Making Waves

A Guide to the Ideas behind *Outside In*

Silvio Levy

With an article by Bill Thurston
Afterword by Albert Marden

The Geometry Center • University of Minnesota

About This Work (Please Read)

Making Waves, as the subtitle implies, is an informal presentation of the mathematical ideas behind the computer-animated movie *Outside In*. Most of the work consists of two parallel strands: the dialog of the movie and an expository text that, while relevant to the corresponding chunk of dialog, can also be read independently. A table of contents can be found on the inside of the back cover.

I had a great deal to say and not much space to say it in, so I have tried to make every word count. As a result, the pace is often swift. **Don't get discouraged if you don't understand a passage right away:** a later rereading will probably make things clearer.

This isn't a textbook: it does not assume a particular level of mathematical sophistication. It doesn't matter if you've never taken calculus or heard of topology before—you'll get something out of it. Just skip over the portions marked ▸◂ , which assume some background knowledge (typically that of a math major or graduate student) and are not required for the understanding of the rest of the text.

Take time to digest the "food for thought" paragraphs, marked . You'll usually be able to figure out the grapes if you've been following the text. The bananas are slippery, and those marked ▸◂ depend on background knowledge. The walnuts are the toughest, but should mostly be within the reach of math graduate students or even advanced math majors. **Warning: A question doesn't necessarily have a single right answer!**

If you feel that something is poorly explained, or if you have questions that are not answered here, feel free to contact me (see page 2). I will reply and perhaps include new material, based on your questions, in the next printing. Corrections are of course welcome.

Silvio Levy
Minneapolis and Berkeley
June 1994 – March 1995

Making Waves

A Guide to the Ideas behind ***Outside In***

Silvio Levy

The Geometry Center
University of Minnesota

With an article by Bill Thurston
Afterword by Albert Marden

A K Peters
Wellesley, Massachusetts

Editorial, sales, and customer service office:

A K Peters, Ltd., 289 Linden Street, Wellesley, MA 02181
phone: 617-235-2210 fax: 617-235-2404

For questions and comments contact the author at The Geometry Center, University of Minnesota, 1300 South Second St., suite 500, Minneapolis, MN 55454, or by e-mail at levy@geom.umn.edu.

"The Geometry Center" is the informal name for the Science and Technology Center for the Computation and Visualization of Geometric Structures, established by the National Science Foundation at the University of Minnesota. The Geometry Center is supported by grant DMS-8920161 from the National Science Foundation, by grant DOE/DE-FG02-92ER25137 from the Department of Energy, by the University of Minnesota, and by Minnesota Technology, Inc.

The printing of *Making Waves* was funded in part by Wolfram Research, Inc. Wolfram Research maintains one of the largest privately funded computer mathematics research organization in the country, and has approximately 200 employees worldwide.

Figures on page 32 courtesy of Tony Phillips and Nelson Max. Used with permission.
Figure on pages 24–25 based on a design by Stuart Levy and Delle Maxwell.
Cover figure and figures next to *Outside In* dialog taken from *Outside In*.
All other figures made by the author (see page 26).
Cover design by the author, Delle Maxwell and Tamara Munzner.

Credits for *Outside In* (video)

Dedicated to Albert Marden
Concept: Bill Thurston
Direction: Silvio Levy, Delle Maxwell, Tamara Munzner
Master Illusionist: Nathaniel Thurston
Technical Shepherd: Stuart Levy
Animation: David Ben-Zvi, Silvio Levy, Delle Maxwell, Daeron Meyer, Tamara Munzner
Additional Animation: Adam Deaton, Dan Krech, Matt Headrick, Mark Phillips
Technical Contributions: Celeste Fowler, Charlie Gunn, Stephanie Mason, Linus Upson, Scott Wisdom
Written by: David Ben-Zvi, Matt Headrick, Silvio Levy, Delle Maxwell, Tamara Munzner, Bill Thurston
Software: RenderMan, Softimage, Mathematica, Geomview, programs in C, C++ and Perl
Still images: George Francis, *A Topological Picturebook*, Springer, 1987
Audio post-production, sound design and mix: Sam Hudson, Bryan Forrester (Hudson–Forrester Studios)
Narration: Karen McNenny, Paul de Cordova
Video post-production: Lamb & Co.
Editing: Audrey Robinson
Technical assistance: Scott Gaff
Special thanks: François Apéry, George Francis, Nelson Max, Richard McGehee, Tony Phillips, Angie Vail

ISBN 1-56881-046-6 (video and paperback)
ISBN 1-56881-049-0 (paperback)

Printed in the United States of America
99 98 97 96 95 10 9 8 7 6 5 4 3 2 1

Y: *Hey, I read somewhere that mathematicians can turn a sphere inside out.*

X: Yes, that's true.

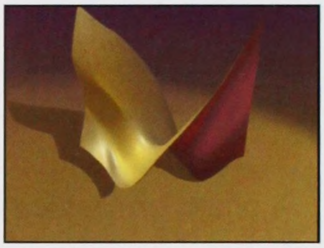
Y: *What's the big deal? Just poke a hole in it and pull it through.*

X: Sure, but the point is to do it without making a hole.

Y: *But then it seems impossible!*

X: You're right, you cannot do it with an ordinary sphere like a basketball. You have to understand the rules of the game: this sphere is made of an abstract elastic material that can stretch and bend and pass through itself. But you cannot rip or puncture this material without destroying it, and you cannot crease it or bend it sharply.

Cast of Characters: Xanthippe and Yorick (usually abbreviated by their initials).



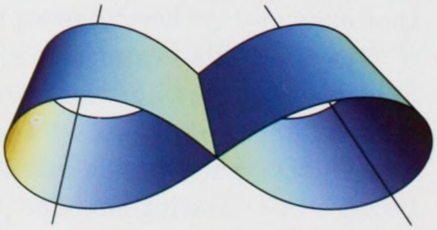
Where do the rules of the game come from?

Just as the rules of baseball have changed over time for balance, and just as you can't write a courtroom thriller based on a case where all the evidence is on one side and the conclusion is foregone, so it is in mathematics: the most interesting situations are those where it's not obvious whether or not something can be done—or where it turns out that a solution exists where none was expected.

If everything is allowed—if you can make holes, for instance—it's obvious that you can "turn the sphere inside out", and *Outside In* would be a very short movie. The question becomes more interesting when we limit the allowed transformations—but not too much. That's where the rules come in.

It's best if the rules are "natural", not far-fetched or contrived. To say that the surface shouldn't be torn or punctured or creased is natural in the context of *differential topology* (see next page)—a tear or a crease changes the character of the surface more drastically than a mere deformation from roundness.

On the other hand, self-intersecting surfaces may not seem natural at first glance. However, by "natural" I don't mean occurring in nature, but rather reasonable from the point of view of analogy or logic. Self-intersecting surfaces are natural by analogy with curves (see page 5), and also because one way that surfaces are defined in mathematics is as the set of points in space satisfying a certain equality—for instance, the sphere is the set of points whose distance to the center is equal to a fixed number. Surfaces formulated in this way often turn out to be self-intersecting, like the one on the right, which has the property that for every point on the surface the product of the distances to the two parallel lines shown is the same.



Y: *If the surface can pass through itself, what's the problem?*

X: Do you think allowing self-intersections makes it easy? Try it.

Y: *I'll push the two halves right through each other.*

X: Be careful. What about that ring around the equator? Remember, you mustn't tear or crease it.

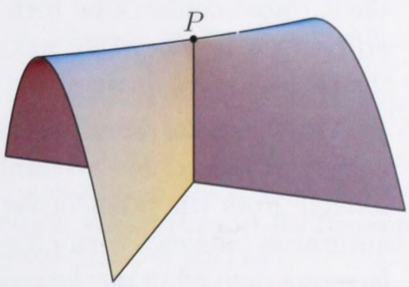
Y: *Argh—let me try again.*

X: That's no good either—you're pinching it infinitely tight.

A pinch of topology

Topology—an important area of mathematics that has applications from subatomic physics to large-scale astronomy—was born when mathematicians realized that there is something behind the notion of nearness that does not depend on measurements. This qualitative view of nearness implies that elastic distortions, compressions or expansions don't change the essence of a shape. By contrast, tearing an object changes it drastically: points on opposite sides of the tear end up apart, no matter how close together they were. Puncturing—that is, taking away a point—is also an essential change. So the rule about not tearing or puncturing the surface is a topological one.

Topology deals with all sorts of surfaces (and also curves—see page 5). A subarea called *differential topology* concentrates on surfaces that are *smooth*. Think of a mesh on the surface, like the latitude-longitude grid on the globe. Smoothness means that the curves of the mesh don't have corners. This corresponds to the intuitive idea of a smooth surface, without creases, kinks, corners, etc. At a crease (such as the equator in the figure above) it's impossible to find a smooth mesh.



It is also impossible to find a smooth mesh around a *pinch point*, like the point P on the surface to the right. This surface is called the *Whitney umbrella*, in honor of Hassler Whitney, one of the creators of differential topology. You can obtain it by taking an X made of two horizontal lines, and dragging it up while at the same time closing the angle between the lines. The center of the X traces the vertical line of self-intersections that ends at P, where the two lines collapse into one.

Any point that cannot be included in a smooth mesh is called a *singular point*. Singular points are not "bad"—in fact they can be quite interesting—but our rules don't allow them because we want our surface to remain smooth. Yorick's attempts to turn the sphere inside out by just pushing the halves through each other fail because he's creating singular points: a crease the first time around, and a pinch point the second time.

Y: *But then there's no way! It's impossible! You have to crease or pinch it to turn it inside out.*

X: It is surprising. But watch this:

Y: . . . *Is this it? Is this a sphere turning inside out?*

X: You bet. That wasn't easy to follow, was it? To figure out what's going on, let's look at something simpler: a circle.

What's in a name?

Among the following pictures, which ones show a curve?

(a) (b) (c) (d) (e) (f) (g)

We all have an intuitive idea of curves, and we don't generally worry too much about how to define them. But when we start studying curves we see that there are many ways in which we can formalize our intuitive concept. Which is the best?

The answer depends on the context. The important thing is to agree beforehand on which idea we have in mind. Mathematicians usually do this by insisting that their objects of discussion have certain properties (such as "continuous" or "smooth"), in order to make it easier to find out things about the objects—much as a telemarketer will buy directories of people with certain characteristics, the better to predict their buying patterns.

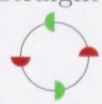
To state these properties it is convenient to think of a curve as something that you trace with a pen. From the time you put the pen to paper to the time you lift it, there is for each instant a corresponding position of the pen. A correspondence of this kind is called a *map*, and is represented by a letter, like f. A map goes from one set (here the interval of time during which the curve is being traced) to another (here the plane). For each element of the from-set (a moment of time), the map gives an element of the to-set (a point in the plane). The curve, then, is the set of points touched by the pen at one time or another—it is the part of the to-set "reached" by the map. ▸ So the curve is the image of $f : [a, b] \to \mathbb{R}^2$. From this point of view, even (e) above is a curve, with $f(t)$ the same for all t. ◂

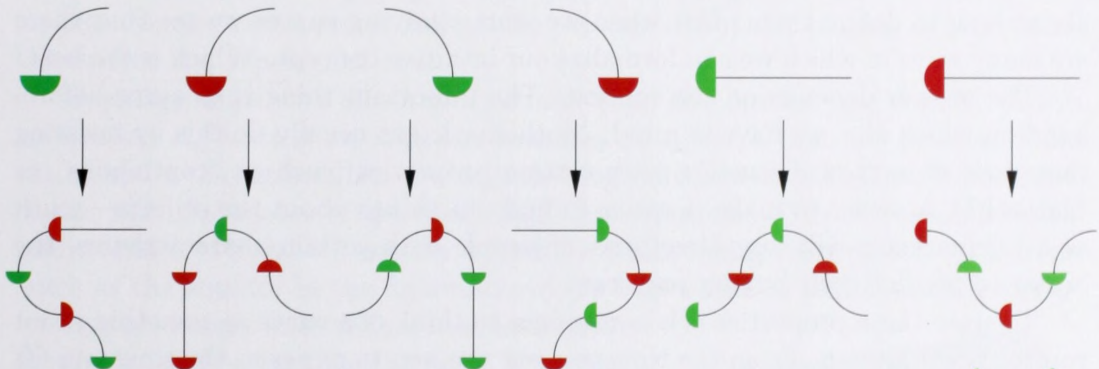
We also say that f is a *parametrization* of the curve, and that the curve is parametrized by an interval of numbers (the from-set).

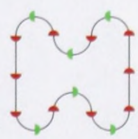
One objection to this point of view is that the parametrization contains much more information than the curve we see on the page, information that is largely arbitrary: for instance, if we trace around a curve twice as fast we get the same set of points, but the parametrization is different. This way to store extra information will actually turn out to be advantageous, but in any case I emphasize that the parametric point of view is only one possible way to formalize our intuitive idea of curves. The important thing is to have the intuitive idea clear.

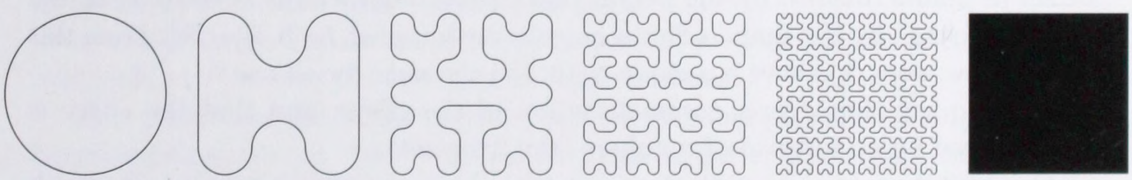
We will always want our curves to be continuous; this means the pen doesn't jump from one position to another instantaneously. (Of course a real pen could not do this anyway, but our idealized pen is not subject to such pedestrian limitations.) ▸So "continuous" properly refers to the parametrization, but we'll use it for the parametrized curve as well. We will do the same thing with other properties.◂

Which among the curves (a)–(g) of the previous page are continuous?

Giuseppe Peano discovered in the late nineteenth century that the restriction to continuous curves still allows many surprises, such as a curve that visits every single point of a square! Here's a recipe for such a curve, adapted from David Hilbert's example [Hilbert 1890, 1891] of a Peano curve. Start with an infinite supply of tinkertoys in six shapes , the length of the straight pieces being twice the radius of the curved ones. Make this initial pattern:  which represents a continuous curve, if we disregard the colored half-disks. Then replace every piece by four smaller ones, as follows:



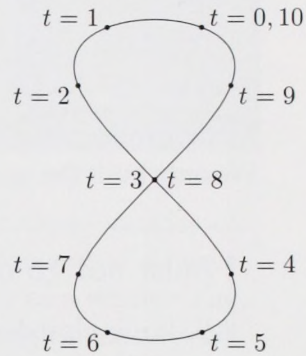
Make sure to take into account the way each piece is turned. This gives , which is another continuous curve. Apply the process again, and keep repeating it *ad infinitum*. Here are the first five steps, and the limit:



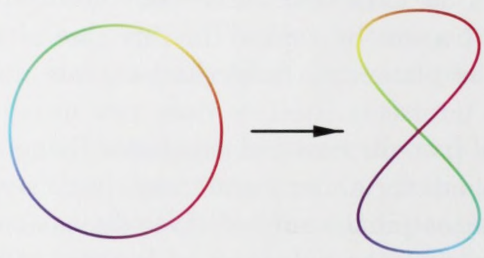
Make some twenty copies of each tinkertoy piece and try out the recipe. Can you find other replacement-based recipes that give interesting curves?

Show that the limit curve is still continuous, and that it does fill the square.

Getting back to properties. Let's agree that from now on "curve" always presupposes "continuous". The next interesting property that a curve may have is closedness—a curve is closed if the pen starts and ends at the same point. At the right is an example of a closed curve, drawn in ten seconds. We mark where we are at the end of every second, and at the end of the ten seconds we're back where we started.



The prototypical closed curve is the circle, and in general "time" for a closed curve should be thought of as a circle, since the situation at the beginning is the same as at the end. Philosophically, this cyclic view of time has often been adopted (see [Gould 1987] for some of its implications). Mathematically, it suggests that we should define a closed curve as something *parametrized by a circle*—in other words, the from-set of the parametrization should be a circle, rather than an interval of numbers. Visually, we can use instead of numbers some naturally cyclic quantity, such as color, as the parameter. There is no longer a starting or an ending point: all parameter values (colors, points of the from-set) are equivalent.



