The Mathematics of Doodling

Ravi Vakil

Abstract. Wondering about a childhood doodle compels us to ask a series of questions, which will lead us on a tour of increasingly sophisticated ideas from many parts of mathematics.

Let me tell you about a doodle I did when I was very small, and the mathematics that flows inevitably from it. It looks like play, but in my mind this is what mathematics is really about: finding patterns in nature, explaining them, and extending them. Mathematics is about asking the right questions, and that's what we will do here, finding interesting questions, and then finding questions behind the questions. These ideas will lead to rather deep mathematics, in lots of fields. I am not an expert in these fields, and you needn't be an expert in a part of mathematics to let your curiosity pull you in. On a related note: in this article, I will ask a lot of questions, and give relatively few answers. And there are many questions that you will think *should* be asked, that I don't—I encourage you to follow up on them, because they may lead you somewhere interesting and unexpected. This article is intended for readers with widely different mathematical backgrounds, so if you come across a notion you have seen, or one you find trivial, please keep reading.

This article is based on the first of the Hedrick lectures at the 2009 Mathfest, in Portland, Oregon.

THE DOODLE. The doodle involves finding some shape on your piece of paper, and then drawing a curve tightly around it, as close as you can. After you've completed the loop, do it again. And again. And again. (See Figures 1 and 2.)

Figure 1. The doodle.

THE QUESTION. I noticed that if I kept on doing the doodle over and over again, no matter what shape I started from, the doodle got "more and more circular." So my first question is: is this true?

Original Question. If we repeat the doodle a lot, does the resulting shape get very circular?

doi:10.4169/amer.math.monthly.118.02.116

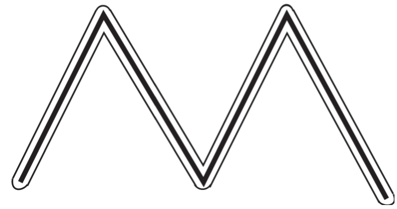

Figure 2. The doodle, once.

That's an imprecise question, but that's how science is done: we observe something qualitative, and then we try to make sense of it mathematically. So we have a prior question: what precisely (mathematically) are we doing?

Here is a reasonable definition of what we are doing.

Definition. Given a plane set X, and a constant $r \geq 0$, define the "radius r neighborhood of X" $N_r(X)$ to be those points within a distance r of some point in X. Formally, define:

$$N_r(X) = \{y : |y - x| \leq r \text{ for some } x \in X\}.$$

We can now reword our question:

Reworded Question. In some sense does

$$N_r(N_r(\cdots(N_r(N_r(X)))\cdots))$$

become more and more circular, as we iterate N_r more and more often?

Of course, we still have to make sense of the phrase "more and more circular." But as always in science, we will let the answer tell us what the question should have been.

There is also the matter of the number r. How relevant is it?

Prior Question. How does our choice of r affect the problem?

For example, which is bigger, $N_1(N_1(X))$ or $N_2(X)$? In fact, the two sets are exactly the same! And as you might expect, this isn't a special fact about $1 + 1 = 2$, so we will want the most general statement that has the same proof.

Theorem 1. *For any X, and any $a, b > 0$, $N_a(N_b(X)) = N_{a+b}(X)$.*

What I love about this fact is that it comes from a natural question that a five-year-old can (and did) appreciate, and yet the answer is in a precise sense equivalent to a fundamental mathematical fact: the triangle inequality. We will show two inclusions: $N_{a+b}(X) \subset N_a(N_b(X))$ and $N_a(N_b(X)) \subset N_{a+b}(X)$.

The first doesn't need the triangle inequality: suppose we can get to a point p from a point in X by taking a step of size at most $a + b$. Then by dividing this step into two in a ratio of b to a we see that we can also get there from a point in X by taking a step of size at most b, followed by a step at most a.

Conversely suppose there is a point p we can get to from a point in X by a step of size at most b followed by a step of size at most a. Then by the triangle inequality, the total distance we've traveled is at most $a + b$, so we can get there in a single step at most that large, concluding the proof.

You can see that Theorem 1 is equivalent to the triangle inequality. Experts may want to consider different spaces than the plane, where the triangle inequality may or may not hold. This leads to a number of very interesting questions.

