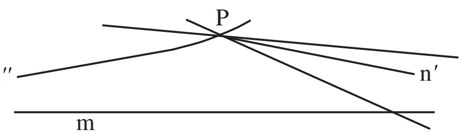

Figure 3 The lines n′ and n″ through P separate the lines through P that meet the line m from those that do not.

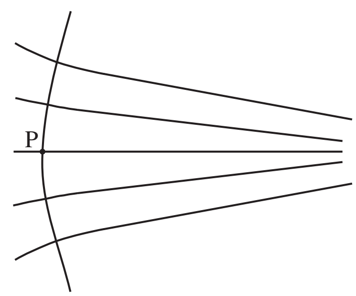

Figure 4 A curve perpendicular to a family of parallels.

Given this new way of talking, it still makes sense to talk of dropping the perpendicular from P to the line m. The left and right parallels to m through P make equal angles with the perpendicular, called the *angle of parallelism*. If the angle is a right angle, then the geometry is Euclidean. However, if it is less than a right angle, then the possibility arises of a new geometry. It turns out that the size of the angle depends on the length of the perpendicular from P to m. Neither Bolyai nor Lobachevskii expended any effort in trying to show that there was not some contradiction in taking the angle of parallelism to be less than a right angle. Instead, they simply made the assumption and expended a great deal of effort on determining the angle from the length of the perpendicular.

They both showed that, given a family of lines all parallel (in the same direction) to a given line, and given a point on one of the lines, there is a curve through that point that is perpendicular to each of the lines (figure 4).

In Euclidean geometry the curve defined in this way is the straight line that is at right angles to the family of parallel lines and that passes through the given

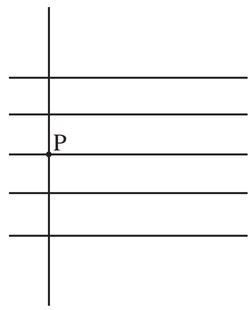

Figure 5 A curve perpendicular to a family of Euclidean parallels.

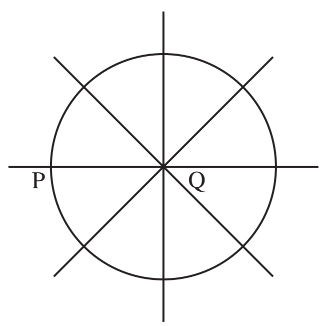

Figure 6 A curve perpendicular to a family of Euclidean lines through a point.

point (figure 5). If, again in Euclidean geometry, one takes the family of all lines through a common point Q and chooses another point P, then there will be a curve through P that is perpendicular to all the lines: the circle with center Q that passes through P (figure 6).

The curve defined by Bolyai and Lobachevskii has some of the properties of both these Euclidean constructions: it is perpendicular to all the parallels, but it is curved and not straight. Bolyai called such a curve an *L-curve*. Lobachevskii more helpfully called it a *horocycle*, and the name has stuck.

Their complicated arguments took both men into three-dimensional geometry. Here Lobachevskii's arguments were somewhat clearer than Bolyai's, and both men notably surpassed Gauss. If the figure defining a horocycle is rotated about one of the parallel lines, the lines become a family of parallel lines in three dimensions and the horocycle sweeps out a bowl-shaped surface, called the *F-surface* by Bolyai and the horosphere by Lobachevskii. Both men now showed that something remarkable happens. Planes through the horosphere cut it either in circles or in horocycles, and if a triangle is drawn on a horosphere whose sides are horocycles, then the angle sum of such a triangle is two right angles. To put this another way, although the space that contains the horosphere is a three-dimensional version of case L, and is definitely not Euclidean, the geometry you obtain when you restrict attention to the horosphere is (two-dimensional) Euclidean geometry!

Bolyai and Lobachevskii also knew that one can draw spheres in their three-dimensional space, and they showed (though in this they were not original) that the formulas of spherical geometry hold independently of the parallel postulate. Lobachevskii now used an ingenious construction involving his parallel lines to show that a triangle on a sphere determines and is determined by a triangle in the plane, which also determines and is determined by a triangle on the horosphere. This implies that the formulas of spherical geometry must determine formulas that apply to the triangles on the horosphere. On checking through the details, Lobachevskii, and in more or less the same way Bolyai, showed that the triangles on the horosphere are described by the formulas of hyperbolic trigonometry.