If the pen never touches the same point twice, the curve it draws is *simple.* A self-intersecting curve, like the figure-eight curve above, is not simple. For a closed curve the pen is at the same point at the beginning and at the end, but since we're thinking of these two times as the same, this doesn't count.

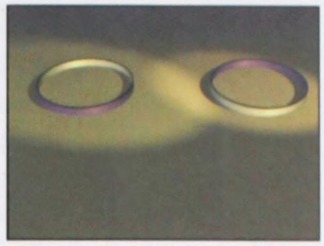
If the movement of the pen has no jerkiness—no instantaneous changes in speed or direction—the curve is *smooth.* This allows the possibility that the pen slows down, stops, and starts again in a different direction, so the curve may have a corner and still be called smooth (because the parametrization is smooth).

If not only the curve is smooth but also the pen never stops moving, even momentarily, the curve is *immersed.* This type of curve cannot have a sharp corner; if it did, the motion would either have to stop or change directions instantaneously. So immersed curves look smooth as well as being traced smoothly. An immersed closed curve is also called an *immersion of the circle.*

► More precisely, I will use the word smooth to mean that, if the curve is given by $f : [a, b] \to \mathbb{R}^2$, the coordinates of f have continuous second derivatives. For a closed curve $f : S^1 \to \mathbb{R}^2$ derivatives are should be taken with respect to the angle parameter on the circle, which is locally defined. The curve is immersed if the derivative of f is never the zero vector. ◄

Which among the curves of page 5 are closed? simple? smooth? immersed?

X: We'll build a vertical wall along the circle so that we can color the two sides differently. Can you gradually turn this circle into this other circle, where the purple and gold sides are reversed, without creating sharp corners?



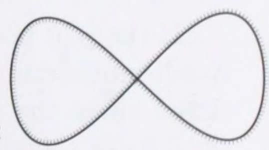
Y: *Of course! I can turn a rubber band inside out.*

X: Remember, we're really trying to turn the circle inside out. We only built the wall so we could see the different sides.

I must not go over Jordan (Deuteronomy 4 : 22)

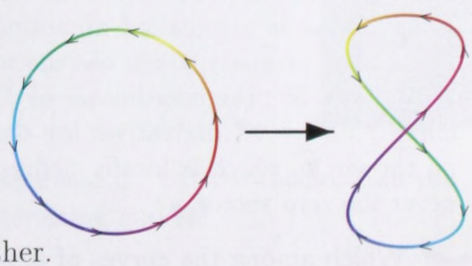
All simple closed curves have an inside and an outside. This intuitively obvious fact is actually not so easy to justify rigorously, and I won't attempt to do it here ▸ (see [Munkres 1975, §§ 8–12 and 8–13], for example). ◂ It is an early theorem of topology, and is associated with the French mathematician Camille Jordan, one of the first people to try to pin down analytically the notion of a curve [Jordan 1893, vol. 1]. For the same reason simple closed curves are also called *Jordan curves*.

It is also true that an immersed curve—not necessarily simple or closed—has two sides. What I mean by "sides" in this case is a bit subtler—it's not merely a division of the plane into inside and outside, because a curve that is not closed $\subset$ or that intersects itself does not usually divide the plane in this way. Flatlanders (two-dimensional creatures living in the plane) can see the sidedness that I'm talking about right away, just as we see the sidedness of a surface in space; but as inhabitants of three-dimensional space we must use indirect ways of visualizing the sidedness of curves. One way is Xanthippe's trick of replacing the curve by a surface. Another is to keep things in the plane, but make the curve into a flatlander's carpet, giving one side a furry texture:



Yet another way to distinguish between the sides is to *orient* the curve, that is, to paint one-way arrows along it. This distinguishes sides if we make the following *agreement*: the right-hand side of every curve (to someone who's going the right way, of course) is purple or furry, while the left-hand side is gold or smooth.

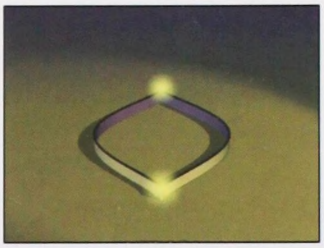
What's nice about this formulation is that the orientation comes free with the parametrization. If the curve is parametrized by time, the arrows indicate direction of movement—they point toward increasing values of the parameter. For a closed curve parametrized by the circle, we must choose beforehand an orientation for the circle; then, to transfer the arrows from the circle to the curve, we look at the way the colors are changing, like this:



Curves that look the same (as sets on the plane) but are traced in opposite directions have opposite orientations; the furry or purple side of one is the gold or smooth side of the other.

Y: *Oh, yes, the wall has to stay vertical. And it can't have creases, but it can pass through itself. Fine, let me try.*

X: Watch out! That was a sharp bend!



Homotopies (regular and decaf)

We're now going to consider curves that change gradually over time—movies of curves, so to speak—and we'll need to move beyond our basic concept of a curve as a point in motion (page 5) to consider the motion of the curve as a whole. The motion of the curve is a concept quite independent of motion along the curve, and to keep the two straight it may be a useful exercise to imagine that there are two "kinds of time", allowed to vary independently—like time within a story that is being narrated in real time.

The mathematical name for a movie of a changing curve, by the way, is a *homotopy* (of curves). The word applies to the evolution itself and also to the sequence of frames of the movie, each of which is a curve. The first and last frames are called the *initial* and *final curves*. They are *homotopic* to one another—and to all the curves in between, but generally what we're interested in are the endpoints of the process.

When necessary, we will distinguish between the two kinds of time by saying "homotopy time" versus "time along the curve", or "homotopy parameter" versus "curve parameter". You can also think of the curve parameter as color (page 7), or simply as a point of an interval or circle, if you are at home with that idea.

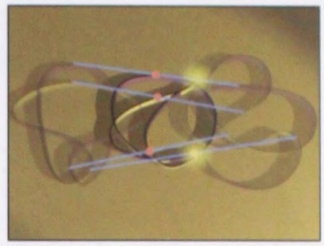
▸ Formally, then, a homotopy of closed curves in the plane is a map F from $S^1 \times [T_{\text{init}}, T_{\text{fin}}]$ to the plane, continuous with respect to both parameters together. Fixing the second parameter T gives a closed curve $f_T : S^1 \to \mathbb{R}^2$, the frame of the movie at time T. We will generally assume that the initial and final times are $T_{\text{init}} = 0$ and $T_{\text{fin}} = 1$. For open curves the situation is the same, with an interval instead of S^1. ◂

In a homotopy of closed curves, all intermediate stages are closed curves by definition. But the definition of homotopy has nothing to say about other properties of curves. For instance, a homotopy between two simple curves may well include curves that self-intersect:

In the same way, a homotopy between two immersions of the circle may include curves that are not immersed, the picture at the top of this page being an example. In a *regular homotopy* between two immersions, however, this doesn't happen: the curve remains an immersion throughout, by definition (see also page 11).

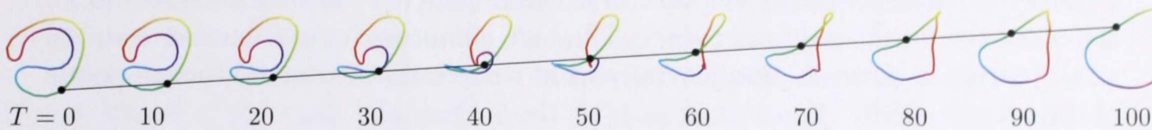
With these definitions, Yorick's task can be put very succinctly: *Find a regular homotopy between a circle traced one way and one traced the other way.*

X: If we could make sharp bends in the material, we'd be able to turn any curve into any other, by moving each point of the initial curve in a straight line toward a target point in the final curve.



The straight-line homotopy

It turns out that, given any two closed curves in the plane, there is always a homotopy between the two. In fact there are many, but one is particularly easy to construct and to explain; it is a kind of weighted average between the two curves. I will describe the frame of this homotopy at time T, where T varies from 0 to 100. For each color (curve parameter), find the corresponding points P and Q in the two given curves. Now find the point that's $T\%$ of the way along the straight line from P to Q; this will be the point of that color on the intermediate curve. Do this for all colors, and you get the whole intermediate curve. Do this for all values of T, and you get the whole homotopy. We call this the straight-line homotopy, since points of the same color describe a straight line as T varies:



This homotopy is likely not to be regular, for the following reason. Since we are averaging points of the same color on the initial and final curves, we are also averaging their velocities (rate of change of position with respect to color). So if points of the same color on the two curves happen to have velocities pointing in diametrically opposite directions, there will be some T for which the two velocities will cancel out. For this value of T the curve will not be an immersion.

Suppose the initial curve is an immersion and the final curve is its mirror image. Draw a few of the curves of the straight-line homotopy. For what value(s) of T is the curve not an immersion? What if the initial and final curves are mirror images as sets but are parametrized differently, as in the figure on the right?

▶ Homotopies of curves in the plane are a special case of the more general concept of a homotopy of maps. If X and Y are topological spaces, a homotopy of maps $X \to Y$ is a map $X \times [0,1] \to Y$, continuous in both variables. We can think of it as a deformation of one map $X \to Y$ (the initial stage of the homotopy) into another (the final stage). In the case of curves in the plane, X is an interval or the circle S^1, and $Y = \mathbb{R}^2$. The idea of a straight-line homotopy applies no matter what the space X is, because it depends only on the fact that two points of $\mathbb{R}^2$ are connected by a unique line. So any two maps $X \to \mathbb{R}^2$ are homotopic. More generally, any two maps $X \to Y$ are homotopic if Y is a *contractible* space (roughly, this means Y can be shrunk to a point). An example of a space that is *not* contractible is the torus (the surface of a donut): a circle that goes around the middle hole can never be homotopic to, say, a "constant curve" like (e) on page 5. ◀

Y: *But I can avoid corners altogether by making a loop smaller and smaller. . .*

X: That's an interesting idea, but pulling a loop tight is not really a gradual change. It's like having a corner in disguise, so it's against the rules.

Y: *Well, if you can't have corners and you can't pull loops tight, I think it's impossible to turn the circle inside out!*

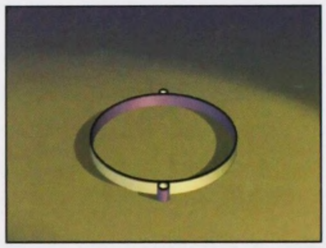
X: Yes! You're right.

Y: *Wait a minute. Am I supposed to believe that you can turn a sphere inside out, but not a circle?*

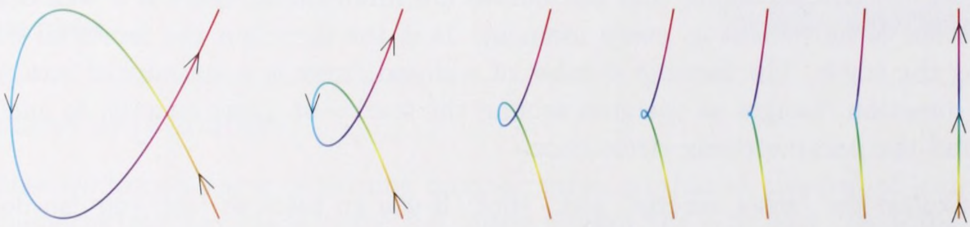
X: Yes! There is something fundamental about curves that would have to change if you were to turn a circle inside out—and that something cannot change under our allowed motions.

Y: *And what's that?*

Yorick gets an A+. Though his idea doesn't work out, it shows resourcefulness and a desire to try new things.



The whole truth and nothing but the truth



Each of the curves in the homotopy sketched above is an immersion—in fact, all but the last are obtained from one another by rescaling and reparametrization. Is the homotopy regular? It might appear so from the discussion on page 9, but I didn't give you quite the whole story there.

It's not hard to see from this figure why Xanthippe might want to disqualify Yorick's idea of pulling a loop tight. When you do that, the direction of movement changes abruptly. In the figure, we're moving *down* at the points parametrized light blue, while there is a loop; but when the loop disappears, we're moving *up* at the corresponding point. We disallow this sort of thing because we don't want jumps in the direction of movement along the curves any more than in their position itself. ▸ To spell it out: *a homotopy of curves is regular if the velocity along the curve is never zero and depends continuously on the homotopy parameter.* ◂

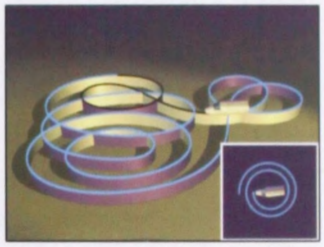
▸ What can you find out about the curvature of a closed curve that undergoes a regular homotopy? ◂

As far as the rules for curves are concerned, this is it—no more surprises. The surprise, of course, is that you can turn the sphere but not the circle inside out through regular homotopies.

X: I'll explain. Imagine a monorail atop the wall. Now the rule about monorail traffic is that the car only travels forward, and it always keeps the purple wall on its right.

This "traffic rule" is equivalent to the agreement of page 8.

We'll use a diagram to monitor the car's direction. On this track, the car is always turning left. As it goes around the circle once, it makes one full turn toward the left.



On a more complicated track, the car might sometimes be turning left and sometimes right, but the net amount of turning after one complete circuit is always some number of full turns in one direction or the other. The number of full turns it makes toward the left is called the curve's turning number. For a curve where there is more turning toward the right than toward the left, the turning number is negative.

Y: *Hey, and if there is no net turning, the turning number is zero!*

X: Right.

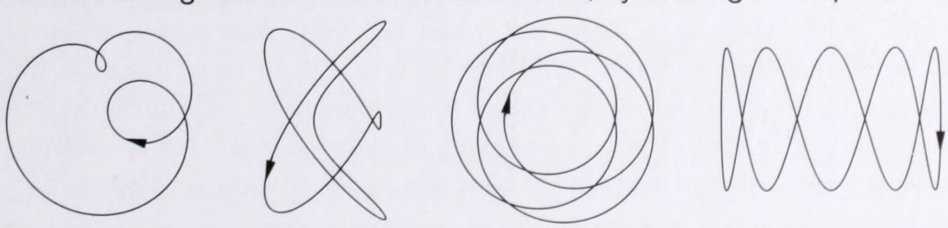
Murder on the Oriented Express

Because we are assuming that our curves are immersions, there is a well-defined direction of movement at every moment. It is the direction the monorail points along the track. The *turning number* of a closed curve is a measure of how much this direction changes as one goes around the track—or, more exactly, as one goes around the parametrizing circle once.

Explain the "more exactly" part. Hint: If you go twice as fast, you can do two circuits in the same amount of time; the parametrized curve is different even though the track is the same. What effect does this have on the turning number?

(See the figure above, and the one at the bottom of the next page.) In the movie, how does the diagram in the inset relate to the main image? Why does it show a spiral instead of a portion of a circle?

Find the turning number of these closed curves, by drawing their spirals:

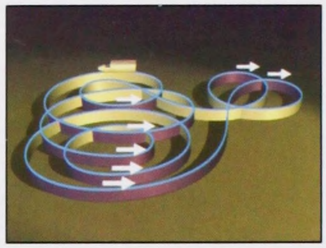


Why is the turning number always a whole number?

What happens to the turning number when the orientation of a curve is reversed? (Remember this means the curve is traveled in the opposite direction, or that its two sides are switched.)

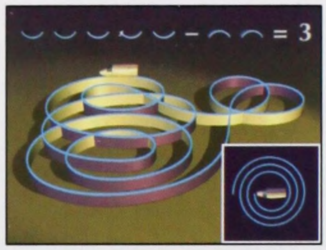
Y: *I had a hard time following the net turning for this winding track.*

X: That's natural. But there's another way to get the right answer: find the spots where you're traveling in a particular direction, like due east.



Y: *Let's see, since we're looking from the south, that would be wherever the monorail is going toward our right—where we see the purple wall face on.*

X: Exactly. At some of these points the track is curving away from us. Viewed from here, it looks like a smile; at these places, the car would be turning left. At others the track looks like a frown, curving toward us; at these points the car would be turning right. The net number of full turns increases when the car passes a smile, and decreases when it passes a frown. Starting at zero... one... two... three... four... three... two... and we finish with three. The turning number is the number of smiles minus the number of frowns.