But let's return to our original question. If there were n parentheses in the reworded question, then we can reword it yet again as:

Reworded question′. In some sense does $N_{nr}(X)$ become more and more circular as n gets large?

We know we've made progress, because our question has gotten shorter! And clearly the "right" question to ask is:

Reworded question″. In some sense does $N_R(X)$ become more and more circular as R gets large?

We are now ready to see that the answer is yes! (And after we have the answer, we will have the proper question.) The key idea is the following. (We will use this simple principle a lot.)

Lemma 2. *If A is contained in B, then $N_R(A)$ is contained in $N_R(B)$.*

A little thought will convince you why it is true: if every point in A is also a point in B, then everything you can get to by a step (of size at most R) from a point in A you can also get to by a step (of size at most R) from a point in B—indeed the same point!

Equipped with this insight, let's answer the question. Fix a point p of X. Let D_t denote the disk of radius t around p. Pick an r such that $X \subset D_r$. Then

$$\{p\} \subset X \subset D_r.$$

Applying the lemma, we find that

$$D_R = N_R(\{p\}) \subset N_R(X) \subset N_R(D_r) = D_{R+r}.$$

So the boundary of the Rth doodle $N_R(X)$ is stuck between two circles around p, one of radius R and one of radius $R + r$. As $R \to \infty$, the ratio of these two radii goes

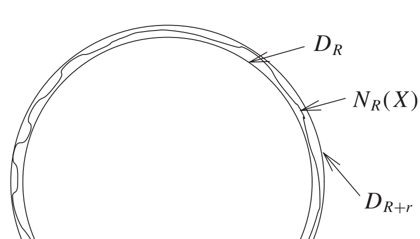

Figure 3. $N_R(X)$ for R huge, seen from a distance.

to 1. So if you were to zoom out and look at $N_R(X)$ from a distance (see Figure 3), it would look very much like a disk. And undeniably this means that it looks more and more circular—we have definitively answered the question, and retroactively made the question precise.

(You may find other questions gnawing at you: what precisely is happening in that narrow annulus between the two circles? Have we answered the question as fully as we might? Is the doodle getting less and less curved as it gets bigger? I encourage you to follow up on these: first figure out the right questions, and then possibly even find some answers.)

Now that we have answered the question that had been plaguing me from a young age, you may think we are done with this doodle, but we are just getting started!

PERIMETERS AND AREAS. The next question is one I saw in high school, although not in the language of doodles.

Question. Let X be a convex polygon (including its interior). What is the perimeter of $N_r(X)$ in terms of X?

In order to answer the question, you first must understand the geometry of the doodle in this situation. The boundary $N_r(X)$ has some straight portions and some circular portions, which suggests that it wants to be cut up in a certain way, shown in Figure 4.

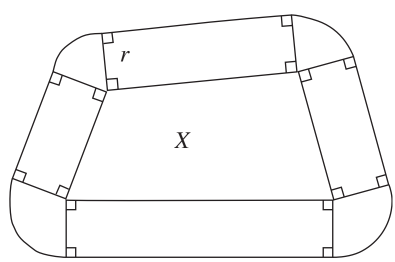

Figure 4. Dissecting $N_r(X)$ when X is a convex polygon.

You will see that the boundary of $N_r(X)$ has a number of circular arcs whose union is precisely a circle of radius r, so their combined length is the circumference of a circle, $2\pi r$. The remainder of the perimeter is a union of straight edges, each parallel to, and equal in length to, a side of X. Thus their total length is precisely the perimeter $\mathrm{Perim}(X)$ of X. We conclude:

$$\mathrm{Perim}(N_r(X)) = \mathrm{Perim}(X) + 2\pi r.$$

If you followed that, you should be able to use Figure 4 to answer the following problem. Here and later, let $\mathrm{Area}(Z)$ denote the area of a plane set.

Problem. If X is a convex polygon (including its interior), show that

$$\mathrm{Area}(N_r(X)) = \mathrm{Area}(X) + \mathrm{Perim}(X)r + \pi r^2. \tag{1}$$

There are a huge number of interesting questions here, and which ones you like will depend on your tastes and what you know. I want to emphasize that these questions are essentially forced on us by the original doodle: of course we would do the doodle around polygons, and of course we would wonder about basic notions such as area and perimeter.