The formulas for spherical geometry depend on the radius of the sphere in question. Similarly, the formulas of hyperbolic trigonometry depend on a certain real parameter. However, this parameter does not have a similarly clear geometrical interpretation. That defect apart, the formulas have a number of reassuring properties. In particular, they closely approximate the familiar formulas of plane geometry when the sides of the triangles are very small, which helps to explain how this geometry could have remained undetected for so long—it differs very little from Euclidean geometry in small regions of space. Formulas for length and area can be developed in the new setting: they show that the area of a triangle is proportional to the amount by which the angle sum of the triangle falls short of two right angles. Lobachevskii, in particular, seems to have felt that the very fact that there were neat and plausible formulas of this kind was enough reason to accept the new geometry. In his opinion, all geometry was about measurement, and theorems in geometry were unfailing connections between measurements expressed by formulas. His methods produced such formulas, and that, for him, was enough.

Bolyai and Lobachevskii, having produced a description of a novel three-dimensional geometry, raised the question of which geometry is true: is it Euclidean geometry or is it the new geometry for some value of the parameter that could presumably be determined experimentally? Bolyai left matters there, but Lobachevskii explicitly showed that measurements of stellar parallax might resolve the question. Here he was unsuccessful: such experiments are notoriously delicate.

By and large, the reaction to Bolyai and Lobachevskii's ideas during their lifetimes was one of neglect and hostility, and they died unaware of the success their discoveries would ultimately have. Bolyai and his father sent their work to Gauss, who replied in 1832 that he could not praise the work "for to do so would be to praise myself," adding, for extra measure, a simpler proof of one of Janos Bolyai's opening results. He was, he said, nonetheless delighted that it was the son of his old friend who had taken precedence over him. Janos Bolyai was enraged, and refused to publish again, thus depriving himself of the opportunity to establish his priority over Gauss by publishing his work as an article in a mathematics journal. Oddly, there is no evidence that Gauss knew the details of the young Hungarian's work in advance. More likely, he saw at once how the theory would go once he appreciated the opening of Bolyai's account.

A charitable interpretation of the surviving evidence would be that, by 1830, Gauss was convinced of the possibility that physical space might be described by non-Euclidean geometry, and he surely knew how to handle two-dimensional non-Euclidean geometry using hyperbolic trigonometry (although no detailed account of this survives from his hand). But the three-dimensional theory was known first to Bolyai and Lobachevskii, and may well not have been known to Gauss until he read their work.

Lobachevskii fared little better than Bolyai. His initial publication of 1829 was savaged in the press by Ostrogradskii, a much more established figure who was, moreover, in St Petersburg, whereas Lobachevskii was in provincial Kazan. His account in *Journal für die reine und angewandte Mathematik* (otherwise known as *Crelle's Journal*) suffered grievously from referring to results proved only in the Russian papers from which it had been adapted. His booklet of 1840 drew only one review, of more than usual stupidity. He did, however, send it to Gauss, who found it excellent and had Lobachevskii elected to the Göttingen Academy of Sciences. But Gauss's enthusiasm stopped there, and Lobachevskii received no further support from him.

Such a dreadful response to a major discovery invites analysis on several levels. It has to be said that the definition of parallels upon which both men depended was, as it stood, inadequate, but their work was not criticized on that account. It was dismissed with scorn, as if it were self-evident that it was wrong: so wrong that it would be a waste of time finding the error it surely contained, so wrong that the right response was to heap ridicule upon its authors or simply to dismiss them without comment. This is a measure of the hold that Euclidean geometry still had on the minds of most people at the time. Even Copernicanism, for example, and the discoveries of Galileo drew a better reception from the experts.

8 Acceptance of Non-Euclidean Geometry

When Gauss died in 1855, an immense amount of unpublished mathematics was found among his papers. Among it was evidence of his support for Bolyai and Lobachevskii, and his correspondence endorsing the possible validity of non-Euclidean geometry. As this was gradually published, the effect was to send people off to look for what Bolyai and Lobachevskii had written and to read it in a more positive light.

Quite by chance, Gauss had also had a student at Göttingen who was capable of moving the matter decisively forward, even though the actual amount of contact between the two was probably quite slight. This was RIEMANN [VI.49]. In 1854 he was called to defend his Habilitation thesis, the postdoctoral qualification that was a German mathematician's license to teach in a university. As was the custom, he offered three titles and Gauss, who was his examiner, chose the one Riemann least expected: "On the hypotheses that lie at the foundation of geometry." The paper, which was to be published only posthumously, in 1867, was nothing less than a complete reformulation of geometry.