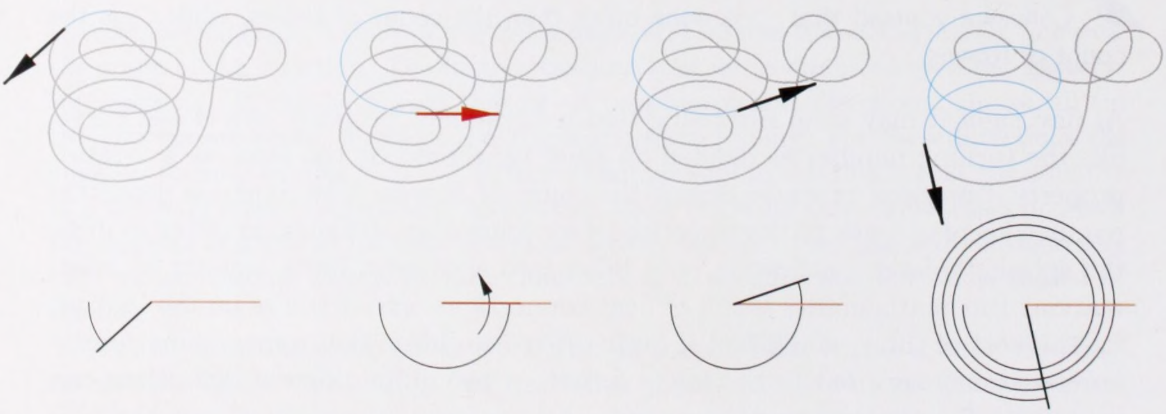
Y: *I see... the turning number measures happiness!*

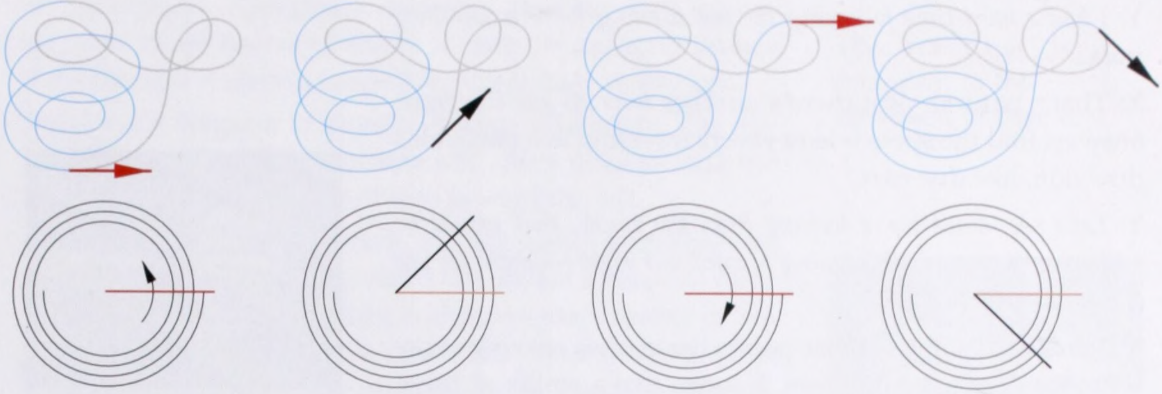
X: If you insist...

The laws of (e)motion

Suppose we have a curve of turning number three, so that at the end of a circuit the monorail car has made three full turns toward the left, and the spiral has spiraled counterclockwise a net total of three times.

Let's monitor each time the spiral reaches a given compass point, which is to say each time the car is moving exactly in a given direction, say due east. We call this direction a marker, and draw it in red so it will stand out:





We now count separately the times when the spiral sweeps past the marker going counterclockwise (as in the first diagram above) and those when it sweeps clockwise (as in the third diagram).

Because the net amount of spiraling is three turns counterclockwise, this must mean that all but three counterclockwise passes are canceled by clockwise passes. We know this is true even if we don't know the actual count of counterclockwise and clockwise passes. That is, we can write

$$\circlearrowleft - \circlearrowright = 3,$$

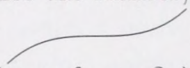
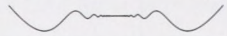
where the symbol $\circlearrowleft$ stands for the number of counterclockwise passes and $\circlearrowright$ for the number of clockwise passes, whatever they may be.

What replaces the equality $\circlearrowleft - \circlearrowright = 3$ when the turning number is something other than three? Can you see why our argument depends on the integerness of the turning number?

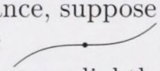
This also means (as Yorick finds out) that if we do know $\circlearrowleft$ and $\circlearrowright$ for a particular curve, we also know its turning number, even without having to draw the spiral.

Convince yourself that $\circlearrowleft$ is none other than the count of smiles, while $\circlearrowright$ is the count of frowns.

At first sight it may seem surprising that a "dynamic" property of a closed curve, like the turning number as defined on page 12, should be the same as a "static" property, the count of smiles minus the count of frowns. The dynamic definition requires keeping track of the direction of movement at all times, in order to draw the spiral; the static definition only takes into account a finite amount of information. But mathematics is full of neat connections, and if you're on the lookout for this sort of thing, you'll find it quite often: two ideas that appear superficially unrelated may turn out to be closely linked, or two quite different definitions can lead up to exactly the same thing.

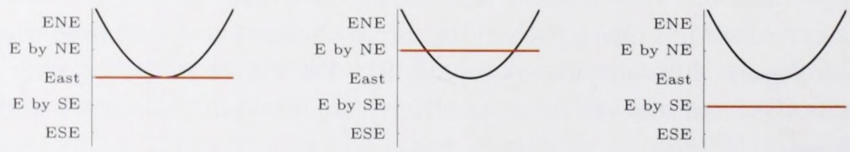
► I've glossed over several points in justifying the equality *turning number* $= \circlearrowleft - \circlearrowright$. For one thing, what happens if the very beginning of the circuit—and therefore also the end—is a smile or frown? And what if the spiral reaches the marker, stops, and backtracks, so the track forms neither a smile nor a frown?  And how can we even be sure that there isn't an infinite number of smiles or frowns? 

The funny situations discussed in the previous paragraph shoot holes in the equality *turning number* $= \circlearrowleft - \circlearrowright$. Can you modify the definition of $\circlearrowleft$ and $\circlearrowright$, or the equality itself, to make it work in every case? Can you think of any other holes?

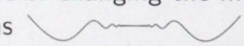
If you tried to crack this walnut, you know that patching the holes in the equality is possible but messy. There is another alternative, namely, to choose the direction marker so that none of the funny situations occur. For instance, suppose that one of the points where the track runs due east is an *inflection point* . If we choose east as the marker, we have a funny situation, but if we choose a slightly different marker, east by northeast, all is well again:

Is all well if we choose east by southeast?

We now draw a graph of the direction of the monorail car as it goes past the inflection point. Time is the horizontal coordinate, and the direction is the vertical coordinate. We draw several copies, with different red horizontal lines representing possible marker directions.

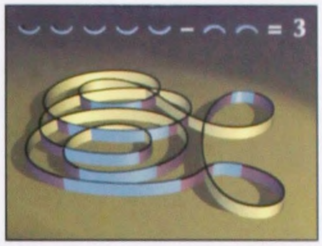


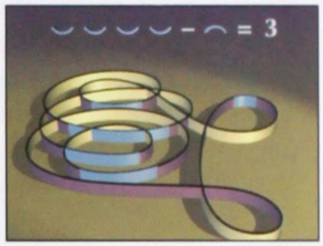
Note that, when we choose east as the marker, the horizontal line is tangent to the graph. When we choose some other direction as the marker, the line intersects the graph transversely, or not at all.

Study the effect of changing the marker slightly when the track has an infinite cluster of smiles and frowns .

In technical language, a smile or frown is a *transverse intersection* of the marker line with the graph of the direction. The funny situations that we've discussed are *nontransverse intersections*. It can be proved that of all possible marker directions, almost all (in the appropriate sense) intersect the graph transversely only. For these directions the equality *turning number* $= \circlearrowleft - \circlearrowright$ is true. Another way to say this is that the equality is *generically* true. "Generically" means that there might be exceptional cases in which the equality is not true (or does not make sense), but that the chance of encountering such an exceptional case is literally infinitesimal if the direction marker is chosen at random.

It may seem like a cop-out to work with an equality that is true most of the time, rather than all the time, but in fact the notion of genericity is very useful and appears over and over in differential topology and other areas of mathematics. We will come back to it on page 22. ◂





X: Now the nice thing about the turning number is that it remains the same when a curve changes according to our rules. Frowns and smiles can appear or disappear, but only in pairs that balance out. The number of smiles minus the number of frowns never changes.

Y: *So a curve can only turn into another curve with the same turning number?*

X: Right: the turning number is the fundamental property I mentioned before. Now what's the turning number for the two circles?

Y: *Hmm... This one has one smile and no frowns, so the turning number is one. And if the gold is outside—one frown and no smiles—minus one! It makes sense—on one curve you're turning left all the time, on the other it's the opposite.*

X: Good. So the reason you cannot turn a circle inside out...

Y: ... *is that that would change the turning number!*

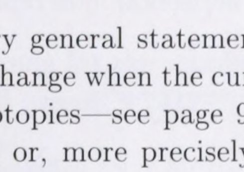
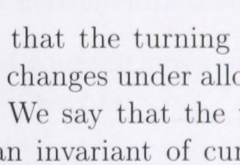
Plus ça change, plus c'est la même chose

Xanthippe makes a very general statement: that the turning number of an immersed curve does not change when the curve changes under allowed motions (also known as regular homotopies—see page 9). We say that the turning number is an *invariant* of curves, or, more precisely, an invariant of curves under regular homotopies.

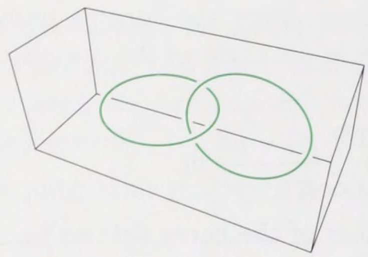
Invariants are very important in topology. Usually, as here, they help rule out the possibility that two objects are the same, or that they can be turned into one another in some precise sense (that depends on the context).

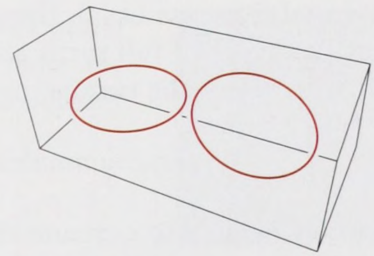
For instance, suppose we are interested in *links*, as in the video *Not Knot* [Gunn and Maxwell 1991] (see also the references mentioned there). A link is simply a collection of *separate* simple closed curves in space. We think of two links as being the same if we can give them the same shape by moving them around in space—without cutting them open, or making strands cross, or anything like that. The link has to remain a link throughout the move.

It is easy to see that the number of pieces in a link is an invariant, so that this green link, which is composed of two curves, can never be the same as the blue one, with three curves:

 $\neq$ 

This is a very coarse invariant; it does not distinguish between the next two links, which are pretty obviously different:

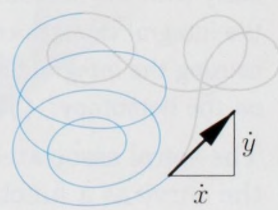


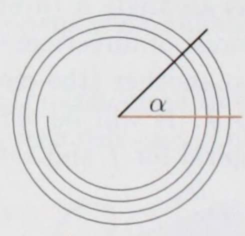


Another invariant comes to the rescue: the *linking number*, which is 1 for the green link and 0 for the red link. This may sound like a pedantic way of saying that the two red loops are not linked at all, while the green ones are. But the point is that the linking number can be defined quantitatively; there is a formula to compute it that gives a definite result. The alternative, trying to take the loops apart by playing around with them, carries no guarantee with it: if we don't succeed in taking the loops apart, is it because there really is no way, or just because we weren't clever enough? Many fascinating mechanical puzzles are based on hard-to-find link manipulations.

Back to plane curves. Xanthippe does not justify her statement about the invariance of the turning number, although the example in the movie makes it intuitively plausible. A convincing justification is most easily given by using calculus to write down a formula for the turning number. If you're not familiar with calculus, skip to "The spiral..." near the bottom of page 18, and take the invariance on faith.

Suppose our curve is parametrized by $f(t) = (x(t), y(t))$, where x and y are real-valued functions of the parameter t. Remember from page 7 that for a closed curve "time" is defined on a circle. But to do any calculations we must work with real numbers, so we think of t as the angle parameter around the circle, going from 0 to 2π (which represents a full circuit).





The angle coordinate α of the endpoint of the spiral at some given time satisfies $\tan\alpha = \dot{y}/\dot{x}$, where $\dot{x} = dx/dt$ and $\dot{y} = dy/dt$ are the coordinates of the monorail car's velocity vector. This does not completely define α, since if we change α by adding or subtracting 2π (a full turn) the tangent does not change. Indeed the whole question is by how many full turns α increases or decreases as time goes from 0 to 2π. Fortunately, we can express the *rate of variation* of the angle α, rather than α itself, in a way that's unambiguous:

$$\frac{d\alpha}{dt} = \frac{d}{dt}\arctan\frac{\dot{y}}{\dot{x}} = \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{\dot{x}^2 + \dot{y}^2}.$$

Note how the fact that the curve is an immersion ensures that the denominator is never zero, so the formula makes sense for all values of t. Now, the total amount of spiraling is the accumulated variation $V = \alpha(2\pi) - \alpha(0)$ of α as t goes from 0 to 2π:

$$V = \int_0^{2\pi} \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{\dot{x}^2 + \dot{y}^2}\,dt,$$

where $\ddot{x}$ and $\ddot{y}$ are, as usual, the second derivatives. Since the turning number is measured in units of full turns and one full turn is an angle of 2π, we also know that V is 2π times the turning number. So

$$\textit{turning number} = \frac{V}{2\pi} = \frac{1}{2\pi}\int_0^{2\pi} \frac{\dot{x}\ddot{y} - \dot{y}\ddot{x}}{\dot{x}^2 + \dot{y}^2}\, dt.$$

Use this formula to compute the turning number of the curve defined by $x(t) = \cos t$ and $y(t) = \sin t$. What is this curve?