Question. Notice that the derivative of the area is the perimeter:

$$\frac{d}{dr}\,\text{Area}(N_r(X)) = \text{Perim}(N_r(X)).$$

Why is this?

You may have noticed this fact for a circle (area πr^2, perimeter $2\pi r$), and that it is false for a square (area s^2, perimeter $4s$), but correct if you take the radius of the square to be the distance from the center to the midpoint of a side (area $4r^2$, perimeter $8r$). (What should the "radius" be for an equilateral triangle to make this work? How about a regular n-gon? What can you say for an arbitrary convex n-gon? Something more general?)

There is something else to observe in the area formula (1): for any given shape X, it is a *quadratic formula in* r, and *the coefficients all have important geometric meaning*: the constant term is the area (not surprisingly); the coefficient of the linear term is the perimeter; and the coefficient of the quadratic term is the important geometric constant π. This behavior—roughly speaking, of certain geometric notions such as areas (or volumes) being polynomials, and their coefficients being very useful—will be a recurring theme in the rest of this article.

LET'S GENERALIZE. Suppose now that X is convex, but not a polygon. It will be convenient to redefine N_r in terms of the normal vector. (For those familiar with this notion: we need to assume that the boundary is differentiable, which immediately excludes our friends the polygons, although you may quickly realize how to extend the definition a little bit in this case.)

As you continue to read, you may want to ask yourself *why* I am changing the definition in the middle of our discussion—changing horses midstream is always dangerous. (Partial answer: it makes proofs easier and more natural, and it will generalize appropriately when we give up the assumption of convexity.) But being flexible with definitions is something we are used to doing when trying to figure out laws of nature.

Informal redefinition. *I walk around the outside of* X counterclockwise, *sticking my right hand out and marking the path with a marker. By a remarkable coincidence, my arm has length precisely* r. *We call the resulting doodle* $N_r(X)$.

(As you read on, you should of course ask about the use of "counterclockwise" or "right hand"—why did I make these choices? What would happen if you did something different? What happens if I kept my arm pointing *forward*—using a tangent vector rather than a normal vector?)

With this new definition, we have a remarkable fact: the perimeter and area formulas still hold:

$$\begin{aligned}\text{Perim}(N_r(X)) &= \text{Perim}(X) + 2\pi r, \\ \text{Area}(N_r(X)) &= \text{Area}(X) + \text{Perim}(X)r + \pi r^2.\end{aligned} \tag{2}$$

Can you prove it?

The String-Around-the-Earth Puzzle. This relates to a well-known puzzle. Suppose string is wrapped tightly around the equator of the Earth. Someone with too much time on his or her hands adds 1 meter more of string, and raises the string to a constant height. How high off the ground is the string?

If you haven't seen this problem before, you should take a break and try to answer it. Before you work it out, you should first test your intuition: do you expect the answer to be closer to a millimeter, a micrometer, or a nanometer (10^{-3} m, 10^{-6} m, or 10^{-9} m)? I won't tell you the diameter of the earth, but you probably have a good enough idea to answer the question. (Once you answer the problem: how can it be seen as a consequence of (2)?)

GENERALIZE FURTHER! Let's give up convexity. What happens to our formulas? How should we change the statement of our "laws of nature" so they remain true in this more general situation? To get an idea, we of course try an example. What happens if our doodle is big enough so it runs into itself, as in Figure 5? It turns out that the perimeter formula still holds, but the area of the region contained "within" the doodle is now not correct without modification. You might quickly see what the answer is—the area in the "overlap" should count doubly, and indeed if we take this definition, the area formula (1) *remains true*. This is nature telling us that this is the correct definition—the notion of *area with multiplicity* is forced upon us by asking this question.

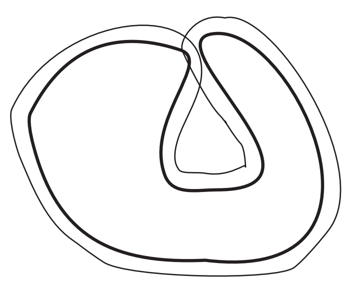

Figure 5. A doodle around a seriously nonconvex shape.