Riemann proposed that geometry was the study of what he called MANIFOLDS [I.3 §§6.9, 6.10]. These were "spaces" of points, together with a notion of distance that looked like Euclidean distance on small scales but which could be quite different at larger scales. This kind of geometry could be done in a variety of ways, he suggested, by means of the calculus. It could be carried out for manifolds of any dimension, and in fact Riemann was even prepared to contemplate manifolds for which the dimension was infinite.

A vital aspect of Riemann's geometry, in which he followed the lead of Gauss, was that it was concerned only with those properties of the manifold that were *intrinsic*, rather than properties that depended on some embedding into a larger space. In particular, the distance between two points x and y was defined to be the length of the shortest curve joining x and y that lay entirely within the surface. Such curves are called *geodesics*. (On a sphere, for example, the geodesics are arcs of great circles.)

Even two-dimensional manifolds could have different, intrinsic curvatures—indeed, a single two-dimensional manifold could have different curvatures in different places—so Riemann's definition led to infinitely many genuinely distinct geometries in each dimension. Furthermore, these geometries were best defined without reference to a Euclidean space that contained them, so the hegemony of Euclidean geometry was broken once and for all.

As the word "hypotheses" in the title of his thesis suggests, Riemann was not at all interested in the sorts of assumptions needed by Euclid. Nor was he much interested in the opposition between Euclidean and non-Euclidean geometry. He made a small reference at the start of his paper to the murkiness that lay at the heart of geometry, despite the efforts of Legendre, and toward the end he considered the three different geometries on two-dimensional manifolds for which the curvature is constant. He noted that one was spherical geometry, another was Euclidean geometry, and the third was different again, and that in each case the angle sums of all triangles could be calculated as soon as one knew the sum of the angles of any one triangle. But he made no reference to Bolyai or Lobachevskii, merely noting that if the geometry of space was indeed a three-dimensional geometry of constant curvature, then to determine which geometry it was would involve taking measurements in unfeasibly large regions of space. He did discuss generalizations of Gauss's curvature to spaces of arbitrary dimension, and he showed what METRICS [III.56] (that is, definitions of distance) there could be on spaces of constant curvature. The formula he wrote down is very general, but as with Bolyai and Lobachevskii it depended on a certain real parameter—the curvature. When the curvature is negative, his definition of distance gives a description of non-Euclidean geometry.

Riemann died in 1866, and by the time his thesis was published an Italian mathematician, Eugenio Beltrami, had independently come to some of the same ideas. He was interested in what the possibilities were if one wished to map one surface to another. For example, one might ask, for some particular surface S, whether it is possible to find a map from S to the plane such that the geodesics in S are mapped to straight lines in the plane. He found that the answer was yes if and only if the space has constant curvature. There is, for example, a well-known map from the hemisphere to a plane with this property. Beltrami found a simple way of modifying the formula so that now it defined a map from a surface of constant *negative* curvature onto the interior of a disk, and he realized the significance of what he had done: his map defined a metric on the interior of the disk, and the resulting metric space obeyed the axioms for non-Euclidean geometry; therefore, those axioms would not lead to a contradiction.

Some years earlier, Minding, in Germany, had found a surface, sometimes called the pseudosphere, that had constant negative curvature. It was obtained by rotating a curve called the tractrix about its axis. This surface has the shape of a bugle, so it seemed rather less natural than the space of Euclidean plane geometry and unsuitable as a rival to it. The pseudosphere was independently rediscovered by LIOUVILLE [VI.39] some years later, and Codazzi learned of it from that source and showed that triangles on this surface are described by the formulas of hyperbolic trigonometry. But none of these men saw the connection to non-Euclidean geometry—that was left to Beltrami.

Beltrami realized that his disk depicted an infinite space of constant negative curvature, in which the geometry of Lobachevskii (he did not know at that time of Bolyai's work) held true. He saw that it related to the pseudosphere in a way similar to the way that a plane relates to an infinite cylinder. After a period of some doubt, he learned of Riemann's ideas and realized that his disk was in fact as good a depiction of the space of non-Euclidean geometry as any could be; there was no need to realize his geometry as that of a surface in Euclidean three-dimensional space. He thereupon published his essay, in 1868. This was the first time that sound foundations had been publicly given for the area of mathematics that could now be called non-Euclidean geometry.