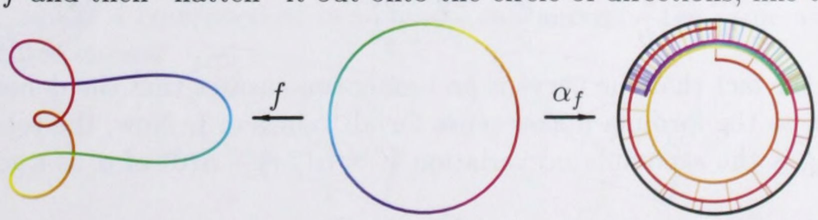
► (For readers who like calculations.) The rightmost curve on page 12 is defined by $x(t) = \cos t$ and $y(t) = .65 \sin 5t$. Plug these expressions into the formula for the turning number and perform the integral. Does the result agree with what you got by drawing the spiral? ◄

The merit of this formula is not so much as a practical way to compute the turning number of particular curves, but rather that it makes it clear that the turning number changes continuously with the curve. That is, if the functions x and y change by a little bit, the integral also changes by (at most) a little bit. But we know that the turning number cannot change by a little bit: the smallest amount by which it can change is 1, because it is always an integer. Therefore it does not change at all! Since any regular homotopy is the accumulation of many tiny changes in the curve, we conclude that *the turning number cannot change at all in the course of a regular homotopy.*

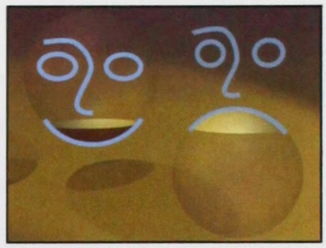
► Where in this proof of the invariance of the turning number did we use the regularness of the homotopy? (If you have trouble, read the beginning of the next question.) ◄

► According to the definition of a regular homotopy on page 9, $\dot{x}$ and $\dot{y}$ change continuously with the homotopy parameter, and this is clearly necessary if we are to conclude that the integral changes continuously. But what about $\ddot{x}$ and $\ddot{y}$—couldn't they change abruptly, causing the integral to jump? Can the proof be fixed without imposing additional conditions on the homotopy? (This requires some nonelementary analysis.) ◄

The spiral associated with a closed curve encodes the direction of movement along the curve as a function of the curve's parameter. If we interpret the direction not as an angle α (a number) but as a point in the circumference of a compass—the circle of directions—we get a map from one circle (which parametrizes the curve) to another (the circle of directions). We call this map α_f if the immersed curve is f. It will be very important later. Another way to think of α_f is to take the spiral for f and then "flatten" it out onto the circle of directions, like this:



► Write down $\alpha_f(t)$ in terms of the derivatives $\dot{x}$ and $\dot{y}$ of the coordinates of f. ◄



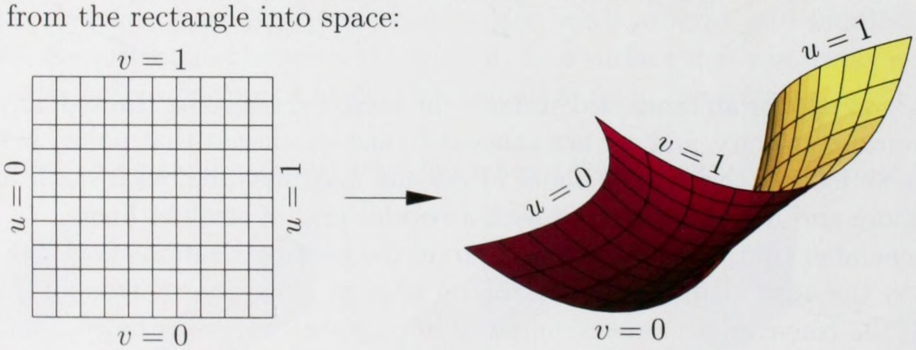
Y: *But wait—doesn't the same argument prove you can't turn a sphere inside out? This sphere has a three-dimensional smile, and this one has a three-dimensional frown. So they have different turning numbers!*

X: Not quite. Your analogy is good…

Some superficial remarks

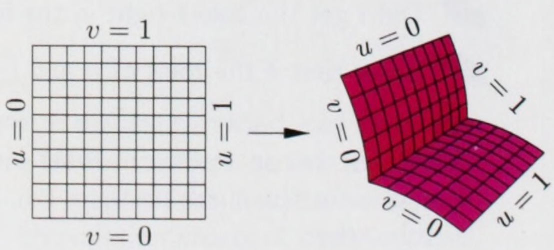
During the making of *Outside In*, something that tripped us up repeatedly was the question of the "turning number" for surfaces. Although we all had read the script dozens of times—and some of us had written it—we would still get confused, even the mathematicians. Only after everyone had worked out a number of examples did the confusion go away. So although in the movie the subject is necessarily treated very briefly, a flesh-and-bone student (as opposed to the fleshless Yorick) would want to go at a more leisurely pace.

I'll start by spelling out what our surfaces are like. On page 5, we adopted a parametric point of view: a curve is something parametrized by an interval of numbers (I'm leaving closed curves aside for the moment). In the same spirit, a surface should be something that can be parametrized by the two-dimensional analog of an interval, which is a rectangle. A point in the rectangle has two coordinates, across and up, or u and v; we draw a mesh to emphasize the coordinates. The mesh can also be used to show the parametrization of the surface, that is, the map from the rectangle into space:



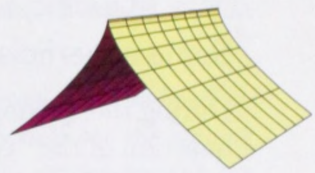
Note that a parametrized surface can intersect itself: this happens when different points in the rectangle map to the same point in space.

We want our surfaces to be *smooth*, in a sense that is reasonably close to our intuition about the word. The surface above is smooth, whereas this one is not: Formally, we say that a parametrized surface is smooth when any smooth curve in the rectangle maps to a smooth curve in space. When a surface is not smooth, this is generally obvious from the mesh lines, as it is on the right: observe how the rectangle's horizontal lines, which are smooth, are mapped to lines that have corners.

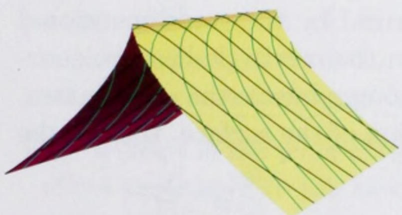


For a smooth curve we have the notion of the velocity of movement along the curve. I mentioned on page 7 that a curve may be smooth (in the sense of a parametrized curve) but nonetheless have a corner, if the pen stops gradually and then starts off in a different direction. To avoid this, we imposed the rule that our curves must be *immersed* (their velocity is never zero).

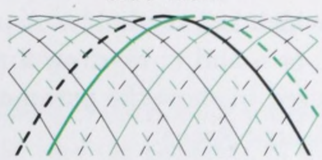
We need an immersion requirement for surfaces too, to rule out things like this:



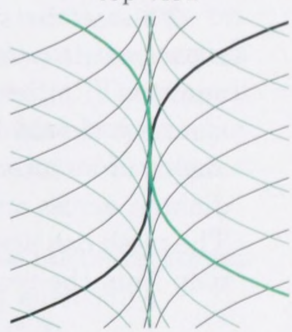
This requirement involves the movement along the curves of the coordinate mesh. The two mesh lines going through a point of the surface determine two velocity vectors. For the surface to be an immersion, neither vector can be zero, otherwise the mesh gets scrunched up, as in the picture just shown (note how the mesh line forming the front edge slows down as it approaches the crease). Even when neither velocity vector is zero, the mesh gets scrunched up if the two velocities point in the same direction or in opposite directions, so we disallow this case too. Note, in the three figures below, which show the same surface of the picture above with a different parametrization, how the mesh lines are immersed, but because they are parallel at certain points the surface is not:



side view



top view



To sum up: for an immersed surface, the mesh curves going through any point have nonzero velocity, and are not tangent to one another at that point. It follows that as we look at the surface under increasing magnification, each small bit of it looks more and more like a plane with a regular grid of straight lines.

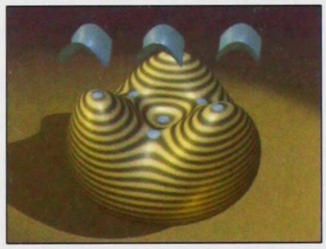
Remember that, for an immersed curve, the parametrization gives a choice of sides (by the agreement, or convention, on page 8). The same is true for surfaces, though the convention is more complicated to state. Suppose the coordinates of the parametrizing rectangle are called u and v. Place yourself so that your line of sight is along the u-vector, and so that the v-vector is pointing up. Then the purple or furry side is on your right. (By the u-vector I mean the velocity vector in the direction of increasing u and fixed v.)

Did I get the colors right in the figures of the previous page?

Show that if the roles of u and v are interchanged, the orientation changes.

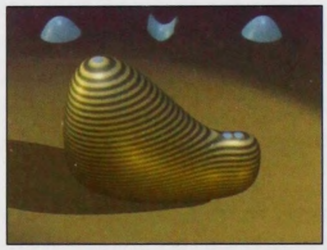
After all this background, we're ready to discuss the smiles and frowns for surfaces. We choose a plane in some direction (this corresponds to choosing a marker direction for the curve on page 13). Let's say this plane is horizontal, but it could be any other.

X: . . . but to make it complete, we must look at a general surface and consider all the points where it is horizontal and gold is on top. We'll draw horizontal stripes to make these points easier to locate. Smiles are like bowls, curving up; frowns are like domes, curving down. But there are other points where the surface is horizontal that are neither bowls nor domes. They are saddles, and look like smiles from one direction and frowns from another. Near a bowl or a dome, the horizontal stripes form rings. Near a saddle, they form an X.



Y: *But how does that change anything? Spheres don't have saddles!*

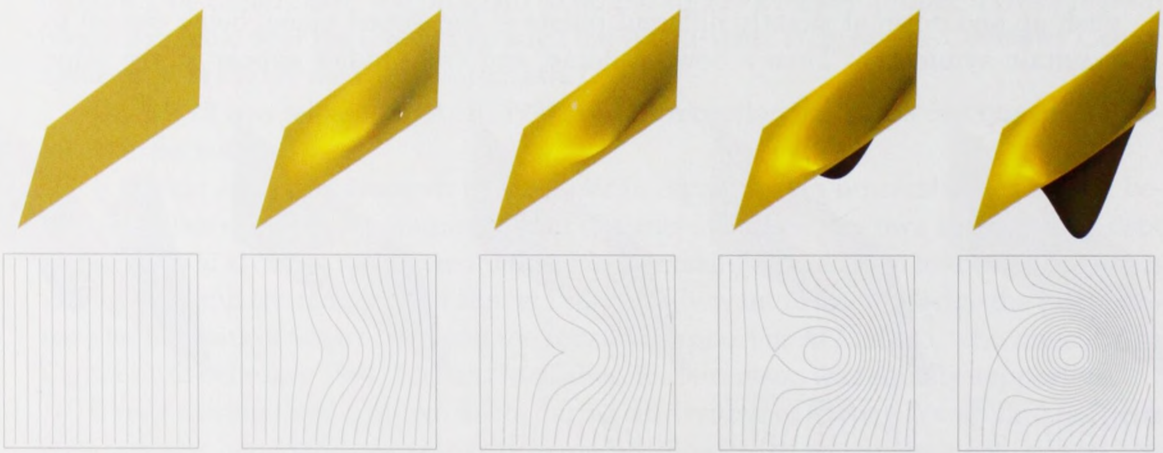
X: Ah, but the point is how these features interact. Look: a dome and a saddle can come together and cancel out. Likewise, a bowl and a saddle can cancel out. But bowls and domes, like electrical charges of the same sign, normally don't get near each other.



An outline of contours

We will look at all the points where the tangent plane is in the same direction as the marker plane—that is, horizontal—and where gold is on top (remember that the coloring of sides follows from the parametrization). As Yorick finds out, these points come (generically) in three types: bowls, domes, and saddles. I will illustrate the interaction between the types using a surface that's basically a sloped plane, with various features added. This will allow us to concentrate on the "local picture".

A bowl and a saddle appear together when I push down on the plane:

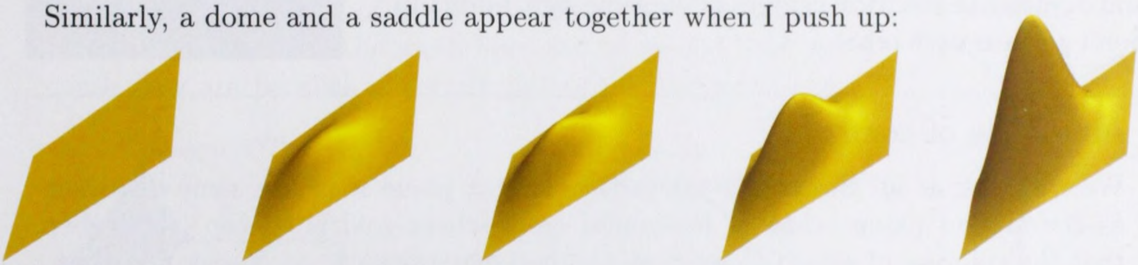


The funny curves above, and also the stripes on the surfaces at the top of the page, are lines of equal elevation, or *contours*. Find a map—in the everyday sense, not the mathematical one—that features such lines. (Navigation charts of shallow water usually have lines of equal depth, and will do just as well.) Locate some domes and saddles by looking at the contours. Note that the saddles are apt to look like this; you need to read between the (contour) lines in order to see the X. Can you find any bowls? Can you tell apart bowls and domes without looking at the elevations?

► In the third map of the previous page, there is a contour line that is not smooth. This is a sign that there is something funny going on at that moment, similar to the situations for curves at the top of page 15. What we have here is a horizontal tangency that's not a bowl, a dome, or a saddle. The bowl and saddle are born together from this point as I keep pushing.

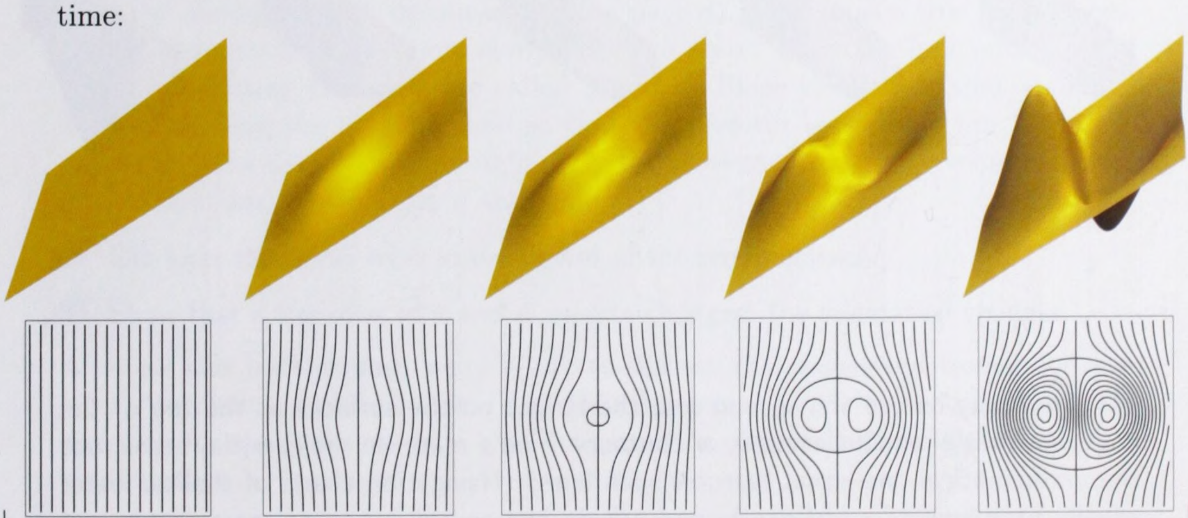
Reread the discussion on page 15, and try to extend it to the case of surfaces. Hint: consider the map that associates to each pair of coordinates (u, v) the normal vector N at the corresponding point of the surface. Bowls, saddles and domes come out as transverse intersections of this map's graph with the plane defined by $N =$ up; other types of horizontal tangencies correspond to nontransverse intersections. For more background on transversality, see [Guillemin and Pollack 1974, pp. 27–32]. ◄

Similarly, a dome and a saddle appear together when I push up:



Draw approximate contour maps for these surfaces.

On the other hand, bowls and domes don't like to get near each other. The only way that we can make them appear or disappear together is with the help of two saddles, and even then we have to be pretty crafty in order to achieve this ► (in other words: generically, this situation does not occur). ◄ One way to do it is to push up and down at slightly different points of the sloped plane, being careful to maintain symmetry. Then a bowl, a dome, and two saddles appear at the same time:



The Making of *Outside In*

In 1987, the National Science Foundation funded the Geometry Supercomputer Project, following a proposal by Al Marden and a group of distinguished mathematicians and computer scientists from several institutions around the country and in Europe. I was given the wonderful opportunity of working for the Project since its foundation. Early in the following year I undertook to make an animation to explain Bill Thurston's sphere eversion method, which I had learned from him years before. Charlie Gunn, the other staff member in those early days, had developed rendering software for the Pixar Image Computer (a dedicated graphics engine), and together we made a five-minute narrated tape showing the basic elements of the eversion.