WHAT ABOUT A FIGURE EIGHT? We have given up convexity, so let's try something weirder still. What happens if we do the doodle around a figure eight, as in Figure 6? We first have to figure out the meaning of the word "around." Given our

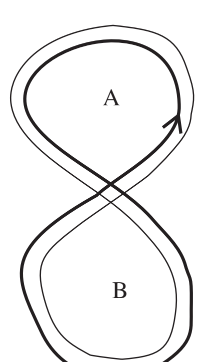

Figure 6. Doodling around a figure eight.

definition of the doodle, we must choose a direction around the figure eight, so we can choose a normal direction. That choice is indicated in the figure.

Once again we get a perimeter formula that is very nice—but it is *different*:

$$\operatorname{Perim}(N_r(X)) = \operatorname{Perim}(X).$$

Why is this true? What "causes" the change from (2)?

And again, we have a good formula for area, *once we define area appropriately*. I'll not tell you the precise answer, but I'll give you a hint: we consider the area of the top part (labeled "A" in Figure 6) with one multiplicity, and the bottom part (labeled "B") with another multiplicity. With this convention, we have a nice formula:

$$\operatorname{Area}(N_r(X)) = \operatorname{Area}(X) + \operatorname{Perim}(X)r.$$

As with the perimeter, this is a different formula than the earlier one (1), but it is different "in the same way" — for example, once again the derivative of the area (with respect to r) is the perimeter.

To shed some light on this, we try to come up with more examples.

THE DOUBLED LOOP. Consider the "doubled loop" of Figure 7. The perimeter formula is again nice, and again different from (2), but now the "error" is opposite that of the figure eight:

$$\operatorname{Perim}(N_r(X)) = \operatorname{Perim}(X) + 4\pi r.$$

And once again, to define area correctly, we need to weight the areas appropriately, giving one weight to region A and one to region B. With this definition, we get a nice area formula:

$$\operatorname{Area}(N_r(X)) = \operatorname{Area}(X) + \operatorname{Perim}(X)r + 2\pi r^2.$$

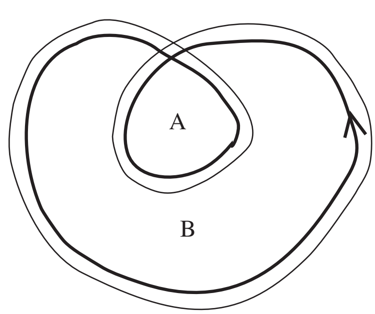

Figure 7. Doodling around a "doubled loop."

What explains the different answers to the original problem, the figure eight doodle, and the doubled loop doodle? How could you immediately tell by looking at a figure? A related question is: what rule tells you how to weight each piece? You know you understand the answers to these questions when you can predict the perimeter and area formulas for the crazy shape in Figure 8.

The answers to both of these questions are in terms of the *winding number*. We have been inevitably led to this fundamental notion in mathematics and physics from this basic doodle, and just asking what happens to basic ideas like perimeter and area when we do it around crazy shapes.

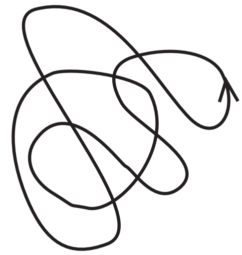

Figure 8. A crazy shape. What are the perimeter and area formulas? What does the area formula *mean*?

SHAPES WITH HOLES. Having explored drawings that cross themselves, let's now restrict ourselves to drawings that don't. What happens to perimeter and area for the shape shown in Figure 9? As you can tell, I first did these doodles in geography class: one can visualize these as landmasses in an ocean, with "holes" that are lakes, that might even have islands in them. The doodles are like ripples in the water. As usual, we orient our border so that the "interior" (land) is always on our left.

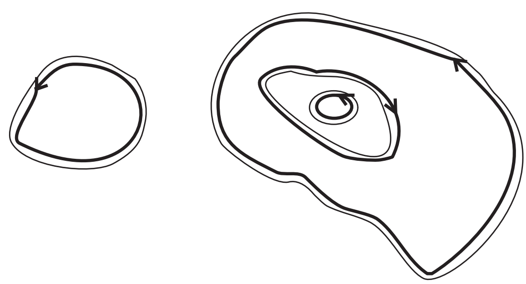

Figure 9. Doodling in geography class.

