1

Beauty and the Beast

in which departures from traditional thinking help expose some unsuspected beauty

Beauty vs. Reality

Teacher. I hear you've been studying conic sections. Conics go back more than 2300 years, and have found their way into almost every area of mathematics and physics. What a beautiful subject!

Philosopher. Beauty was my vision. Then came the reality.

Teacher. What do you mean?

Philosopher. My journey into conics started very nicely. But then I got to thinking about how the whole subject ought to be perfectly unified. It was then that I began running into a number of simple-sounding questions. They all ought to have simple and obvious answers—I just can't answer them!

Teacher. What got you to thinking that the whole subject should be perfectly unified?

Philosopher. It happened when I saw that among nondegenerate conics, there's "conservation of information," implying that they're all essentially equivalent.

Teacher. I don't understand.

Philosopher. Well, every conic is the intersection of some double cone and a plane. If the plane goes through the cone's vertex, then the conic is *degenerate*—like a line, two lines or a point, for example. It's when the plane misses the vertex that you get nondegenerate conics—true ellipses, parabolas or hyperbolas.

Teacher. I agree with that.

> Euclid defined conics using only a single cone. It was Apollonius (ca. 262–ca. 190 BCE) who introduced the double cone, giving the hyperbola its two branches.

Philosopher. If you keep track of where the vertex was, then for nondegenerate conics you can go back and reconstruct the original cone—just pass lines through the vertex and the points of the conic. You might need to take the topological closure of this, but the essential thing is that you can always retrieve the cone. (Topological closure is defined in Appendix 2.)

Cone comes from the Greek *konos*, meaning "pine cone"; typically, an end of a pine cone has a conical shape. *Konos* is thought to be also related to our word "hone." A pencil that is honed to a point has a conical shape—think of a pencil sharpener.

Apollonius
ca. 262–ca. 190 BCE

Apollonius of Perga wrote an extraordinary treatise, *Conic Sections*. This eight-book masterpiece completely superseded works on conics by earlier geometers such as Euclid, Aristaeus and Menaechmus. It was Apollonius who introduced today's method of defining a conic by slicing a double cone with a plane. Before him, geometers used three separate "flavors" of single-napped cone, having vertex angle less than, equal to or greater than a right angle; the slicing plane was then always perpendicular to a line in the cone. The respective section was then an ellipse, parabola or one branch of a hyperbola. Apollonius realized that a hyperbola needed to have two branches, and his cone sections naturally produce them. In many ways Apollonius' work represented, in its originality, depth and completeness, the pinnacle of ancient mathematics. It earned him the nickname "The Great Geometer."

Teacher. This doesn't work for degenerate ones?

Philosopher. No. For example, there are lots of different cones intersecting a plane in a given point. Likewise, many cones can intersect a plane in one line. Same thing for two crossing lines. But nondegenerate conics? They're all blood brothers of each other—Figure 1.1 suggests the idea:

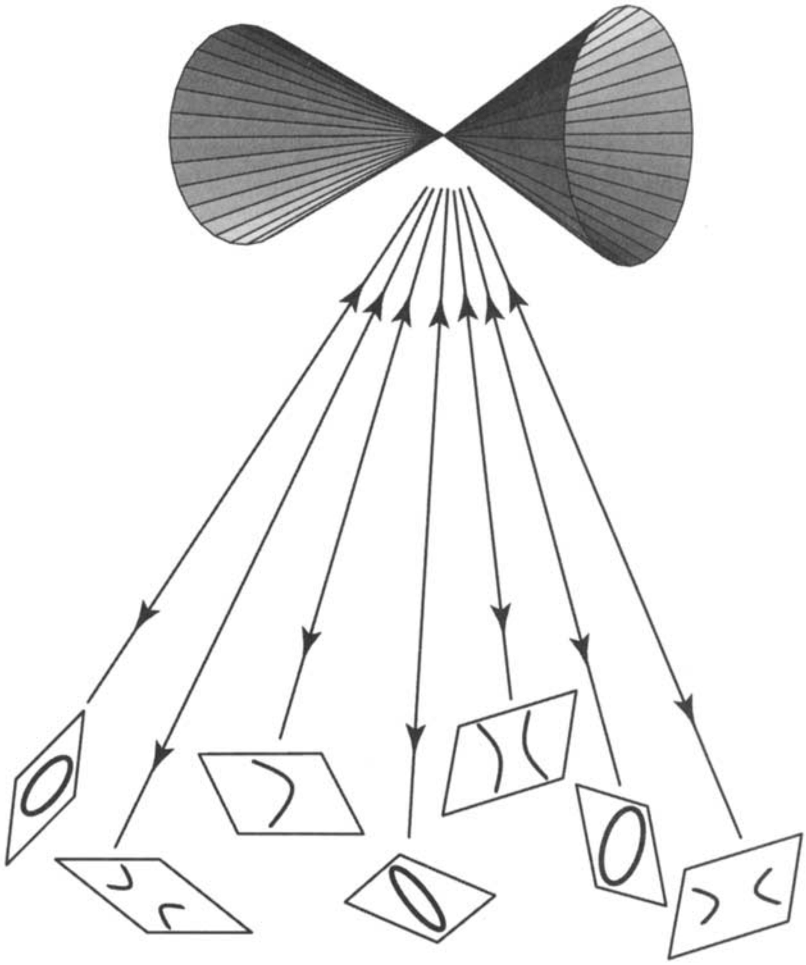

FIGURE 1.1. At the top is the parent cone and at the bottom, a sampling of nondegenerate conics. Any down arrow represents taking a nondegenerate slice of the cone. The up arrow from any one of the slices represents reconstructing the parent cone. All these nondegenerate conics are equivalent in the sense that any two of them are connected by a two-way path passing through the parent cone.

At the top we see the parent cone and below that, a sampling of ellipses, parabolas and hyperbolas that are plane sections of the cone. One could call them "children." Any downward arrow represents taking a section; any upward arrow represents reconstructing the cone from that section. The plane cutting the cone becomes a "screen," giving a particular view of the conic. Starting at the bottom of the figure, we can go up to the parent cone and then back down to another conic, and in this sense, the two conics are *equivalent*. From the standpoint of screens, they're simply different views of the same thing.

Teacher. This is a nice observation, but how does this lead to questions you can't answer?

Philosopher. Well, if all those nondegenerate conics are equivalent, then whatever you say about one should make sense with any other. So, for example, any ellipse encloses an area. But hyperbolas don't! One could say "Well, they just don't—that's obvious!" But I can't believe that this represents the ideal state of any theory. There's a big information gap.

Teacher. Your point of view could generate a lot of questions.

Philosopher. Tell me about it! As just one example, hyperbolas have asymptotes, so *why don't ellipses*? Again, the obvious answer: "That's a nonsensical question. Ellipses are bounded—there can't be asymptotes!" But if ellipses and hyperbolas are in fact different views of the same thing, and no information is lost, then where, exactly, do the asymptotes disappear to? Or how about this: when you start sketching an ellipse from its standard-form equation, it is natural to draw the "fundamental rectangle" that encloses it. I don't see a rectangle of *any* sort enclosing a hyperbola!

Teacher. I don't either, but maybe your questions aren't posed right.

Philosopher. I've wondered about that. But there are other, more specific and clear-cut questions, too. One of them, for example, comes from an ancient method of generating an ellipse—the "stakes and string" method: drive two stakes into the ground, and to each stake, tie one end of a piece of string. Now, keeping the string taut, trace out an ellipse. The stakes are the ellipse's foci, and the taut string segments can be thought of as the path of a photon as it goes from one focus to a point on the ellipse, reflects off the ellipse and proceeds to the other focus.

Student. This has a practical application: imagine an enormous ellipse, with the sun at one focus and the earth at the other. Here on earth, just a tiny part of this ellipse, maybe a foot or two wide, can intercept and reflect to the focal point enough radiant energy from the sun to start a fire!

> It was Johannes Kepler (1571–1630) who, in 1604, added *focus* to the vocabulary of conics. The term comes from the Latin for *hearth* or *fireplace*, which represents a room's focal point or center of interest.
> It does not refer to heat or burning.

Philosopher. But again, I run into a problem if I ask, "Why doesn't this work for a hyperbola?" If I release a photon from F_1, it bounces off the mirrored hyperbola and heads for "outer space." Forget about starting any fire!

> Conic sections were largely forgotten in the centuries following the ancient Greeks. It was only in Kepler's decades-long struggle to determine planetary paths that he turned—mostly in desperation—to those ancient curiosities. To him, the ellipse was a definite step down from his beloved "perfect" circles.