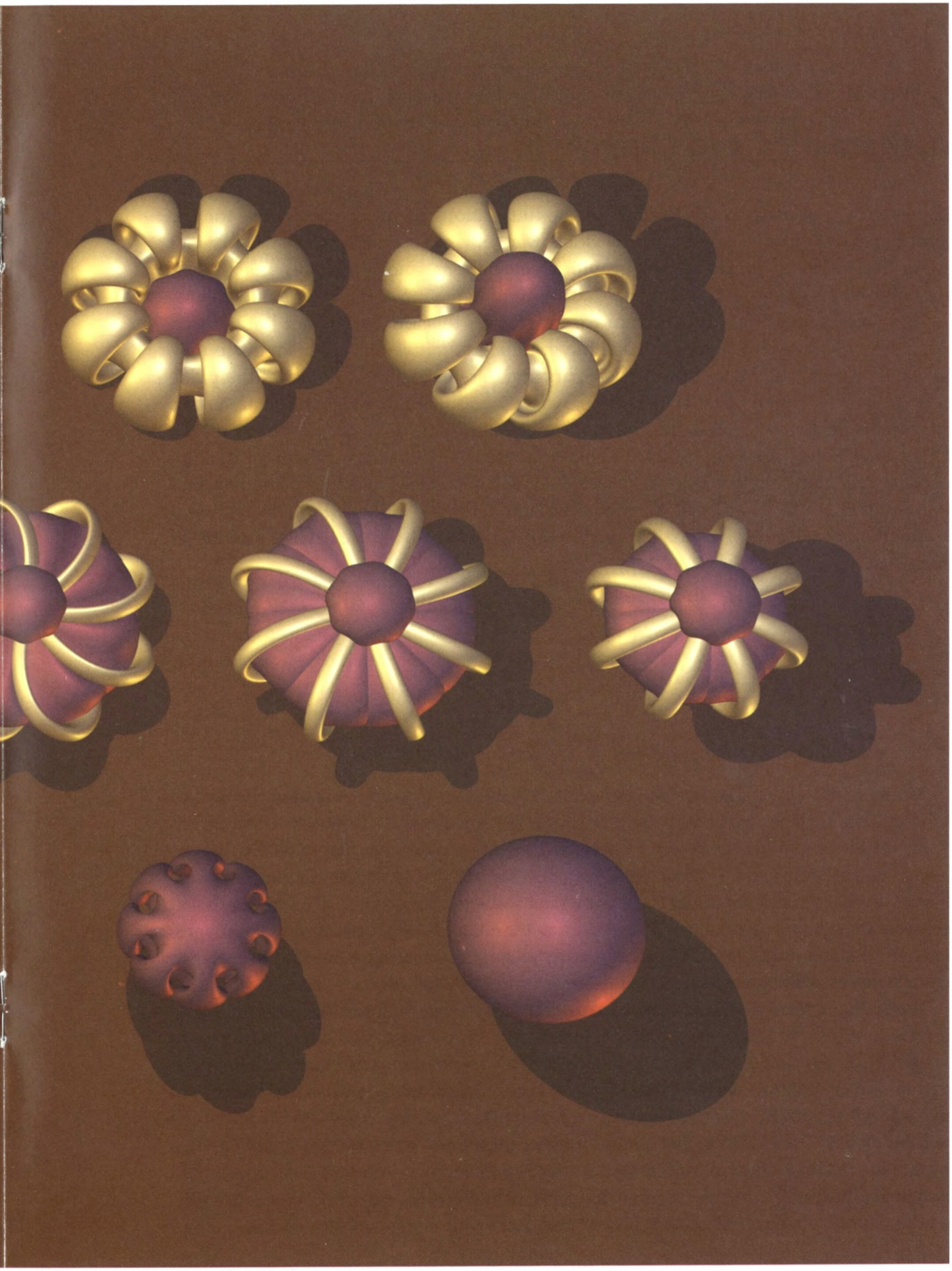
More pressing projects soon claimed our attention, however, and work on what eventually became *Outside In* was not attempted again until early 1991, when David Ben-Zvi and Matt Headrick, then students at Princeton University, started working with me. By then software and hardware had evolved significantly, and it did not seem unreasonable to aim for a production-quality, feature-length animation, especially in view of the success of *Not Knot*, which had just been released, and of the transformation of the Geometry Supercomputer Project into the Geometry Center, one of whose raisons d'être is the communication of mathematics to a wide public.

Delle Maxwell, one of the directors of *Not Knot*, and Tamara Munzner, from the Geometry Center staff, agreed to direct the new movie with me. More people joined the team in 1992, notably Nathaniel Thurston (Bill's son), then on leave from his graduate studies at Berkeley. Nathaniel wrote or rewrote much of the software used in the animation, including its centerpiece: the program to do the sphere eversion proper. (This program is now available over the Internet through the World-Wide Web; see the Geometry Center home page at http://www.geom.umn.edu/.)

Outside In was released in April 1994, and its reception has surpassed our most optimistic expectations.

It is clear from this synopsis that *Outside In* represents an unusual collaboration between mathematicians, programmers, and designers. Its directors owe an enormous debt of gratitude to all those who were involved in its making. In addition to the movie credits (see page 2) and the names cited above, I specifically want to acknowledge the extraordinary helpfulness of Stuart Levy and the constant support of Al Marden, who directed the Center until February 1994. Richard McGehee, his successor, was equally supportive.

I have accumulated my own debts during the writing of *Making Waves*: first to Sheila Newbery, for her triple role as a critical reader, a source of encouragement, and an understanding wife; then to David Epstein, George Francis, Tamara Munzner, Tony Phillips, and Bill Thurston, whose comments on various drafts have made a great difference. I also thank the publishers, Alice and Klaus Peters, for their patience during the long gestation of this work.



Mathematica and *Outside In*

Mathematica was used to create most of the illustrations in this book, except for the centerfold and the frames taken from the video. Mathematica is a comprehensive software system for symbolic, numerical and graphical calculations, and it includes an excellent programming language and a front-end that allows the creation of interactive documents.

One great strength of Mathematica for graphics is that the user is not limited to a fixed menu of plots, but is free to combine graphics primitives in order to create new objects. For example, Mathematica has a function ParametricPlot that draws the plane curve defined by some formula $(x(t), y(t))$, where x and y are real-valued functions of a real variable. Several aspects of the plot, such as line thickness, can be controlled by means of built-in options. More importantly, although there is no built-in provision for making the color vary along the curve, as on pages 7–11, Mathematica's graphics primitives and programming language allowed me to write a new function, RainbowPlot, that draws a parametric plot with color as the parameter. I can now easily create a large number of such plots:

RainbowPlot[{Cos[t], Sin[t]}, {t, 0, 2 Pi}]

RainbowPlot[{2 Sin[t], Sin[2 t]}, {t, 0, 2 Pi}]

Mathematica also played a role in *Outside In* itself. It was used in creating the shapes of the striped surfaces (page 21), the belt (page 38), and several of the curves on which the monorail runs. These geometric objects were then interactively placed in the desired position using Geomview, a viewing program developed at the Geometry Center, and were photorealistically rendered using the commercial software Renderman. The software that generates the sphere eversion itself is written in the C and C++ programming languages, but again Mathematica was helpful in the early stages of modeling, since it allowed us to obtain samples of the geometry with little programming effort.

While I have tried to give a taste of the use of Mathematica in the areas of graphics and programming, the program has many other strengths. Indeed, it is used by professionals in every discipline where mathematics plays a role, from engineering and physics to financial analysis and cryptography. The reader interested in more information about Mathematica can contact Wolfram Research via e-mail at info@wri.com, or call 1-800-441-MATH, or visit the Mathematica World Wide Web home page at http://www.wri.com/. There are also over a hundred textbooks on Mathematica and its applications: I particularly recommend [Gray and Glynn 1991, Maeder 1991, Wagon 1991, Wickham-Jones 1994, Wolfram 1991], and the periodicals *The Mathematica Journal* and *Mathematica in Education*.

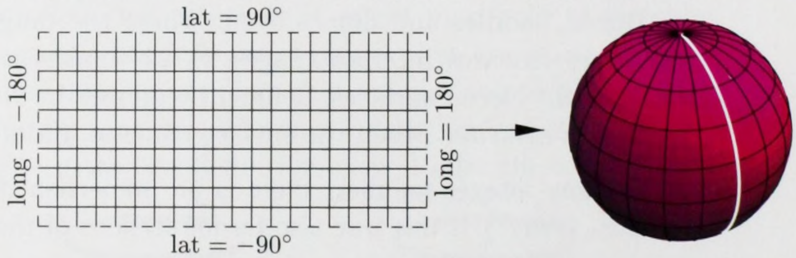
X: The unchanging number for surfaces, then, is this: add domes and bowls, and subtract saddles. This number is 1 for the sphere no matter which face is out!

This number is unchanging for *closed* surfaces.

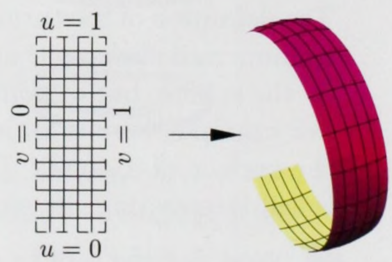
The global village

The surfaces of pages 19–22 have edges, coming from the sides of the parametrizing rectangle. It is easy to see that domes, bowls and saddles can individually "fall off the edge" of such an open surface. To get a turning number that is invariant we need to take *closed surfaces*, which are analogous to closed curves.

Closed surfaces can't be described by a single coordinate mesh. For instance, take a sphere, which is after all a pretty simple surface. Everyone knows that we can use geographic coordinates, latitude and longitude, to name points on the surface of the earth. But not all points. First, we can't sensibly assign a longitude to the poles. Second, as we fly around the world, somewhere over the Pacific Ocean we jump from longitude 180° E (or +180°) to 180° W (or −180°). In these cases, the parametrization fails to account for the nearness of points on the surface, and loses its mathematical usefulness. The upshot is that, to name all points on the sphere, we need at least two *local coordinate systems*. We can use the latitude-longitude coordinate system to name points away from the left-out meridian and from the poles.



But we also need another coordinate system covering at least a strip around the left-out meridian and the poles: Now all points of the sphere can be named in at least one, and sometimes in both, coordinate systems. The sphere itself can be thought of as being made up of two overlapping pieces, or *coordinate patches*, glued together in an appropriate way.



Now, we understand the round sphere pretty well, and we know that the coordinate systems are not really an intrinsic part of it (though they are undeniably useful for calculations). Moreover, what interests us most are changing surfaces, and it is unpleasant to have to fuss about coordinates throughout the evolution. So it's best to imagine a model sphere—the Platonic sphere, as it were—and a map from this model sphere into three-dimensional space. The resulting set of points in space is a surface *parametrized by the sphere*.

This may sound complicated, but it's the same idea that's on page 7, where I defined a closed curve as something parametrized by a circle. Our intuition about the circle and closed curves is pretty good, so I didn't include at that point any mention of local coordinates for the circle, only some "deep thoughts" about cyclic time. But for surfaces it's more important to do things carefully.

Anyway, we want our closed surfaces to have the same sorts of properties discussed on page 20 for surfaces parametrized by a rectangle: they should be smooth and immersed. These are *local* properties, that is, properties that can be tested by looking at a small piece. Once you choose a local coordinate system on the sphere, you can pretend that (a piece of) the sphere is a rectangle, and apply the definition of smoothness or immersion to a surface parametrized by the sphere just as if it were parametrized by a rectangle.

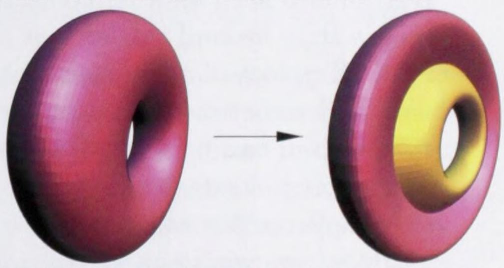
So, to spell it out: an *immersion of the sphere* is a map from the sphere into space that is smooth (lines of the sphere's local coordinate meshes are mapped to smooth curves), and is locally an immersion (the mesh doesn't get scrunched up). An immersion of the sphere looks smooth in the everyday sense. It has a tangent plane everywhere, just as an immersed curve has a tangent line everywhere.

Now I have defined fairly rigorously the surfaces that are the main subject of the movie. A continuous change in the surface is a *homotopy* of the surface. The homotopy is *regular* if the surface remains an immersion throughout, and if the velocities along the mesh also change continuously with the homotopy parameter. This last condition is the same as saying that the direction of the tangent plane changes continuously with the homotopy parameter.

Bowls, saddles and domes occur where the tangent plane is horizontal. The way these features interact (pages 21–22) makes it plausible that the "turning number" for closed surfaces, defined by adding bowls and domes and subtracting saddles, is invariant when the surface changes under regular homotopies.

For any integer number, there is an immersion of the circle with that turning number. (Why?) Is this true also for immersions of the sphere?

The definition of the turning number via bowls, domes and saddles can be extended to immersed surfaces that are not parametrized by the sphere, but by some other closed surface. For example, here is an immersion of the torus, the surface of a donut. (The immersed surface is self-intersecting: its cross section is an 8.)



Work out the turning number for this surface.

Suppose you count as smiles and frowns for a closed curve *all* the points where the tangent is east-west, regardless of which way the train is moving. What do you get by subtracting frowns from smiles? Try a few examples. Same question for closed surfaces: If you count *all* points where the tangent plane is horizontal regardless of which face is up, what do you get by adding bowls and domes and subtracting saddles?

This question is the province of *Morse theory*, an important area in differential topology. See [Milnor 1963] for a good introduction. Incidentally, there is a neat movie by Marston Morse, the inventor of this theory, about the alliterative version of bowls, domes and saddles [Morse 1966].

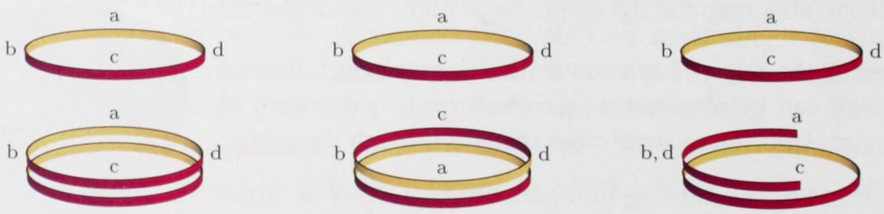
► The turning number for any immersion of a sphere with g handles is $1 - g$. ◄

All things . . . near or far / To each other linkèd are

To wrap up this discussion, I will talk about a very important concept in differential topology, the *degree* of a map, which is really where the turning number of closed curves and surfaces comes from. Before reading this section, go back to pages 13–14, and reread also the second-to-last paragraph on page 18 ("The spiral. . . "). Try to understand what the map α_f introduced there does, and what the figure shows. If you can't understand it, even after some thinking, you should ask for help or skip to page 31.

The map α_f takes a circle to another. Whenever we have a map between two smooth closed objects of the same dimension (two closed curves, or surfaces, or higher-dimensional things), we can ask "How many times does the map wrap one object around the other?"

More precisely: let's call the objects X and Y, and suppose they are painted gold and purple on opposite sides. ► (That is, X and Y are oriented, compact, smooth manifolds. We also assume Y is connected.) ◄ Let's also take a smooth map g from X to Y. If we choose almost any point y of Y, there will be finitely many points of X mapping to y (possibly zero!). At some of these points g might map gold side to gold side, *preserving orientation*, and at others it might map gold to purple, *reversing orientation*. The degree is the number of points that are mapped to y preserving orientation, minus the number of points that are mapped to y reversing orientation. The wonderful thing is that this difference is independent of the point y that we chose on Y! To try to make this definition clear, I offer three pictures:



In each picture, the top curve is the from-circle X, the bottom curve is the to-circle Y, and the middle curve shows the action of the map (by matching letters). The vertical dimension has no meaning: it just helps visualize the map. The coloring of the from-circle and the to-circle is given a priori, but the coloring of the middle curve is determined by the map, just like the arrows on the to-curve of the figure at the bottom of page 8. The map on the left preserves orientation (colors match); the one in the middle reverses orientation; and the one on the right preserves orientation in the arc abc and reverses it in the arc cda.

Find the degrees of these maps. Try to think of other maps of the circle to itself, and find their degrees.

Suppose f is an immersed closed curve, and let α_f play the role of g in the definition of the degree. Convince yourself that the degree of α_f is the turning number of f. Hint: the role of y is played by the marker direction of page 13.

Draw the α_f's corresponding to the first two maps of the previous page, and find their degrees. Is α_f defined for the third map?

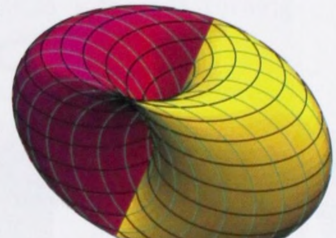
If an immersed closed curve f undergoes a regular homotopy, the corresponding α_f changes continuously. (Why?) It is a fundamental fact of differential topology that the degree of a map does not change under continuous changes (homotopies). This gives another proof that a regular homotopy does not change the turning number. (Actually, is this proof really different from the one on pages 17–18?)