In order to answer such a general question, as always in science we consider the simplest case that we have not seen before, where there is one island with one lake (Figure 10). Then once again we get nice formulas (and all area multiplicities are one):

$$\mathrm{Perim}(N_r(X)) = \mathrm{Perim}(X)$$
$$\mathrm{Area}(N_r(X)) = \mathrm{Area}(X) + r\,\mathrm{Perim}(X).$$

You may be able to prove this *given* the original perimeter and area formulas (2) and (1), applying it separately to each of the two "coastlines."

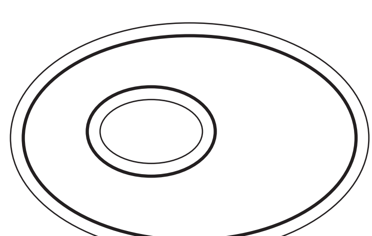

Figure 10. Doodling around an island with a lake.

If you can, then you should be able to work out the perimeter and area formulas for the archipelago of Figure 9, and you may be surprised to find out that we have inevitably been led to a notion from topology not usually seen by doodlers. The area formula is:

$$\mathrm{Area}(N_r(X)) = \mathrm{Area}(X) + \mathrm{Perim}(X)r + \chi(X)\pi r^2,$$

where χ is the number of pieces minus the number of holes. This is the Euler characteristic of X! (I've not told you the perimeter formula for $\mathrm{Perim}(N_r(X))$—what is it?)

MORE DIMENSIONS. Another natural way to generalize mathematical ideas is by changing the dimension. We will proceed directly to three dimensions, but you may want to work out the one-dimensional case yourself.

I want to start with convex polyhedra, which means that I will revert to our original definition: $N_r(X)$ once again is the set of points of distance at most r from some point in X. To see how this notion behaves, we begin with a particularly simple polyhedron. Let X be a box of length ℓ, width w, and height h. The analogues of perimeter and area are now surface area and volume respectively. We will denote them $\mathrm{Surf}(X)$ and $\mathrm{Vol}(X)$.

Problem. Show that

$$\mathrm{Vol}(N_r(X)) = \mathrm{Vol}(X) + \mathrm{Surf}(X)r + (\ell + w + h)\pi r^2 + \frac{4}{3}\pi r^3. \tag{3}$$

Once again, this is a problem I saw in high school, although at the time I had no idea that there was any connection to my earlier doodling musings. (If you are stuck, here is a hint: think about the argument in the two-dimensional case.)

The volume formula (3) has a feature that is interesting in light of our two-dimensional discussion: it is a cubic in r, and some of the coefficients are geometrically interesting. Indeed if X is a convex body in general, $\mathrm{Vol}(N_r(X))$ is cubic, and you may be able to guess many of the coefficients:

$$\mathrm{Vol}(N_r(X)) = \mathrm{Vol}(X) + \mathrm{Surf}(X)r + (???)r^2 + \frac{4}{3}\pi r^3.$$

We like the constant term (the volume) and the linear term (the surface area), and the cubic term is the volume of the unit sphere (in analogy with the two-dimensional case). But what is the quadratic term? In the case of the box, it has a simple description, but this doesn't generalize well. (What is it for the sphere?) If we believe that volume, surface area, and the unit sphere are important notions, we should believe that this new mystery number—some function of X—should be important as well. I'll mention a couple of places where it comes up naturally—one fun and then one serious—and then we'll figure out how to interpret it, by considering the generalization to an arbitrary number of dimensions.

A BEAUTIFUL RUSSIAN PROBLEM. Its first appearance for me was in one of my favorite problems. I'd heard it by word of mouth, and had heard that it originated in St. Petersburg. I've since seen it in Peter Winkler's book [**3**, p. 63], where it is said to have originated in Moscow. I will avoid this delicate geopolitical issue, and just describe the problem.