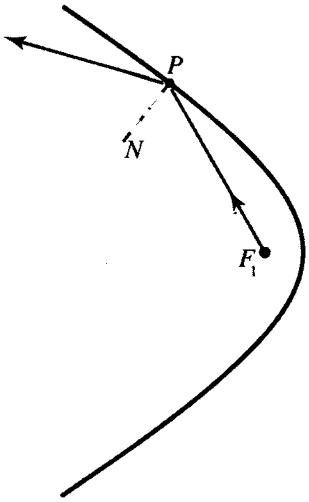

FIGURE 1.2. A photon released from focus F_1 reflects off a hyperbola the "wrong way"—it heads for the void rather than toward the other focus.

Student. But there is a standard picture for hyperbolas:

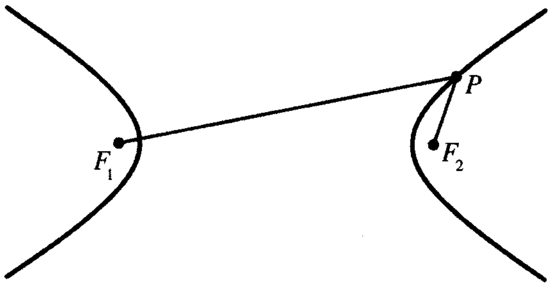

FIGURE 1.3. One can draw a "stakes and rubber band" model of a hyperbola. In it, $F_1P - PF_2 = K$, where the constant K is the shortest distance between the branches.

The segments seem to represent light paths for hyperbolas, too.

Philosopher. But you've got the photon piercing the left-branch mirror so it can get to P!

Student. We could throw that branch away, perhaps ...

Philosopher. That doesn't seem fair, but even if you did, then when the photon reaches P and reflects, it again heads off into outer space. To make things work, you've got to force the photon to go *backwards* to reach the other focus.

That is, to make the standard picture work, you have to break two physical laws: pierce one mirror, and make the photon move backwards as soon as it reflects off the other. That's crazy! So often, when the ellipse performs like a beauty, the hyperbola acts like a beast.

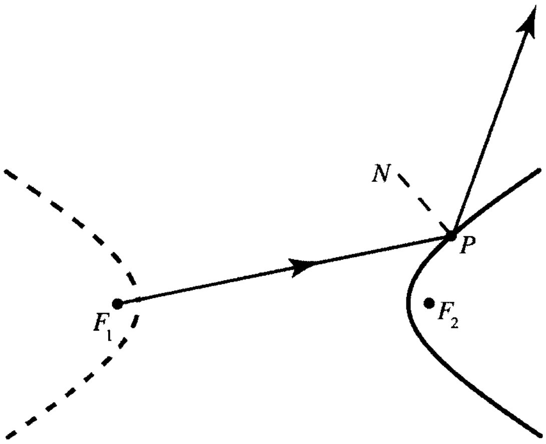

FIGURE 1.4. Even if the photon goes through the left branch, it still reflects off the right branch the wrong way. It would have to travel backwards to reach F_2.

Teacher. I admit there seems to be a problem here. Your contention is that because conics are in their innermost nature unified, what works for the goose, should work for the gander?

Philosopher. Yes. To me, the standard picture is wrong.

Teacher. But it's so well established!

Philosopher. It sure is. But I'm thinking that might be the fault of man, not of Mathematics itself. When I say this, I'm viewing things through "Platonic eyeglasses": Mathematics with a capital "M" is what's objectively out there, independent of anything we human beings might know of it. In contrast, mathematics with a lower-case "m" is what we happen to see or understand of it. We devise definitions, conventions, modes of representation and so on, in our struggle to understand Mathematics. The problem is that sometimes our constructs get set in concrete long before we arrive at a truer, more complete picture, and then those conventions work against us.

Teacher. I see your point. I guess that in some cases, those conventions could have worked against us, actually impeding the discovery process! When I was very young, I had an experience that sounds a lot like this.

Student. Do I hear a story coming? I love stories!

Teacher. It happened when I started learning about decimals in school. I knew then that ten has one zero, a hundred has two, a thousand, three, and so on. And then this teacher starts saying that a *tenth* doesn't have *any* zero, a *hundredth* has only *one*, a thousandth only two, and so on. I got very upset, and I told her she had to be wrong. She assured me that this is really the way things are. I felt betrayed by math. I loved its consistency, and the way you could figure things out, and this seemed like a darker side or something. Only much later did I have enough perspective to put my finger on the problem: **The decimal point is always misplaced!**

Philosopher. That's a pretty brave statement.

Teacher. Yes, but it means that Mathematics is still wonderful; mathematics may have betrayed me, but Mathematics didn't.

Student. You have to tell me how the decimal is misplaced!