In 1871 the young KLEIN [VI.57] took up the subject. He already knew that the English mathematician CAYLEY [VI.46] had contrived a way of introducing Euclidean metrical concepts into PROJECTIVE GEOMETRY [I.3 §6.7]. While studying at Berlin, Klein saw a way of generalizing Cayley's idea and exhibiting Beltrami's non-Euclidean geometry as a special case of projective geometry. His idea met with the disapproval of WEIERSTRASS [VI.44], the leading mathematician in Berlin, who objected that projective geometry was not a metrical geometry: therefore, he claimed, it could not generate metrical concepts. However, Klein persisted and in a series of three papers, in 1871, 1872, and 1873, showed that all the known geometries could be regarded as subgeometries of projective geometry. His idea was to recast geometry as the study of a group acting on a space. Properties of figures (subsets of the space) that remain invariant under the action of the group are the geometric properties. So, for example, in a projective space of some dimension, the appropriate group for projective geometry is the group of all transformations that map lines to lines, and the subgroup that maps the interior of a given conic to itself may be regarded as the group of transformations of non-Euclidean geometry: see the box on p. 94. (For a fuller discussion of Klein's approach to geometry, see [I.3 §6].)

In the 1870s Klein's message was spread by the first and third of these papers, which were published in the recently founded journal *Mathematische Annalen*. As Klein's prestige grew, matters changed, and by the 1890s, when he had the second of the papers republished and translated into several languages, it was this, the *Erlanger Programm*, that became well-known. It is named after the university where Klein became a professor, at the remarkably young age of twenty-three, but it was not his inaugural address. (That was about mathematics education.) For many years it was a singularly obscure publication, and it is unlikely that it had the effect on mathematics that some historians have come to suggest.

9 Convincing Others

Klein's work directed attention away from the *figures* in geometry and toward the *transformations* that do not alter the figures in crucial respects. For example, in Euclidean geometry the important transformations are the familiar rotations and translations (and reflections, if one chooses to allow them). These correspond to the motions of rigid bodies that contemporary psychologists saw as part of the way in which individuals learn the geometry of the space around them. But this theory was philosophically contentious, especially when it could be extended to another metrical geometry, non-Euclidean geometry. Klein prudently entitled his main papers "On the so-called non-Euclidean geometry," to keep hostile philosophers at bay (in particular Lotze, who was the well-established Kantian philosopher at Göttingen). But with these papers and the previous work of Beltrami the case for non-Euclidean geometry was made, and almost all mathematicians were persuaded. They believed, that is, that alongside Euclidean geometry there now stood an equally valid mathematical system called non-Euclidean geometry. As for which one of these was true of space, it seemed so clear that Euclidean geometry was the sensible choice that there appears to have been little or no discussion. Lipschitz showed that it was possible to do all of mechanics in the new setting, and there the matter rested, a hypothetical case of some charm but no more. Helmholtz, the leading physicist of his day, became interested—he had known Riemann personally—and gave an account of what space would have to be if it was learned about through the free mobility of bodies. His first account was deeply flawed, because he was unaware of non-Euclidean geometry, but when Beltrami pointed this out to him he reworked it (in 1870). The reworked version also suffered from mathematical deficiencies, which were pointed out somewhat later by LIE [VI.53], but he had more immediate trouble from philosophers.

Their question was, "What sort of knowledge is this theory of non-Euclidean geometry?" Kantian philosophy was coming back into fashion, and in Kant's view knowledge of space was a fundamental pure a priori intuition, rather than a matter to be determined by experiment: without this intuition it would be impossible to have any knowledge of space at all. Faced with a rival theory, non-Euclidean geometry, neo-Kantian philosophers had a problem. They could agree that the mathematicians had produced a new and prolonged logical exercise, but could it be knowledge of the world? Surely the world could not have two kinds of geometry? Helmholtz hit back, arguing that knowledge of Euclidean geometry and non-Euclidean geometry would be acquired in the same way—through experience—but these empiricist overtones were unacceptable to the philosophers, and non-Euclidean geometry remained a problem for them until the early years of the twentieth century.