A Russian train company has a rule: you are not allowed packages (in a box) whose sum of dimensions (length plus width plus height) is more than 1 m. Is it possible to cheat by taking an illegal box, putting it in a larger box that is legal? Suppose X and Y are boxes, and X is contained in Y ($X \subset Y$). Let ℓ_X, w_X, and h_X refer to the length, width, and height of X, and similarly for Y. Is it true that $\ell_X + w_X + h_X \leq \ell_Y + w_Y + h_Y$?

Of course, you can't win by nestling one box in the corner of the other, so that the three "axes" of the boxes line up, as then $\ell_X \leq \ell_Y$, $w_X \leq w_Y$, and $h_X \leq h_Y$. But what happens if X has a drastically different "shape" from Y, and is placed in some tilted way?

The answer must of course be "no": we know from experience that on mathematical issues, Russians know what they are doing. Here is why.

If $X \subset Y$, then $N_r(X) \subset N_r(Y)$ (for any r), essentially by Lemma 2. From our formula for the volume of the doodle around a box (3), this means:

$$\begin{aligned} &\mathrm{Vol}(X) + \mathrm{Surf}(X)r + (\ell_X + w_X + h_X)\pi r^2 + \frac{4}{3}\pi r^3 \\ &\quad \leq \mathrm{Vol}(Y) + \mathrm{Surf}(Y)r + (\ell_Y + w_Y + h_Y)\pi r^2 + \frac{4}{3}\pi r^3. \end{aligned}$$

We cancel the common cubic terms, leaving:

$$\begin{aligned} &\mathrm{Vol}(X) + \mathrm{Surf}(X)r + (\ell_X + w_X + h_X)\pi r^2 \\ &\quad \leq \mathrm{Vol}(Y) + \mathrm{Surf}(Y)r + (\ell_Y + w_Y + h_Y)\pi r^2. \end{aligned}$$

Now we notice that both sides are quadratic in r (and the coefficients are constants of the problem). So for the right side to remain at least as big as the left side as r gets large, the quadratic coefficient of the right side must be at least as large as that of the left:

$$\ell_X + w_X + h_X \leq \ell_Y + w_Y + h_Y.$$

And we're done!

HILBERT'S THIRD PROBLEM. This "linear invariant" turns up in fancier places as well. It is clear that if you have two polygons of different areas, then you can't dissect the first into pieces, and rearrange them to form the second. (By "dissect," I mean with a finite number of straight line cuts—no Banach-Tarski funny business allowed!) This is because we have an invariant—area—preserved by dissections. It is a fun (and nontrivial) problem to show that this is the *only* invariant: given two polygons of the same area, it *is* possible to dissect the first into pieces and reassemble them to form the second. (Possible hint: show that you can dissect any triangle and rearrange the pieces into a square of the same area. Show that you can dissect two squares and rearrange the pieces into a single square of the same area.)

Hilbert's third problem (the easiest, or at least the first to be solved) was whether the same is true in three dimensions: is volume the only dissection invariant? It turns out that there is precisely one more invariant—the *Dehn invariant*—and this invariant is a close relative of our one-dimensional invariant! And then it is possible to check that the Dehn invariants of a regular tetrahedron and a cube of the same volume are different—it is impossible to dissect a regular tetrahedron and rearrange the pieces to form a cube.

IN DIMENSION n. In dimension n, this polynomiality behavior of the (hyper)volume of $N_r(X)$ still holds, except now there are more mysterious coefficients:

$$\text{HyperVol}(N_r(X)) = a_0 + a_1 r + \cdots + a_n r^n.$$

In light of the situation in dimensions two and three (and one), you might guess (correctly) that the constant term is the hypervolume, and the linear term is the hypersurface area, and that the top term is the volume of the unit ball in n-space.

We can interpret all the higher invariants in a consistent way that makes them more down-to-earth, but which forces us to reinterpret notions we thought we understood well. I learned this point of view from G.-C. Rota [**1**].

The trick is that we need to reinterpret invariants we thought we knew well in completely new ways.

A NEW TAKE ON THE ONE-DIMENSIONAL INVARIANT IN DIMENSION 2 (“PERIMETER”). Here is a remarkable fact.

Theorem 3. *The average length of the shadow of a convex region of the plane, multiplied by π, is the perimeter!*

The shadow of a bounded convex region of the plane is a line segment. By “average length of a shadow” I mean the following. Consider the shadow of the body in a random direction, and average over all directions (Figure 11). This can be made more precise.