► Express the turning number of an immersion $f : X \to \mathbb{R}^3$, where X is a closed surface, as the degree of a certain map $\alpha_f : X \to S^2$. Hint: look at unit normal vectors to $f(X)$. ◄

► The first two maps of the previous figure represent immersions of the circle: the identity map and the reflection in the x-axis. They differ by a change of orientation, and have opposite turning numbers (see top of this page). Analyze the situation for two immersions of the sphere that differ by an orientation: the identity map and the reflection in the xy-plane. Can you generalize to spheres of higher dimensions? ◄

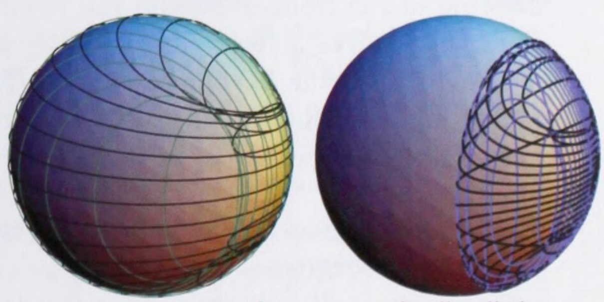
► I will mention one more guise of the degree: the linking number of two loops (page 17). To define this number, we consider two moving points, one on each loop, and take the vector going from the first to the second, like this:

This vector depends on two parameters, the positions of the two points along the corresponding loops. In symbols, if the loops are given by maps $f_1 : S^1 \to \mathbb{R}^3$ and $f_2 : S^1 \to \mathbb{R}^3$, we get a map from the Cartesian product $S^1 \times S^1$ (which is the torus) into $\mathbb{R}^3$ minus the origin, taking a pair (u, v) to $f_1(u) - f_2(v)$. The image of this map for the green loops above looks like this:



Where is the origin in relation to this Fabergé egg? Explain how the egg can be obtained as a collection of rigid images of either original loop. Draw some cross sections through the center.

As in the case of the turning number, what interests us is the direction of this vector, so we replace $f_1(u) - f_2(v)$ by a unit vector in the same direction, or, equivalently, project the egg onto a sphere centered at the origin (much as we projected the spiral onto a circle in defining α_f, on page 18). This gives a map $g : S^1 \times S^1 \to S^2$ from the torus into the unit sphere. The linking number is the degree of this map!



Here is a representation of the map g (by means of lines of constant u and lines of constant v) for the two pairs of loops of page 17. Which is which? Can you read off from these pictures the degree of each g?

True or false: The degree of g is the number you get if you travel along a ray from the origin to a point far away, and tally all the times that you cross the surface defined by $f_1(u) - f_2(v)$, counting up or down depending on which side of the surface you see as you approach. ◄

Y: *OK. I'm willing to believe that turning numbers don't prevent the sphere from turning inside out, as they do the circle. But that doesn't mean you can actually do it.*

X: We'll get to that. I know it's hard to see. Steve Smale proved it was possible in theory in 1957, but it took seven years before Arnold Shapiro found a practical way to do it. Since the problem remained hard to visualize, more methods were invented later, by Bernard Morin and several others. I'll show you Bill Thurston's method, invented in 1974.

This should have been "several", not "seven".

A brief history of sphere eversions

This section owes greatly to George Francis, whose writings, illustrations, and critical reading of an early draft were very helpful. (See Chapter 6 of the richly illustrated book [Francis 1987] and its extensive bibliography.) Tony Phillips also made significant contributions. Any errors are, of course, my own.

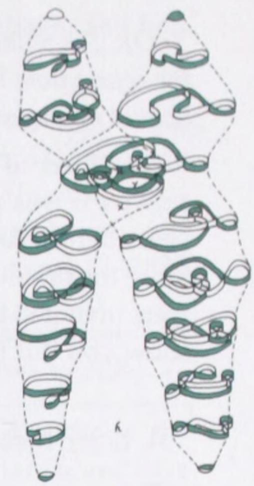
The history of sphere eversions starts in 1957, when Stephen Smale proved a very general fact about immersions of spheres ▸(see bottom of next page).◂ One consequence of his proof is that there should be a way to turn the sphere inside out by a regular homotopy.

For a little while, this claim met with skepticism. The mathematican Raoul Bott, who had been Smale's graduate adviser and who is one of the founders of differential topology, flatly told Smale that he was wrong, and explained why he thought so. Later he became persuaded that Smale's reasoning was correct, but he, like many other mathematicians, was still frustrated by the inscrutability of Smale's proof, and wished to see a more direct sphere eversion.

"In principle it is possible to piece together the myriad minute geometric constructions prescribed by [Smale's] proof to assemble an explicit visualization of an eversion. This strategy is far from practical" [Francis 1987]. It is akin to describing what happens to the ingredients of a soufflé in minute detail, down to the molecular chemistry, and expecting someone who has never seen a soufflé to follow this "recipe" in preparing the dish.

In 1961, Arnold Shapiro invented the first explicit eversion, but did not publish or divulge it widely. He did explain it to the French mathematician Bernard Morin, who passed it on to his compatriot René Thom, and eventually this eversion became more widely known thanks to Morin and George Francis, and especially to the article [Francis and Morin 1979]. Morin, incidentally, is blind, and the fact that he was one of the first people to understand how a sphere can turn inside out is both a tribute to his ability and a convincing proof that "visualization" goes far beyond the physical sense of sight.

The first time that most mathematicians and the public at large became aware of an explicit eversion was when Tony Phillips, aroused by an exchange of letters with Thom, worked out the details of what he thought was Shapiro's eversion (though later it turned out to be a different one). Phillips published a beautifully



written article in *Scientific American* [Phillips 1966], aimed at a wide audience and culminating with a series of pictures representing various stages of the evolution. Here is one of his original drawings:

The publication of these pictures dispelled the mysterious and paradoxical reputation of the problem; but filling in the missing levels in Phillips' pictures and tracking them through their implied deformations was not an easy task. In fact, a complete mental picture of any eversion is very hard to keep in mind, and clear printed images are even harder to draw, because of the many layers of surface: the most successful expositions of the everting sphere concentrate on local details, and rely on the viewer's powers of synthesis to piece the details together. People began searching for simpler and more symmetrical solutions.

Morin, in particular, devised in 1967 a new eversion that was simpler than all the preceding ones in terms of the number of crossings. Charles Pugh made wire-mesh models of various stages of it, and Nelson Max digitized these models and used them as the basis for the movie [Max 1977], an early triumph of computer

animation. The rendering of the evolving sphere was done by Jim Blinn; a frame is shown at the left. (Shortly after their digitization Pugh's models were stolen from the Berkeley campus, where they had been on display; they have never been recovered, but their digital shadows live on in Max's movie.) Morin himself described his eversion method in a series of short notes and in a popular-audience article, all skillfully illustrated by Jean-Pierre Petit [Morin and Petit 1978a, 1978b, 1978c, 1980]. One of these notes gave for the first time algebraic equations that might be used for a full-fledged computer rendition of the process.

Morin also found the smallest number and types of "events" unavoidable in any eversion. Together with François Apéry, he created an algebraic-analytic eversion using only the minimal number of events [Apéry 1992]. He also performed a similar feat in the polyhedral realm (where the surfaces of interest are flat-sided rather than smooth), turning a "sphere" with twenty triangular sides inside out in the most economical way possible [Morin 1995]. Models for this eversion, made by Richard Denner, are on exhibition at the Palais de la Découverte, in Paris.

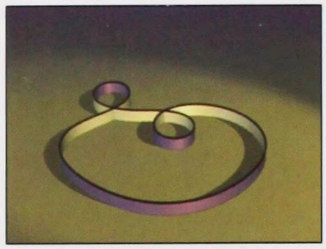
In the mid-seventies, Bill Thurston developed his idea of corrugations. This gives another route for the eversion, as explained in *Outside In*, and it can also be applied to other more general situations (see page 40).

► (Very technical.) Smale's original result [Smale 1958] says that if N is a C^2 manifold of dimension at least three, and $F_2(N)$ is the bundle of two-frames of N, the regular homotopy classes of C^2 immersions of S^2 in N based at $y_0 \in F_2(N)$ are in one-to-one correspondence with elements of $\pi_2(F_2(N), y_0)$. Our particular case is $N = \mathbb{R}^3$. Since $F_2(N)$ is homeomorphic to SO(3), and since $\pi_2(\mathrm{SO}(3)) = 0$, this implies that all immersions of S^2 in $\mathbb{R}^3$ are regularly homotopic. ◄

X: Let's go back to curves for a bit. Remember that this circle can only be changed into curves of turning number 1?

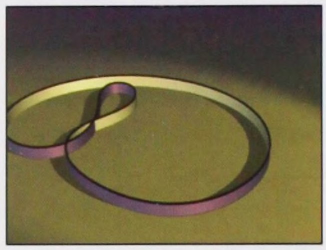
Y: *Still not allowing sharp corners, right?*

X: Of course. Now can the circle be turned into any curve of turning number 1, say this one?



Y: *Let's see: I'll try to go backwards from this curve to the circle... I think I got it. There.*

X: Excellent. Now try this one.



Y: *I'll undo this loop first... and push this fold back... Now here... Here we go.*

X: Very good! And this one?

Y: *Whoa! You're not going to ask me to do every single curve of turning number one, are you?*

X: Of course not. What we need is a general method. Do you remember the simple way to transform one curve to another when sharp bends are allowed?

Y: *Yes. You just go straight from one to the other.*

The Whitney–Graustein theorem

The next section details Thurston's argument for showing that, if two immersed closed plane curves have the same turning number, they can be transformed into one another by means of a regular homotopy. This theorem, credited by Hassler Whitney to William Caspar Graustein [Whitney 1937], has had several proofs, and [Francis 1969] discusses in some detail the many interesting ideas that go into them. You may also enjoy the movie [Max 1973], which illustrates nicely a more "classical" proof— ▸essentially the one given in many textbooks, such as [Berger and Gostiaux 1988]. (If you look up this book, be aware that it erroneously leaves out the part about the homotopy being regular; but the proof in 9.4.13 is right).◂

Note that the converse statement—if two immersed closed curves in the plane are regularly homotopic, they have the same turning number—is what we proved before, on page 16. Together the two results mean that *the turning number classifies immersed closed curves in the plane up to regular homotopy.*

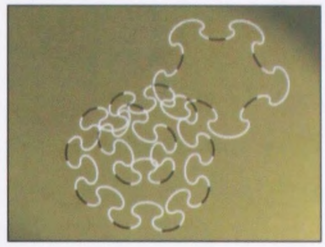
Suppose our immersed closed curves are f and g. We want to find a regular homotopy between them: that is, intermediate immersions h_s, for all s between 0 and 1, with the property that the position $h_s(t)$ and the direction of movement along h_s vary continuously with s. We already know the straight-line homotopy (page 10), where each point of an intermediate curve h_s lies a certain fraction of the way between corresponding points of f and g. This satisfies all the conditions except that h_s may not always be an immersion. (In fact, as we know, if f and g have different turning numbers, there *will* be some h_s's that are not immersions.)

X: That's the one. When the curves have the same turning number, this method can be adapted to work without sharp bends. The trick is to add waves to the curve.

Y: *Can we do it on a simpler one?*

X: Sure. We start by marking small pieces of the curve that will serve as guides for the transformation. We'll concentrate on these segments now. We move the centers of the guide segments straight toward their final destinations on the circle, without any rotation. Next, we rotate the guides so that they are lined up with the circle.

These guides represent a *direction map*.



Y: *OK, what about the parts in between?*

X: That's where the waviness comes in. We make the connecting segments between adjacent guides bulge out into corrugations. This allows the segments to move freely around each other, as long as they remain more or less parallel.

Y: *Oh, I see—the guides can move around without creating sharp bends.*

X: Correct. Here is the transformation of the whole curve.

Y: *The original curve, in blue, develops sharp corners, but the wavy curve is springy enough to remain smooth throughout.*

The curve in blue is called the *base curve*.

Proof of Whitney–Graustein (don't be fazed)

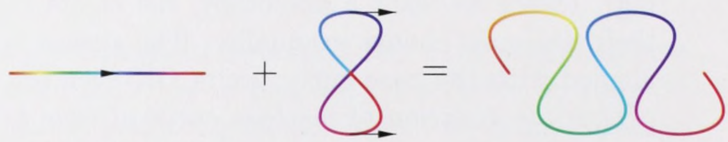
Now suppose that f and g have the same turning number. We will construct a more sophisticated homotopy from f to g than the straight-line homotopy, in a way that guarantees that the intermediate stages are always immersions.

This new homotopy will consist of four phases: you can also think of them of four mini-homotopies in a row. Obviously, each phase should be a regular homotopy in its own right. In phase I, we make waves in f, so it comes to resemble the inside layer of corrugated cardboard. These corrugations come from "adding 8s" to the curve in a sense that I will specify presently, so the corrugated curve at the end of this phase will be called f_8. In phases II and III we go from f_8 to g_8, which is to g as f_8 is to f. Finally, by shrinking the eights we get to g; this is phase IV. The whole process is summarized in the following table, whose various entries will become clear as we go along:

phase of homotopy intermediate curve	f	$\xrightarrow{\text{I}}$	f_8	$\xrightarrow{\text{II}}$	?	$\xrightarrow{\text{III}}$	g_8	$\xrightarrow{\text{IV}}$	g
base curve	f		f	average	g		g		g
direction map (guides)	(α_f)		α_f		α_f	average	α_g		(α_g)
size of 8s	0	grow	max		max		max	shrink	0

Each intermediate curve is the sum of two curves: a "base curve" and an "eights curve". The base curve is either f or g or an average of them, and in the movie it is shown faintly in blue. The eights curve is a figure 8 drawn over and over again. By sum I mean that we add the x-coordinates of corresponding points on the two curves, then add the y-coordinates, and the result is a point on the combined curve. One can also think in terms of relative motion. Imagine you are walking along the base curve, carrying a racing track in the shape of an 8 with a toy car darting around at great speed. The absolute trajectory of the car—what one would see if you and the track became invisible, and only the car remained—is exactly the sum of the base curve (your absolute motion) and the eights curve (motion of the car relative to you).

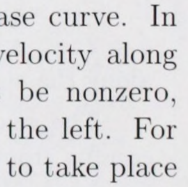
When we place the eights so their direction at the bulges is the same as the direction of the base curve, we get a corrugated curve, as shown at the right. Other relative directions lead to different appearances for the combined curve:



But note that all these combined curves are immersions. That's exactly why we are using the eights: if they are large enough, the velocity associated with them is also large, and can never cancel out the velocity of the base curve. In other words, the overall velocity, which is the vector sum of the velocity along the base curve and the velocity of the eights curve, will always be nonzero, as shown on the left. For this masking to take place it is enough that the smallest possible speed of the eights be greater than the largest possible speed of the base curve.

In each of the figures above, I pretended that all the eights pointed the same way. But what we really want to do is use eights that point in different directions, and control the homotopy by choosing the directions carefully. That is the point of the guide segments of the movie: they encode the directions (at the bulges) of the eights that we are adding to the base curve. More exactly, for each stage of the homotopy there is a *direction map* α that associates to a parameter value t an angle $\alpha(t)$. This angle says how you should be holding the toy track as you pass milepost t along your way.

Once we take into account the direction map—the directions of the guide strips, in the movie—the eights curve doesn't necessarily look like a figure 8 drawn over and over again. Because the angle $\alpha(t)$ usually depends on t, the eights curve looks more like this:



A calculation ▸ detailed below ◂ shows that, *when the eights are large, it doesn't matter what the direction map α is, so long as $\alpha(t)$ doesn't change too fast with t* (that is, so long as we keep adjacent guides roughly parallel, as Xanthippe says).

Phases II and III of the homotopy we are constructing take advantage of this fact. In phase II, we move the base curve from its initial position f to its final position g by a straight-line homotopy. The eights curve does not change. In phase III, we turn the guides from their initial to their final directions, using some care (page 37). The base curve does not change. Throughout both these phases, we don't have to worry about whether the wavy curve is immersed: it automatically is, provided we made the corrugations big enough in phase I.

But what happens in phases I and IV, when the eights are growing and shrinking? (Since we want a homotopy, the eights can't suddenly appear full-grown: their size must change gradually.) The answer is that, when the guide strips are aligned with the base curve, we're OK even for small eights! More formally: *if $\alpha(t)$ is the direction of the base curve at time t, the combined curve is again an immersion, whatever the size of the eights.* Note that this way of choosing the direction map α is exactly the construction of α_f at the bottom of page 18.