Teacher. Well, the only essential role the decimal point plays is to identify the "units" position. For historical reasons—writing sequentially, printing or typesetting in a line—the decimal point got put to the right of the unit place. But it really shouldn't go there, any more than it should go to the left. One symmetric, fair place would be *over* the number. Then $1\dot{0}$ is ten and $\dot{0}1$ is a tenth. In fact it could go under, and it doesn't even have to be a point. It could be a bar, or the number in the units position could be written in bold, or color, or even with a circle around it—anything to make it special, but in a symmetric way:

$$1\,0\,0\,\overset{*}{0} \qquad \overset{*}{0}\,0\,0\,1$$

$$1\,0\,0\,\underline{0} \qquad \underline{0}\,0\,0\,1$$

$$1\,0\,0\,\boxed{0} \qquad \boxed{0}\,0\,0\,1$$

FIGURE 1.5. Any of these would be more symmetric than decimal notation for one thousand and one thousandth.

I get the feeling that, analogously, some of your conundrums may be coming from various long-established conventions and assumptions, just like the decimal point.

Philosopher. Wouldn't that be incredible? I'd be very obliged to begin exploring this approach with you. Maybe I can get unstuck! If we succeed, we'll have a more beautiful theory of conics.

> Compared to the decimal point, base ten is older and even more arbitrary. Arising from the peculiarity of ten digits (fingers), our base was set in stone long before Gottfried Leibnitz (1646–1716) discovered base two in 1679. Base two is less efficient than base ten but is mathematically more natural.

Viewing Screens

Teacher. Last time, we referred to the cutting plane as a "screen," and that got me thinking. It seems that the vertex of the double cone can be regarded as a light source, each line consisting of two oppositely-directed light rays. The plane is like the camera's film, and the conic is the resulting photograph.

Student. Sort of like shining a flashlight on a flat ceiling? You can get a circle, ellipse, or parabola.

Teacher. Yes, and you can get a hyperbola, too. Use a "double flashlight" like this:

FIGURE 1.6. Two conical beams of light emanating from a "double flashlight."

Hold it so both beams hit the ceiling, and you'll get the two branches of a hyperbola.

There's a nice continuity too: if the cone of light is circular and the screen is perpendicular to the cone's axis of symmetry, the image is circular. But when you turn the flashlight a little bit away from perpendicularity, you introduce some distortion, and the image is elliptical. As you continue this you get more and more distortion, and turned far enough, a light ray finally becomes parallel to the ceiling and what you see is a parabola.

Philosopher. So if the ceiling were a large piece of film, a parabola can be regarded as a grossly-distorted picture of a circle?

Teacher. Exactly. And by turning the camera even further to get a hyperbola, the information has been so tremendously distorted that one has to think a bit to realize that it's just a different view of a circle.

Philosopher. So undistorted pictures of the circle are actually quite rare—you get them only when the film happens to be perpendicular to a circular cone's axis of symmetry.

Student. If being perpendicular is so good, why not just curl the film somehow, so the light rays always hit it at right angles?

Teacher. A camera with bent film?!

Student. Why not?

Philosopher. Maybe you have an idea here! I once visited a cinema where the room was circular, and there were projectors aimed at successive sections of the room. It gave a 360° view of everything. It was pretty impressive.

Teacher. You're saying the walls played the role of film bent into a cylinder? Interesting. But the light rays from the projector couldn't have hit every point of the screen orthogonally.

Philosopher. Come to think of it, we'd have perpendicularity on at most one circle of the cylindrical wall, wouldn't we?

Teacher. Yes. You know, in less than a minute I find myself changing from disbelief to now wondering about even *more* bending! A cylinder curves in only one direction, so to speak, and we need a room curving uniformly in *all* directions.

> The Euclidean plane often keeps us from seeing larger truths about conics.

Student. I once visited a planetarium. The whole place was the top half of a sphere.

Philosopher. Aha! Maybe we could let the screen—the film—be spherical. If the vertex of the circular cone—our light source—is at the center of the sphere, then no matter how we angle that source of light, we always record without distortion. If the cone is circular, we get two opposite circles. If the cone is elliptical, defined, say, by

$$z^2 = \frac{x^2}{2^2} + \frac{y^2}{1^2},$$

then we get two elongated curves on opposite sides of the sphere.

Teacher. I wonder if there exists a truly *ideal* screen... If there is such a screen, it should have no distortion, and record exactly one point for each double ray through the vertex. The plane not only creates distortion, it also fails to record light rays parallel to it. The sphere has no distortion, but as far as intersections are concerned, it's almost too much of a good thing—it records *two* points for every double ray!