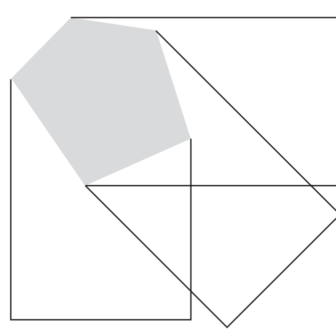

Figure 11. The perimeter of a convex region of the plane is the average length of its shadow, times π.

For example, the shadow of a circle of radius r in *any* direction is $2r$, so the average length of a shadow is certainly $2r$, implying that the perimeter is $2\pi r$, which we know well.

To test your understanding of this fact, try to solve the following using Theorem 3.

Problem. What is the average length of the shadow of a line segment of length 1? (What is the perimeter of the line segment?)

If you also solve this by calculus without invoking Theorem 3, you will see that Theorem 3 is subtle. (But conversely, you may be able to use the above problem to prove Theorem 3 in the case of convex polygons. There is a very clever proof of it in general using the trivial case of the circle.)

If you solve the problem, you’ll be able to understand Buffon’s famous experimental method of determining the value of π.

Buffon's Needle Problem (1733). Suppose the floor is marked with parallel lines spaced 1 inch apart, and a 1-inch needle is randomly dropped on the floor. What is the probability that it will cross one of the lines?

To test your answer, you can actually perform this experiment. If you do it 1000 times, you should see that the needle crosses a line roughly 637 times.

A NEW TAKE ON THE TWO-DIMENSIONAL INVARIANT IN DIMENSION 3 ("SURFACE AREA").

Theorem 4. *Consider the average area of the shadow of a convex region of three-space, and multiply by* 4*. The result is the surface area.*

For example, a sphere of radius r has a shadow of area πr^2 in every direction, so that's certainly the average. Therefore, its surface area is $4\pi r^2$.

Hard problem. Prove this for a box.

Here is a hypothetical possible application of this method. One way in which we find new planets orbiting distant stars is if we are lucky enough for the planet's orbit to pass between us and that star. We get a clue to the existence of the planet by observing that the star appears to dim periodically, and the data of how it dims gives us further information, such as the planet's period of revolution. But you can imagine that it could give more information too, for example information on how spherical the planet might be (if there is some wobbling in the dimming), and possibly even its period of rotation if it is not spherical (the period of the wobbling). You could imagine that Theorem 4 could give a good approximation of its surface area too.

THE ONE-DIMENSIONAL INVARIANT IN DIMENSION 3 ($\pi(\ell + w + h)$ FOR A BOX), AND THE GENERAL CASE. There is an analogous description of our new "mystery invariant" in dimension 3: given a convex body X, take the average length of its shadow on a 1-dimensional screen. Multiply by a certain constant.

Problem. What is this constant? (Hint: consider the case of a sphere.)

You may now be ready for the general case:

Problem. Give an interpretation of the k-dimensional invariant in dimension n.

POLYNOMIALITY AND THE MODULI SPACE OF CURVES. In case you think this is all just fun and games, and unrelated to how mathematics is actually done on the cutting edge, you are only half right: it *is* fun and games, but these insights turn up even in current research.

I have seen the idea of taking advantage of unexpected polynomiality in a number of places, including in my own work on the moduli space of curves, but I am not going to discuss that here. Instead, I would like to tell you about some remarkable recent developments in an area where it has been very fruitful, which will allow some name-dropping. This is not the place for me to be precise and rigorous, so I will content myself with trying to get across the general sense of the field. (For a more technical expository article about some of this material, in which polynomiality makes an appearance, see [**2**].)

Connected one-dimensional manifolds are (at least in a naive sense) boring—topologically, they are just circles or line segments. Dimension two is more interesting—topologically, there are many. One is shown in Figure 12. Surfaces have further structure if you include the notion of distance, or (essentially equivalently) a complex structure. This gives the notion of a *Riemann surface*. There are in general many Riemann surfaces of each topological type, and they are parametrized by something called a *moduli space*. In algebraic geometry, this space (in the case of compact surfaces) was constructed by Mumford in the early 1960s, and in 1969 Deligne and Mumford gave a beautiful compactification. This space is of central importance—any statement about it is a universal statement applying to all families of surfaces.