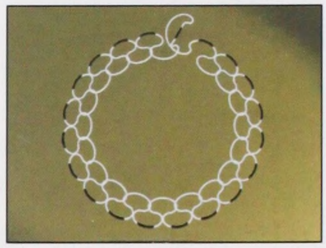
▸ Here now are the promised calculations that prove the italicized statements above. Suppose we make our eights twice as tall as their are broad; other ratios would work too, with the estimates coming out slightly different. A single eight, of half-width w, is described by $(-w\sin 2\theta,\ 2w\sin\theta)$, where θ ranges from 0 to 2π as we go around once. If we're adding n eights to the base curve, the phase angle θ along the eights curve is $\theta = 2\pi nt$, where t is the parameter along the base curve. So the instantaneous position along the eights curve is $(-w\sin 2\theta,\ 2w\sin\theta)$ rotated α.

The velocity along the eights curve can be written as the sum of two vectors, one coming from the variation in α and the other from the variation in θ. (Product rule for the derivative!) The length of the first vector is at most $2w\,|\alpha'|$. The length of the second is $4\pi nw\sqrt{\cos^2 2\theta + \cos^2\theta}$. This square root is always greater than $\frac{1}{2}$; therefore, if we choose n such that $\pi n > 2\,|\alpha'|$, we ensure that the first vector is at most half as long as the second. Their sum, the velocity along the eights curve, will have length greater than πnw (the other half of the length of the second vector). So provided the maximum velocity along the base curve does not exceed πnw, the composite curve is an immersion. This shows things work when w is large and α doesn't change too fast. (Note that since the base curve is a weighted average, with positive weights, of f and g, we can bound its velocity by $\max(\|f'\|, \|g'\|)$).

To show they work when w is arbitrary and α is given by the direction of the base curve, assume without loss of generality that $\alpha = 0$ at the moment of interest, so the velocity along the base curve points along the positive x-axis. Then the velocity along the eights curve equals $4\pi nw(-\cos 2\theta,\ \cos\theta)$ plus a component due to the variation in α, of length at most $2w\,|\alpha'|$. Using the condition $\pi n > 2\,|\alpha'|$, one can check as before that the velocity along the eights curve avoids the negative x-axis, and so cannot cancel the velocity along the base curve. Once more, the composite curve is an immersion.

To summarize, we compute from f and g the maximum values of $\|f'\|$, $\|g'\|$, and $|\alpha'|$; then we choose w and n in such a way that $2|\alpha| < \pi nw$ and $\max(\|f'\|, \|g'\|) < \pi nw$. ◂

X: We have to keep adjacent guides roughly parallel as we rotate them to align with the circle. This is possible as long as the turning number of the original curve is one.

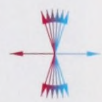


Y: *Why can't we align the guides if the turning number isn't one?*

X: Watch what happens when we try to turn a figure eight into a circle. ... And here both the initial and final curve have turning number zero. Using this method, or others, you can always transform one curve into another with the same turning number. This is called the Whitney–Graustein theorem.

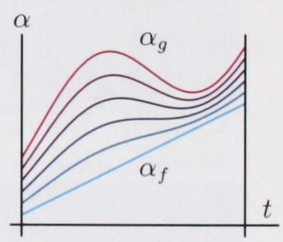
The domino effect

To finish the proof of the Whitney–Graustein theorem, I still have to explain how to change the direction map from α_f to α_g in phase III. The table on page 34 uses the word "average", but that is just a mnemonic. In fact we cannot average angles as if they were numbers, since angles are only defined up to multiples of 360°. These two directions might be thought of as being 30° and −30°, in which case their "average" would be 0°, like this: But we could think of the same directions as 30° and 330°, in which case their "average" would be 180°: We could agree to always place the average "between" the two original directions, but this average can jump when the directions vary continuously:



It is here that we need the assumption that f and g have the same turning number; when this happens, we can choose determinations consistently for all angles, and average them. ▸Another way to put this is: if f and g have the same turning number, α_f and α_g are homotopic as maps from the circle to itself.◂

Let f and g be parametrized by time t, ranging from 0 to 1, and let's think of the graphs of α_f and α_g versus t (see figure). The turning number of f is the difference $\alpha_f(1) - \alpha_f(0)$, in units of full turns. If this is also the turning number of g, we can slide α_f to α_g, keeping the difference between the values at $t = 1$ and $t = 0$ fixed; then each intermediate graph defines a direction map. If the turning numbers are different, we cannot do the sliding in such a way that the difference between the endpoints is always an integer number of full turns.



We can most easily do the sliding by taking the straight-line homotopy between α_g and α_f. This turns out to be visually confusing when applied to the base curve, because all the eights turn at the same time. For clarity, then, we used a different homotopy in *Outside In*, one that changes only a small piece of the direction map at a time. This gives the "domino effect" of the movie. ▸This homotopy is given by the weighted average $\alpha_f + k_s(\alpha_g - \alpha_f)$, where k_s has approximately these graphs for increasing values of s:


◂

Y: *And what does this have to do with the sphere?*

X: A lot. Think of the sphere as a stack of circles, deformed into a barrel shape and closed off by caps above and below. Just as we made our curves more pliable by dividing them into guide segments connected by waves, we divide the barrel into guide strips that alternate with wavy strips. The waviness dies out at the top and bottom, so as to match the caps.

This page and the next don't have explanations, because the material is covered in detail in the movie. For an informal exposition of the general technique of corrugations, turn to page 40.

Y: *Hmm, this is going to get complicated...*

X: Then, for now, let's look at a single guide strip,along with the caps. Start by pushing the two caps past each other.

Y: *Before, when I pushed the poles through, it made a crease!*

X: Stop before the crease, when the guide has a loop in the middle. Now we turn the two caps in opposite directions, because we want to convert the loop in the middle to twisting at the ends.

Y: *Oh I know—it's like a belt! If you put a loop in the middle and pull the ends tight, the loop turns into twisting!*

Try the belt trick yourself!

X: Right. Then you can straighten out the belt by turning each end half a turn in opposite directions. To finish the eversion, we just need to push the middle of the guide strip back through the center of the sphere.

Y: *Hmm. Can I see how two guide strips interact?*

X: Sure. You can see that there are two places where the strips intersect near the central axis.

Y: *And the gold sides that started facing out are now facing in.*

X: Here is the whole process with all the guides.

Y: *The polar caps just move up and down and then rotate into place. Ah, that's why they don't require any springiness.*

X: Exactly. Now let's look at two guides and the corrugation between them, from a pole to the equator. This chunk is the fundamental building block of the eversion: the whole sphere is made from sixteen rotated copies of this piece.

The number of copies is not very important, but the fewer the copies, the bigger the corrugations have to be.

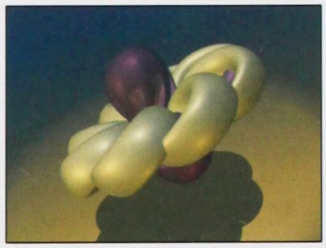
Y: *That looks pretty complicated.*

X: Yes, but the corrugation is just following the twisting of the guide strips that you saw before.

Y: *Can I see that from pole to pole?*

X: Yes. The corrugation provides flexibility between the guides so that their motion does not create any pinches or creases, just like the waves in the curve that we saw before.

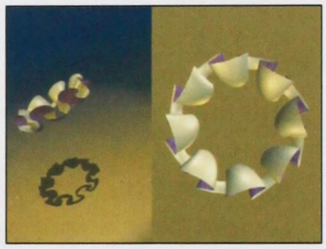
Y: *Let me see the whole thing!*



X: We corrugate the connecting strips between the guides, and push the caps past each other. We twist the caps to undo the middle loops, and push the equator across the sphere. Finally, we uncorrugate.

Y: *I still don't understand. Is there some other way to look at this?*

X: OK, we'll divide the sphere into thin horizontal ribbons. We'll look at one ribbon at a time. You can see the north pole push down into the south. A ribbon near the pole is rather tame: the guide segments keep their position relative to one another, and the corrugations never get very deep. Ribbons closer to the equator are wilder, so we'll split the screen to see what's going on. On the right, the camera tracks the ribbon from above, so its apparent size does not change. This overhead view highlights the symmetries that are hidden in the side view on the left, where we see the position of the ribbon in space. At the equator, the ribbon just twists and doesn't move up or down.



Y: *Wait a minute. This ribbon looks just like the wall under the monorail, and it's turning inside out. You'd finally convinced me that that was impossible!*

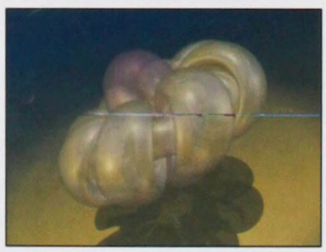
X: I'll play that again. Remember that our walls represented circles and had to stay vertical. But here the ribbon can twist around in space, because it's part of a sphere.

Another way to understand the eversion is to progressively build up the surface of the sphere at a few important stages. This is the corrugation phase... Now we've just pushed the caps through each other... This is the middle of the twisting phase: we can see the complex activity at the equator... At the end of the twisting phase, the corrugations have nearly become figure eights... Here we're in the middle of pushing horizontally through the center of the sphere... Finally, we show the uncorrugation phase. The sphere is now entirely purple.

Y: *Wow. I think I'm ready to see the whole thing again.*

X: Here goes!

Y: *You were right: you can turn a sphere inside out without poking holes or creasing it, even though you can't do it for a circle. This is great—somebody should make a movie about this stuff!*



(Groan...)

Making Waves: The Theory of Corrugations

by Bill Thurston

Corrugations are a technique that can be used to create and control immersions in a direct and flexible way. With this technique, materials like curves and surfaces are made springy, and can be moved about and bent at will.

The method of corrugations was something that I worked out for my own satisfaction. Although I had read various mathematical discussions of immersions (including Smale's original paper), and had used some of the further related developments of Phillips, Gromov, Haefliger and others in some of my work on foliations, and even though I had studied Charlie Pugh's wire models of the sphere turning inside out, I felt a lack in my understanding. The logical steps in Smale's analysis of immersions of spheres were clear and constructive; nonetheless, after reading this proof, I did not have a coherent mental movie. Pugh's models were beautiful, well-made, and each transition was clear. But I learned from them that my three-dimensional visual perception was highly conditioned by what is familiar: contrary to my prior intuition, it was hard for me to apprehend at a glance the self-intersecting surfaces that were so wonderfully represented. It was also hard for me, even after carefully studying them and following the sequence step by step, to assemble them into a coherent story or a mental movie. Even with all this external assistance, I was not satisfied with my understanding, for it did not seem sufficiently alive and direct.

When I discovered corrugations, I was elated. They gave me a clear, compelling and coherent method to see, prove and understand things that had previously stumped my insight. I explained it to other mathematicians and gave talks about it, but I was disappointed that a lot was lost in the translation—it seemed to come across as complicated and obscure, rather than simple and compelling. There isn't much mileage in obscure-seeming proofs of previously proved theorems, so I put the topic on the back burner.

Communication, and failures in communication, are very important to mathematics, far more important than I realized when I started my career. Each person's mind is a world to itself; the channels of communication between these worlds are very limited and very selective, compared to what goes on inside. My insight was an insight to me. It would not necessarily have been much of an advance if I had been able to more clearly take in other mathematicians' ways of thinking, and, conversely, it did not immediately translate into insight for other mathematicians.

Outside In was an effort involving a whole team of talented people at the Geometry Center, by using a different, emerging medium, to communicate a dimension of insight different from that typically conveyed by a mathematical paper.

Turning the sphere inside out is one example and application of a very general and powerful technique. To understand how this works in a setting more general than *Outside In*, you must create a work in a medium that is much more versatile than computer animation, although harder to transmit: your mind. This is the subject of the next section.

Corrugations in two and more dimensions

by Bill Thurston

▸ In this section, which necessarily requires some more background than the rest of *Making Waves*, I will outline how corrugations can be used to prove the Hirsch–Smale immersion theorem, which goes like this: If M and N are smooth manifolds, and if $\dim M < \dim N$ or if M has no compact component, the space of immersions $M \to N$ is homotopy equivalent to the space of bundle maps $TM \to TN$ that are injective on each fiber. Smale had done the case that M is a sphere; Hirsch proved the result for arbitrary M (though not quite in the generality stated here) soon thereafter [Hirsch 1959].

The use of corrugations to control immersions of a curve in the plane was dealt with in some detail in *Outside In*, and is also discussed on pages 33–36. Bring to mind the image of a toy track shaped like an 8, moving along a curve. A small car races around on the track, moving faster around the track than the track itself moves, thereby guaranteeing that the curve it traces is an immersion. By adjusting the size and orientation of the 8-shaped track, one can interpolate between any two curves of the same turning number.

Using corrugations and some basic topological techniques, one can in fact see that the space of immersions of the circle in the plane is homotopy equivalent to the space of maps of a circle (the parameter circle) to a circle (the circle of directions). This, in turn, is homotopy equivalent to a countable union of circles, one for each turning number. (If you start with a standard round circle, you can go once around the homotopy circle of immersions by rotating the circle once around itself in place. However, for a circle with turning number two, this same procedure of moving the map once around its image goes twice around the homotopy circle of immersions, while for a circle with turning number zero, the same procedure is homotopic to doing nothing. To go once around the homotopy circle of immersions starting from an arbitrary given immersion, you can first corrugate the immersion, then turn the orientation of the 8's around 360°, then uncorrugate.)

In general, the data needed to control immersions are:

- the domain: a differentiable manifold M of dimension m;
- the range: a differentiable manifold N of dimension $n \geq m$;
- a *position map*: a continuous map $f : M \to N$; and
- a *direction map*: a lifting of f to a bundle map $F : T(M) \to T(N)$ that is a linear, injective map when restricted to any fiber of $T(M)$.

Whenever one has an immersion $f : M \to N$, there is an associated bundle map, where f is the position map and its derivative $F = T(f)$ is the direction map. In this situation, the pair (f, F) is called *integrable*, since f is the integral of F. If you start with arbitrary data, this situation is very atypical. However, the immersion theorem stated above asserts that, up to homotopy, any set of data is integrable. In fact, suppose we are given manifolds M and N, where for simplicity we assume M is compact. Then the set of immersions with the C^1 compact-open topology form a topological space $\mathcal{I}(M, N)$, and the set of all possible choices of data with the compact-open topology forms another topological space $\mathcal{D}(M,N)$. The immersion theorem asserts that the derivative $T : \mathcal{I}(M,N) \to \mathcal{D}(M,N)$ is a homotopy equivalence.

The figure 8 is a particular immersion of the circle in the plane having the property that its tangent vector has net turning number zero. Let's move up in dimension to the case of immersions of surfaces in space. We'll substitute for the 8 a particular immersion of the torus in space, having the analogous property that its tangent plane field is homotopic to the horizontal plane field along the immersion. We can construct such an immersion by taking a figure 8 and sweeping it through space through a larger figure 8 path. The plane of the smaller 8 is held perpendicular to the larger 8, while the plane of the larger 8 is perpendicular to the x-axis for the smaller 8. We will call this immersion the figure 8^2.

Its most straightforward application is to the construction and analysis of immersions of the torus T^2 in space. Suppose we are given a position map $f : T^2 \to \mathbb{R}^3$ and a direction map $F : T(T^2) \to T(\mathbb{R}^3)$, both differentiable. For each point $x \in T^2$, the restriction of F to the fiber at x is an affine map of $\mathbb{R}^2 = T_x(T^2)$ to R^3; extend this to construct an affine map $A_x : \mathbb{R}^3 \to \mathbb{R}^3$ by mapping the third coordinate in the direction perpendicular to the first two. Make the size of the derivative in the third coordinate the average of the size of the derivatives in the first two coordinates.

By composing the model figure 8^2 with the affine maps A_x, we now obtain what could be called a figure 8^2 field along $f(T^2)$: for each point $x \in T^2$, a copy of the figure 8^2, based at $f(x)$, and with its orientation in space determined by F. In other words, we have a map $B : T^2 \times T^2 \to \mathbb{R}^3$, where the second factors are mapped as figure 8^2-shaped immersions.

Now, just as we did for curves in the plane, we can construct an immersion of the torus that winds very quickly around the figure 8^2's, while slowly traversing the positions specified by f. We can think of this mapping $T^2 \to T^2 \times T^2$ as the graph of the n^2-fold covering map that stretches by a factor of n in each coordinate. The tangent planes for the image are nearly "vertical", that is, nearly tangent to the second factors, provided n is large. Therefore, the composition with B is an immersion, provided n is sufficiently large; how large depends on the derivative of f and the value and derivative of F.

The immersions resulting from this construction are doubly corrugated: they have large corrugations running in one direction, with smaller corrugations that run in a transverse direction superimposed.