Mathematicians could not in fact have given a completely rigorous defense of what was becoming the accepted position, but as the news spread that there were two possible descriptions of space, and that one could therefore no longer be certain that Euclidean geometry was correct, the educated public took up the question: what was the geometry of space? Among the first to grasp the problem in this new formulation was POINCARÉ [VI.61]. He came to mathematical fame in the early 1880s with a remarkable series of essays in which he reformulated Beltrami's disk model so as to make it *conformal*: that is, so that angles in non-Euclidean geometry were represented by the same angles in the

Cross-ratios and distances in conics. A projective transformation of the plane sends four distinct points on a line, A, B, C, D, to four distinct collinear points, A′, B′, C′, D′, in such a way that the quantity

$$\frac{AB}{AD}\frac{CD}{CB}$$

is preserved: that is,

$$\frac{AB}{AD}\frac{CD}{CB} = \frac{A'B'}{A'D'}\frac{C'D'}{C'B'}.$$

This quantity is called the *cross-ratio* of the four points A, B, C, D, and is written CR(A, B, C, D).

In 1871, Klein described non-Euclidean geometry as the geometry of points inside a fixed conic, K, where the transformations allowed are the projective transformations that map K to itself and its interior to its interior (see figure 7). To define the distance between two points P and Q inside K, Klein noted that if the line PQ is extended to meet K at A and D, then the cross-ratio CR(A, P, D, Q) does not change if one applies a projective transformation: that is, it is a *projective invariant*. Moreover, if R is a third point on the line PQ and the points lie in the order P, Q, R, then $\mathrm{CR}(A,P,D,Q)\,\mathrm{CR}(A,Q,D,R) = \mathrm{CR}(A,P,D,R)$. Accordingly, he defined the distance between P and Q as $d(PQ) = -\frac{1}{2}\log \mathrm{CR}(A,P,D,Q)$ (the factor of $-\frac{1}{2}$ is introduced to facilitate the later introduction of trigonometry). With this definition, distance is additive along a line: $d(PQ) + d(QR) = d(PR)$.

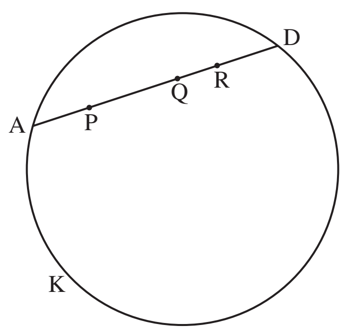

Figure 7 Three points, P, Q, and R, on a non-Euclidean straight line in Klein's projective model of non-Euclidean geometry.

model. He then used his new disk model to connect complex function theory, the theory of linear differential equations, RIEMANN SURFACE [III.79] theory, and non-Euclidean geometry to produce a rich new body of ideas. Then, in 1891, he pointed out that the disk model permitted one to show that any contradiction in non-Euclidean geometry would yield a contradiction in Euclidean geometry as well, and vice versa. Therefore, Euclidean geometry was consistent if and only if non-Euclidean geometry was consistent. A curious consequence of this was that if anybody *had* managed to derive the parallel postulate from the core of Euclidean geometry, then they would have inadvertently proved that Euclidean geometry was inconsistent!

One obvious way to try to decide which geometry described the actual universe was to appeal to physics. But Poincaré was not convinced by this. He argued in another paper (1902) that experience was open to many interpretations and there was no logical way of deciding what belonged to mathematics and what to physics. Imagine, for example, an elaborate set of measurements of angle sums of figures, perhaps on an astronomical scale. Something would have to be taken to be straight, perhaps the paths of rays of light. Suppose, finally, that the conclusion is that the angle sum of a triangle is indeed less than two right angles by an amount proportional to the area of the triangle. Poincaré said that there were two possible conclusions: light rays are straight and the geometry of space is non-Euclidean; or light rays are somehow curved, and space is Euclidean. Moreover, he continued, there was no logical way to choose between these possibilities. All one could do was to make a convention and abide by it, and the sensible convention was to choose the simpler geometry: Euclidean geometry.

This philosophical position was to have a long life in the twentieth century under the name of *conventionalism*, but it was far from accepted in Poincaré's lifetime. A prominent critic of conventionalism was the Italian Federigo Enriques, who, like Poincaré, was both a powerful mathematician and a writer of popular essays on issues in science and philosophy. He argued that one could decide whether a property was geometrical or physical by seeing whether we had any control over it. We cannot vary the law of gravity, but we can change the force of gravity at a point by moving matter around. Poincaré had compared his disk model to a metal disk that was hot in the center and got cooler as one moved outwards. He had shown that a simple law of cooling produced figures identical to those of non-Euclidean geometry. Enriques replied that heat was likewise something we can vary. A property such as Poincaré invoked, which was truly beyond our control, was not physical but geometric.