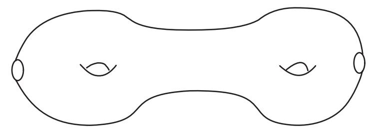

Figure 12. A surface of genus 2 with two punctures.

In 1991, Witten made a remarkable conjecture, giving very strong information about its shape. (Somewhat more formally: a consequence of his conjecture is an effective recursion by which you can intersect all cohomology classes that seem to simply come from geometry, to get numbers.) The conjecture was a bolt from the blue at the time: it connected the moduli space to very distant and seemingly unrelated parts of mathematics. As a sign of the richness of Witten's conjecture, there have been a large number of proofs (perhaps even ten, depending on how you count), and *every single one* has shed considerable new light. The first few seemed to be unrelated to each other, and the most recent few have really started to tie together our understanding of the magic behind Witten's conjecture.

The first proof was by Kontsevich in 1992; the second was by Okounkov and Pandharipande. (Mumford, Deligne, Witten, Kontsevich, and Okounkov all won Fields Medals.) The next proof was, remarkably, by a graduate student, Maryam Mirzakhani (now a colleague of mine, a full professor at Stanford). Her argument, in hyperbolic geometry, was dramatically different from the other proofs, and only now are we beginning to understand how it fits in with the rest. Her beautiful insight involved uncovering an unexpected polynomiality, somewhat related in spirit to what we have been discussing. I will attempt to get across the key idea, at the cost of mutilating the substance.

Instead of considering compact surfaces, consider surfaces with n disks cut out, of various perimeters p_1 through p_n. One advantage of considering these more general objects is that you can sew two holes of the same perimeter together (thus sewing together two different surfaces, or sewing one surface to itself). If you fix the sizes of the holes (the p_i's), the moduli space naturally has a measure on it, and you can measure its "volume."

Mirzakhani's work contains a number of miracles. First, she shows that you could compute the volume inductively (roughly, on the number of holes of the surface) by sewing. Second, she shows that the volume is a *polynomial* in the p_i—a magic polynomiality. And we know that if a polynomial turns up, we should hope that its coefficients have good geometric meaning. Mirzakhani's third miracle is that the leading coefficients are precisely the numbers that are the subject of Witten's conjecture. And her fourth miracle is that her recursions translate directly into the recursions predicted by Witten.

Within the structure of her argument we see a number of themes we have already stumbled upon: (i) finding that a volume has an unexpected polynomial structure, (ii) figuring out how to compute the polynomial, possibly by cutting and pasting, (iii) realizing that the coefficients of the polynomial have important and surprising geometric meaning, and (iv) getting at some of that information by an asymptotic argument (as we did with the Russian problem).

CONCLUSION. We started with a single simple doodle, and just followed where it took us. It led us inevitably (at least implicitly) through beautiful elementary problems, the triangle inequality, area with multiplicity, winding numbers, topology, differential geometry, physics, a Hilbert problem, work of a number of Fields medalists, and current research in algebraic and hyperbolic geometry. In some sense our journey is a metaphor for mathematical exploration in general.

ACKNOWLEDGMENTS. I've heard the above stories from too many sources to list—friends, teachers, students, colleagues, acquaintances, books. So I will just thank everyone en masse, and apologize to anyone I've left out.

REFERENCES

1. G.-C. Rota, Geometric probability, *Math. Intelligencer* **20**(4) (1998) 11–16. `doi:10.1007/BF03025223`
2. R. Vakil, The moduli space of curves and its tautological ring, *Notices Amer. Math. Soc.* **50** (2003) 647–648.
3. P. Winkler, *Mathematical Mind-Benders*, A K Peters, Wellesley, MA, 2007.

RAVI VAKIL received his B.Sc. from the University of Toronto in 1992, and his Ph.D. from Harvard in 1997. He taught at Princeton and MIT before moving to Stanford in 2001. He is in algebraic geometer.
Dept. of Mathematics, Stanford University, Stanford CA 94305
vakil@math.stanford.edu