If we start with an immersion and perform this construction using its derivative as direction map, the original immersion is regularly homotopic to the doubly corrugated form. A good way to see this is to start with a plane, since any surface is approximately planar in a close-up view. First a single set of corrugations grow in one direction; really, this is just the one-dimensional case cross a line. When these corrugations are full-grown, a second set of smaller corrugations grow, transverse to the first.

With these ingredients, it is now easy to show that regular homotopy classes of immersions of the torus in space are in one-to-one correspondence with homotopy classes of direction maps. The doubly corrugated map construction shows that, for every homotopy class of direction maps, there is at least one regular homotopy class of immersions, so what remains is to prove that whenever two immersions a and b have homotopic direction maps, a is regularly homotopic to b. To do this, simply start from a, introduce double corrugations, move the double corrugations to follow the homotopy between the two direction maps, and then remove the double corrugations to obtain b.

It turns out that there are exactly four homotopy classes of directions maps, hence four regular homotopy classes of immersions of the torus in space. The different immersions can be distinguished by looking at small annular neighborhoods of various curves on the torus. As we saw in *Outside In*, when a curve with turning number one is thickened into three dimensions, it can be turned inside out, that is, it is regularly homotopic to a thickened curve of turning number -1. However, it is not regularly homotopic to a thickened curve of turning number 0. In fact, there are exactly two regular homotopic classes of immersions of an annulus in space: one contains the thickenings of immersed curves in the plane with even turning number, and the other contains the thickenings of immersed curves of odd turning number. Going back to the torus, if you consider a meridian circle and a longitude circle on the torus, the annular neighborhoods of each of these circles can independently be in either of the two regular homotopy classes, making four possibilities in all. (See the figure on page 28 for the (even, odd) case.)

The key property of the figure 8^2 immersion is that it represents the trivial homotopy class of direction maps: that's what makes it possible to regularly homotope an immersion to the modified form doubly corrugated via the moving figure 8^2. In fact, given that the immersion theorem is true, it follows that any other immersion X whose direction map has the trivial homotopy class would work in the proof. To use X to obtain a non-circular proof of the immersion theorem for toruses, we would just have to describe an explicit regular homotopy of an immersion to a form doubly corrugated via X, rather than via 8^2.

One more trick is needed to use double corrugations to construct and analyze immersions of a more general surface M in space. The idea is that that, given a direction map, we can first make a small homotopy so that it becomes integrable in the neighborhood of a finite number of points; in fact, we can make it come from a linear immersion in small neighborhoods of points. In the complement of these neighborhoods, we can consistently choose a pattern for double corrugation. One way to achieve this is to choose two closed one-forms α and β with isolated zeros, such that $\alpha \wedge \beta$ is nonzero wherever α or β is nonzero. (This can be easily done graphically, or analytically by choosing a Riemannian metric, letting α be any nonzero harmonic one-form and β its harmonic conjugate). Perturb α and β so that they have rational periods, and so that the integral of the forms along any arc joining zeros is also rational. Multiply them by a constant so that they have integral periods and so that also the integral along any arc joining zeros is an integer. Then integration of (α, β) defines a map of M^2 to the torus $\mathbb{R}^2/\mathbb{Z}^2$ taking every zero of the forms to $(0,0)$.

Given a direction map that is integrable near the zeros of the one-forms, we can now use these data to define a figure 8^2 immersion of the torus for each point on M^2 except at the zeros of the one-forms. Note that the size of the toruses goes to zero at these exceptional points. Just as in the case of the torus, we can now construct an immersion of M^2 that runs quickly around the figure 8^2's while traversing slowly through space as specified by the position map.

Further generalizations to obtain the full immersion theorem for immersions of M into N are not hard: they just involve additional layers of technique. First suppose that the dimension n of N is strictly greater than the dimension m of M. Higher-order m-fold corrugations are completely analogous to double corrugations, and are based on inductively creating a particular figure 8^m immersion of the m-torus in $(m+1)$-space by sweeping the figure 8^{m-1} through a figure 8. To apply this in the case that $n > m + 1$, given a direction map, one must choose an $(m+1)$-st dimension in N. This, along with problems stemming from the topology of M, can be handled in the spirit of partitions of unity, by starting with local coordinate systems in M cross the Grassmannian of m-planes in N, where the choices can readily be made. The transitions between different local coordinate systems can be constructed by arranging that whenever two coordinate systems intersect, one has much higher-frequency corrugations than the other. Start by defining the lower-frequency multiply-corrugated immersions, then modify in the overlap by letting high-frequency corrugations build up over the transition zone where the two coordinate charts overlap.

The case that M has the same dimension as N but has no compact component is handled by finding a lower dimensional spine for M, that is, a cell complex K of lower dimension inside M such that M is diffeomorphic to a regular neighborhood of K. It is possible to make sense of the notion of immersions of K in N. These can be constructed and controlled by the same method of multiple corrugations, using induction on the dimension of cells of K. With the right definition, an immersion of K immediately extends to an immersion of a regular neighborhood, thus giving an immersion of M.

For the purposes of *Outside In*, only single corrugations were needed. If you have a direction map which is half-integrable—integrable in one set of directions—you only need one kind of corrugations. For the particular case of turning the sphere inside out, it is fairly easy to devise a homotopy of the direction map from the identity to the direction map for the antipodal map that always remains integrable in the direction of the longitude curves of the sphere. This semi-integrable homotopy of direction maps is what is visually depicted by the guide strips in *Outside In*.◂

Bibliography for *Making Waves*

[Abbott 1884] Edwin Abbott, *Flatland: a romance of many dimensions*, Seeley, London, 1884. Reprinted by Dover, New York, 1992.

[Apéry 1992] François Apéry, "An algebraic halway model for the eversion of the sphere", *Tohoku Math. J.* **44** (1992), 103–150.

[Berger and Gostiaux 1988] Marcel Berger and Bernard Gostiaux, *Differential Geometry: Manifolds, Curves and Surfaces*, Springer, New York, 1988.

[Francis 1969] George K. Francis, "The folded ribbon theorem", *Trans. Amer. Math. Soc.* **141** (1969), 271–303.

[Francis 1987] George K. Francis, *A Topological Picturebook*, Springer, New York, 1987.

[Francis and Morin 1979] George K. Francis and Bernard Morin, "Arnold Shapiro's eversion of the sphere", *Math. Intelligencer* **2** (1979), 200–203.

[Gould 1987] Stephen Jay Gould, *Time's Arrow, Time's Cycle: Myth and Metaphor in the Discovery of Geological Time*, Harvard University Press, Cambridge, MA, 1987.

[Gray and Glynn 1991] Theodore W. Gray and Jerry Glynn, *Exploring Mathematics with Mathematica*, Addison-Wesley, Reading, MA, 1991.

[Guillemin and Pollack 1974] Victor Guillemin and Alan Pollack, *Differential Topology*, Prentice-Hall, Englewook Cliffs, NJ, 1974.

[Gunn and Maxwell 1991] Charlie Gunn and Delle Maxwell, *Not Knot* (video), A K Peters, Wellesley, MA, 1991. Produced by the Geometry Center, University of Minnesota.

[Hirsch 1959] Morris W. Hirsch, "Immersions of manifolds", *Trans. Amer. Math. Soc.* **93** (1959), 242–276.

[Jordan 1893] Camille Jordan, *Cours d'analyse de l'École polytechnique* (second edition), Gauthier-Villars, Paris, 1893.

[Maeder 1991] Roman E. Maeder, *Programming in Mathematica* (second edition), Addison-Wesley, Reading, MA, 1991.

[Max 1973] Nelson Max, *Regular Homotopies in the Plane, parts 1 and 2* (video), International Film Bureau, Chicago, 1973.

[Max 1977] Nelson Max, *Turning a Sphere Inside Out* (video), International Film Bureau, Chicago, 1977.

[Milnor 1963] John Milnor, *Morse Theory*, Princeton University Press, Princeton, NJ, 1963.

[Morin 1995] Bernard Morin, "The eversion of the cuboctahedron", to appear in *Hokkaido Math. J.*

[Morin and Petit 1978a] Bernard Morin and Jean-Pierre Petit, "Équations du retournement de la sphère", *Comptes Rendus Acad. Sci. Paris* **287** (1978), 879–882.

[Morin and Petit 1978b] Bernard Morin and Jean-Pierre Petit, "Le retournement de la sphère", *Comptes Rendus Acad. Sci. Paris* **287** (1978), 791–794.

[Morin and Petit 1978c] Bernard Morin and Jean-Pierre Petit, "Problématique du retournement de la sphère", *Comptes Rendus Acad. Sci. Paris* **287** (1978), 767–770.

[Morin and Petit 1980] Bernard Morin and Jean-Pierre Petit, "Le retournement de la sphère", pages 32–45 in *Les Progrès des Mathématiques*, Pour la Science/Belin, Paris.

[Morse 1966] Marston Morse, *Pits, Peaks and Passes* (video), Math. Assoc. of America, 1966.

[Munkres 1975] James Munkres, *Topology: A First Course*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[Phillips 1966] Anthony Phillips, "Turning a surface inside out", *Scientific American* **214** (May 1966), 112–120.

[Smale 1958] Stephen Smale, "A classification of immersions of the two-sphere", *Trans. Amer. Math. Soc.*, **90** (1958), 281–290.

[Wagon 1991] Stan Wagon, *Mathematica in Action*, W. H. Freeman, New York, 1991.

[Whitney 1937] Hassler Whitney, "On regular closed curves in the plane", *Compos. Math.* **4** (1937), 276–284.

[Wickham-Jones 1994] Tom Wickham-Jones, *Computer Graphics with Mathematica*, TELOS/Springer, Santa Clara, CA, 1994.

[Wolfram 1991] Stephen Wolfram, *Mathematica: A System for Doing Mathematics by Computer* (second edition), Addison-Wesley, Reading, MA, 1991.

Afterword: Why *Outside In?*

by Albert Marden
Founding Director, Geometry Center
Professor of Mathematics, University of Minnesota

Outside In follows *Not Knot* as part of a new genre of computer-generated videos introduced by the Geometry Center. Traditionally, mathematical communication has been narrowly targeted to one particular group: researchers, schoolchildren, undergraduates, and so on. In contrast, the Center videos are designed to bring subjects on the frontier of contemporary research to a full spectrum of viewers. *Not Knot* was an introduction to the sweeping new theory of hyperbolic structures that has revolutionized the mathematician's understanding of three-dimensional spaces. *Outside In* works as a gentle introduction to the ideas of differential topology, concentrating on an explanation of the surprising and influential discovery of the sphere eversion. To better understand the purpose of these animations, it is worth reflecting on the philosophy and process of their creation.

Computer graphics

Pat Hanrahan, computer graphics guru and Academy Award winner, once pointed out that his field is admirably well-suited to the communication of mathematics, both sciences being, albeit in different ways, abstract mediums for expression. Building mathematical models and simulating real-world phenomena such as turbulent fluid flow involves fascinating and challenging science. Yet it is in the realm of "pure geometry" that Hanrahan made his remark, and it is also there that computer use is most novel.

For over 2500 years, even before Plato's time, geometric discovery and understanding have inspired architecture, art, astronomy and engineering. The line of discovery continues to our own century, which began with the work of Poincaré and Einstein, and continues unbroken to our day. Powerful theories have been built that do not depend on our being able to physically observe the phenomena they describe.

As mathematicians we can often, by analogy and extrapolation from what we do experience, "see" abstract objects in our mind's eye: noneuclidean geometries, spaces of four and more dimensions, and so on. Computer graphics then allows us to share some of this vision, as if it too were part of our everyday world. We can play four-dimensional Tetris or see what it would be like to live in hyperbolic space, because these experiences, which are very real in the mathematician's imagination, can be modeled and displayed on a computer screen and shared with others. We can even look around inside four-dimensional space-time to see the theorized black holes and big bangs. Indeed, the distinction (if any!) between mathematical and physical reality is more blurred than ever.

A proof by video?

This question is not meant literally. Indeed, a proof must be written out in detail, so that the argument can be carefully checked for soundness by independent experts. It is not a rare occurrence that a mathematician, after relishing the pure joy of thinking he or she has made a wonderful discovery, has to make a quick retreat in abject misery after meticulous checking reveals a mistake in the reasoning.

Yet, how do we learn about mathematical discoveries, incorporate them into our thinking, or find the inspiration to make them in the first place? It is not just a matter of reading through pages of carefully reasoned technical jargon, although much hard work along these lines is certainly necessary to learn the tools of the trade. Specialization has to be complemented with discussions, lectures and readings on a less technical level, in order for one to see the "big picture". This is what allows us to apply to one problem the approaches and techniques developed for others, sometimes superficially quite different, and so reap the benefits of our collective work to the fullest.

Because our brains are built to process visual input quickly, animated images can be especially effective in imparting insight and understanding without the distraction of technical details. Seeing is often the first step to believing.

With *Outside In*, one of our hopes was that, after seeing the video, an expert differential topologist could in principle write out a formal proof. One reason for showing the everting sphere from so many perspectives is to convince the skeptical mathematician that it is free of creases or pinch points. And, as Thurston remarked, there is more in the video than the technique for this particular eversion: "Viewing between the frames, you can see a proof of the immersion theorem [of Smale and Hirsch]." In fact, in *Outside In* and *Making Waves* Thurston's technique of corrugations is publicly divulged for the first time. Although the technique is presented in the context of a particular application, Thurston believes it has wide applicability in differential topology.

Communication of mathematics

All this suggests how useful a video can be for disseminating the essential ideas of a theorem among mathematicians and math graduate students. But *Outside In* is aimed also at a much broader audience. For undergraduates, what better way is there to introduce central notions of mathematics such as the turning number, homotopies, and self-intersections, than with a video that shows them in action and brings to life the dry words of the text? Beyond that, *Outside In* was designed to be accessible to a broad segment of the general public: for example to those who rather liked mathematics in school, ten or twenty years ago, but have gotten their hands around little math of substance since then; or even to those who have associated math only with dreaded word problems on a test, and would like to see a different side of it. The central goal of *Outside In* was to help cast contemporary mathematics in its proper role as a part of scientific and cultural life. Not a modest dream!

Communication is a subtle process. We all have our ideas about what constitutes good exposition; but what we should not lose sight of is that each person in the audience has distinct life experiences and ways of receiving information, and every additional channel—visual, intuitive, even tactile—will only improve the chances that our message will be absorbed.

For its creators, the video has additional value. For them, the production complements the classroom experience as a source of training and sensitivity for effective communication. The making of *Outside In*, and also of *Not Knot*, was a seminal educational encounter for all of us involved in it.

Their sponsorship of these animations shows the prescience of the Geometry Computing Group (the team of mathematicians and computer scientists who founded the Geometry Center and its predecessor, and led them from 1985 to 1994), of the National Science Foundation, and of the University of Minnesota. These finely crafted videos will be part of the permanent mathematical literature. Rapidly improving technology, growing graphics expertise among mathematicians, and the positive experience with *Not Knot* and *Outside In* will assure continued expansion of the genre. A lot more people will be seeing a lot more mathematics.

Index

Contents

The award-winning computer animation *Outside In* explains the amazing discovery, made by Steve Smale in 1957, that a sphere can be turned inside out by means of smooth motions and self-intersections. It convincingly demonstrates how valuable visualization can be in the communication of mathematics.

With dialogue and exposition accessible to all who have some interest in mathematics, *Outside In* builds up to the grand finale—Bill Thurston's "corrugations" method of turning the sphere inside out—by discussing the related case of closed curves (which generally *cannot* be turned inside out) and by using everyday analogies such as train tracks, belts, smiles and frowns—all richly animated and complete with sound effects.

The present book, *Making Waves*, includes the movie's lively dialogue and expands upon the ideas introduced in the film from a mathematical and historical perspective. It doesn't assume a particular degree of mathematical sophistication, because it is masterfully organized on different levels so concepts can be assimilated at a rate based on the reader's individual background and interests.

For the more mathematically adventurous, *Making Waves* includes an article by Bill Thurston himself, on how the method of corrugations can be applied to similar problems in any dimension.

Running time of *Outside In*: 20 minutes

Also produced by the Geometry Center and distributed by A K Peters: *Not Knot*

Distributed by
A K Peters, Ltd
289 Linden Street
Wellesley, MA 02181

ISBN 1-56881-046-6 (video + paperback)
ISBN 1-56881-049-0 (paperback)