10 Looking Ahead

In the end, the question was resolved, but not in its own terms. Two developments moved mathematicians beyond the simple dichotomy posed by Poincaré. Starting in 1899, HILBERT [VI.63] began an extensive rewriting of geometry along axiomatic lines, which eclipsed earlier ideas of some Italian mathematicians and opened the way to axiomatic studies of many kinds. Hilbert's work captured very well the idea that if mathematics is sound, it is sound because of the nature of its reasoning, and led to profound investigations in mathematical logic. And in 1915 Einstein proposed his general theory of relativity, which is in large part a geometric theory of gravity. Confidence in mathematics was restored; our sense of geometry was much enlarged, and our insights into the relationships between geometry and space became considerably more sophisticated. Einstein made full use of contemporary ideas about geometry, and his achievement would have been unthinkable without Riemann's work. He described gravity as a kind of curvature in the four-dimensional manifold of spacetime (see GENERAL RELATIVITY AND THE EINSTEIN EQUATIONS [IV.13]). His work led to new ways of thinking about the large-scale structure of the universe and its ultimate fate, and to questions that remain unanswered to this day.

Further Reading

Bonola, R. 1955. *History of Non-Euclidean Geometry*, translated by H. S. Carslaw and with a preface by F. Enriques. New York: Dover.

Euclid. 1956. *The Thirteen Books of Euclid's Elements*, 2nd edn. New York: Dover.

Gray, J. J. 1989. *Ideas of Space: Euclidean, Non-Euclidean, and Relativistic*, 2nd edn. Oxford: Oxford University Press.

Gray, J. J. 2004. *Janos Bolyai, non-Euclidean Geometry and the Nature of Space*. Cambridge, MA: Burndy Library.

Hilbert, D. 1899. *Grundlagen der Geometrie* (many subsequent editions). Tenth edn., 1971, translated by L. Unger, *Foundations of Geometry*. Chicago, IL: Open Court.

Poincaré, H. 1891. Les géométries non-Euclidiennes. *Revue Générales des Sciences Pures et Appliquées* 2:769–74. (Reprinted, 1952, in *Science and Hypothesis*, pp. 35–50. New York: Dover.)

——. 1902. L'expérience et la géométrie. In *La Science et l'Hypothèse*, pp. 95–110. (Reprinted, 1952, in *Science and Hypothesis*, pp. 72–88. New York: Dover.)

II.3 The Development of Abstract Algebra

Karen Hunger Parshall

1 Introduction

What is algebra? To the high school student encountering it for the first time, algebra is an unfamiliar abstract language of x's and y's, a's and b's, together with rules for manipulating them. These letters, some of them variables and some constants, can be used for many purposes. For example, one can use them to express straight lines as equations of the form $y = ax + b$, which can be graphed and thereby visualized in the Cartesian plane. Furthermore, by manipulating and interpreting these equations, it is possible to determine such things as what a given line's root is (if it has one)—that is, where it crosses the x-axis—and what its slope is—that is, how steep or flat it appears in the plane relative to the axis system. There are also techniques for solving simultaneous equations, or equivalently for determining when and where two lines intersect (or demonstrating that they are parallel).

Just when there already seem to be a lot of techniques and abstract manipulations involved in dealing with lines, the ante is upped. More complicated curves like quadratics, $y = ax^2 + bx + c$, and even cubics, $y = ax^3 + bx^2 + cx + d$, and quartics, $y = ax^4 + bx^3 + cx^2 + dx + e$, enter the picture, but the same sort of notation and rules apply, and similar sorts of questions are asked. Where are the roots of a given curve? Given two curves, where do they intersect?

Suppose now that the same high school student, having mastered this sort of algebra, goes on to university and attends an algebra course there. Essentially gone are the by now familiar x's, y's, a's, and b's; essentially gone are the nice graphs that provide a way to picture what is going on. The university course reflects some brave new world in which the algebra has somehow become "modern." This *modern* algebra involves abstract structures—GROUPS [I.3 §2.1], RINGS [III.81 §1], FIELDS [I.3 §2.2], and other so-called objects—each one defined in terms of a relatively small number of axioms and built up of substructures like subgroups, ideals, and subfields. There is a lot of moving around between these objects, too, via maps like group homomorphisms and ring AUTOMORPHISMS [I.3 §4.1]. One objective of this new type of algebra is to understand the underlying structure of the objects and, in doing so, to