Philosopher. Yes, I agree.

Teacher. However, I have an idea. Topologists have long used the trick of "identifying points"—for example, one can identify points to represent topological two-dimensional spaces such as a cylinder, Möbius strip, torus, sphere, the projective plane or a Klein bottle using a flat piece of paper, as in Figure 1.7.

I'd like to make a proposal:

Let our viewing screen be a sphere with opposite points identified.

With this definition of viewing screen, there's no distortion and every double ray intersects the screen in exactly one point.

Philosopher. This sounds like a promising idea.

Teacher. In a sphere with opposite points identified, any double cone intersects the sphere in two opposite curves, each pair of diametrically opposite (that is,"antipodal") points being identified.

Student. But we don't go around carrying globes with curves drawn on them! We use ordinary paper, like anybody else.

Teacher. I have to agree—all this is three-dimensional, and we need to describe exactly how to draw sphere-plus-curve on a piece of paper. We need an "official agreement" of some sort. Let me propose this:

- Use a unit sphere centered at $(0, 0, 0)$ in (x, y, z)-space;
- Place the cone's vertex at $(0, 0, 0)$;
- Vertically project the top half ($z \geq 0$) of the sphere to the (x, y)-plane. Of course, this projects any curve in the top hemisphere to the (x, y)-plane, as well.

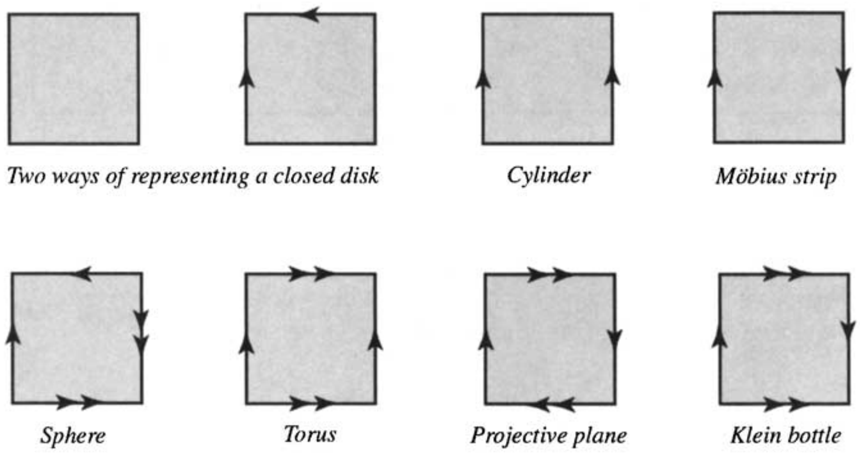

FIGURE 1.7. Various ways of identifying the sides of a square create an assortment of familiar topological two-dimensional spaces. Two sides are *identified* if they're marked with arrows of the same kind (both single-headed or both double-headed)—that is, we consider that the sides are glued together so that the arrow directions agree. Representing such two-dimensional topological spaces using identifications is useful because physically making such surfaces can be challenging or impossible in ordinary 3-D space. For example, at the easy end, to make a Möbius strip one need only "stretch the square," replacing it by a rectangle long enough so the paper won't tear when twisted. But neither the projective plane nor the Klein bottle can be realized in 3-D without introducing self-intersections. Notice that those surfaces in the top row all have some unmarked sides; after the gluing process, these unmarked sides remain, forming a boundary. On the other hand, all four sides of each figure in the bottom row have arrows, so each of these figures is without boundary. The Möbius strip, projective plane and Klein bottle are all nonorientable. For topological two-manifolds, this turns out to be equivalent to "one-sidedness," familiar in the Möbius strip. Each figure in the bottom row has a genus: from left to right it is 0, 1, 1, 2.

The image of the hemisphere in the (x, y)-plane is the "universe" in which we represent all conics, and any conic in that universe is the vertical projection of a curve on the sphere.

Philosopher. Your projection includes the hemisphere's equator, and its opposite points are identified. Does that mean the "universe" is actually a closed disk with opposite boundary points identified?

Teacher. Precisely! A closed disk with opposite boundary points identified is topologically identical to a sphere in which each pair of diametrically-opposite points gets identified. As yet another way of looking at it, one could add a point at infinity to the "end" of each 1-subspace of the (x, y)- plane and set up a topological equivalence between this extended plane and our disk.

Philosopher. Well, I'm looking at that assortment of two-dimensional topological spaces in Figure 1.7, and it appears that what you're describing is topologically the same as the pro-