

# HOW NOT TO BE WRONG



THE POWER *of*  
MATHEMATICAL THINKING

JORDAN  
ELLENBERG

# HOW NOT TO BE WRONG

The Power of Mathematical Thinking

• • • • •

JORDAN ELLENBERG

THE PENGUIN PRESS

NEW YORK

2014

THE PENGUIN PRESS  
Published by the Penguin Group  
Penguin Group (USA) LLC  
375 Hudson Street  
New York, New York 10014



USA · Canada · UK · Ireland · Australia  
New Zealand · India · South Africa · China  
penguin.com  
A Penguin Random House Company

First published by The Penguin Press,  
a member of Penguin Group (USA) LLC, 2014

Copyright © 2014 by Jordan Ellenberg

Penguin supports copyright. Copyright fuels creativity, encourages diverse voices, promotes free speech, and creates a vibrant culture. Thank you for buying an authorized edition of this book and for complying with copyright laws by not reproducing, scanning, or distributing any part of it in any form without permission. You are supporting writers and allowing Penguin to continue to publish books for every reader.

Grateful acknowledgment is made for permission to reprint excerpts from the following copyrighted works:

“Soonest Mended” from *The Double Dream of Spring* by John Ashbery.  
Copyright © 1966, 1970 by John Ashbery. Reprinted by permission of Georges Borchardt, Inc., on behalf of the author.

“Sitting on a Fence,” words and music by Ian Cullimore and Paul Heaton.  
Copyright © 1986 Universal Music Publishing Ltd. and Universal / Island Music Ltd. All rights in the United States and Canada controlled and administered by Universal Polygram International Publishing, Inc. All rights reserved. Used by permission. Reprinted by permission of Hal Leonard Corporation.

LIBRARY OF CONGRESS CATALOGING-IN-PUBLICATION DATA

Ellenberg, Jordan, 1971- author.

How not to be wrong : the power of mathematical thinking / Jordan Ellenberg.  
pages cm

Includes bibliographical references and index.

ISBN 978-1-59420-522-4 (hardback)

1. Mathematics—Miscellanea. 2. Mathematical analysis—Miscellanea. I. Title.  
QA99.E45 2014  
510—dc23 2014005394

Printed in the United States of America

1 3 5 7 9 10 8 6 4 2

DESIGNED BY AMANDA DEWEY

## CONTENTS

WHEN AM I GOING TO USE THIS? 1

### PART I

## Linearity

One. LESS LIKE SWEDEN 21

Two. STRAIGHT LOCALLY, CURVED GLOBALLY 31

Three. EVERYONE IS OBESE 50

Four. HOW MUCH IS THAT IN DEAD AMERICANS? 62

Five. MORE PIE THAN PLATE 77

### PART II

## Inference

Six. THE BALTIMORE STOCKBROKER AND  
THE BIBLE CODE 89

Seven. DEAD FISH DON'T READ MINDS 102

Eight. REDUCTIO AD UNLIKELY 131

Nine. THE INTERNATIONAL JOURNAL  
OF HARUSPICY 145

Ten. ARE YOU THERE, GOD? IT'S ME,  
BAYESIAN INFERENCE 163

### PART III

## Expectation

Eleven. WHAT TO EXPECT WHEN YOU'RE  
EXPECTING TO WIN THE LOTTERY 195

Twelve. MISS MORE PLANES! 233

Thirteen. WHERE THE TRAIN TRACKS MEET 253

### PART IV

## Regression

Fourteen. THE TRIUMPH OF MEDIOCRITY 295

Fifteen. GALTON'S ELLIPSE 311

Sixteen. DOES LUNG CANCER MAKE YOU  
SMOKE CIGARETTES? 347

PART V  
Existence

*Seventeen.* THERE IS NO SUCH THING AS  
PUBLIC OPINION 365

*Eighteen.* "OUT OF NOTHING I HAVE CREATED  
A STRANGE NEW UNIVERSE" 393

HOW TO BE RIGHT 421

*Acknowledgments* 439

*Notes* 443

*Index* 459



## WHEN AM I GOING TO USE THIS?

Right now, in a classroom somewhere in the world, a student is mouthing off to her math teacher. The teacher has just asked her to spend a substantial portion of her weekend computing a list of thirty definite integrals.

There are other things the student would rather do. There is, in fact, hardly anything she would *not* rather do. She knows this quite clearly, because she spent a substantial portion of the previous weekend computing a different—but not *very* different—list of thirty definite integrals. She doesn't see the point, and she tells her teacher so. And at some point in this conversation, the student is going to ask the question the teacher fears most:

*“When am I going to use this?”*

Now the math teacher is probably going to say something like:

“I know this seems dull to you, but remember, you don't know what career you'll choose—you may not see the relevance now, but you might go into a field where it'll be really important that you know how to compute definite integrals quickly and correctly by hand.”

This answer is seldom satisfying to the student. That's because it's a lie. And the teacher and the student both know it's a lie. The number of adults who will ever make use of the integral of  $(1 - 3x + 4x^2)^{-2} dx$ , or



the formula for the cosine of  $3\theta$ , or synthetic division of polynomials, can be counted on a few thousand hands.

The lie is not very satisfying to the teacher, either. I should know: in my many years as a math professor I've asked many hundreds of college students to compute lists of definite integrals.

Fortunately, there's a better answer. It goes something like this:

"Mathematics is not just a sequence of computations to be carried out by rote until your patience or stamina runs out—although it might seem that way from what you've been taught in courses called *mathematics*. Those integrals are to mathematics as weight training and calisthenics are to soccer. If you want to play soccer—I mean, *really play*, at a competitive level—you've got to do a lot of boring, repetitive, apparently pointless drills. Do professional players ever *use* those drills? Well, you won't see anybody on the field curling a weight or zigzagging between traffic cones. But you do see players using the strength, speed, insight, and flexibility they built up by doing those drills, week after tedious week. Learning those drills is part of learning soccer.

"If you want to play soccer for a living, or even make the varsity team, you're going to be spending lots of boring weekends on the practice field. There's no other way. But now here's the good news. If the drills are too much for you to take, you can still play for fun, with friends. You can enjoy the thrill of making a slick pass between defenders or scoring from distance just as much as a pro athlete does. You'll be healthier and happier than you would be if you sat home watching the professionals on TV.

"Mathematics is pretty much the same. You may not be aiming for a mathematically oriented career. That's fine—most people aren't. But you can still do math. You probably already *are* doing math, even if you don't call it that. Math is woven into the way we reason. And math makes you better at things. Knowing mathematics is like wearing a pair of X-ray specs that reveal hidden structures underneath the messy and chaotic surface of the world. Math is a science of not being wrong about things, its techniques and habits hammered out by centuries of hard work and argument. With the tools of mathematics in hand, you can understand the world in a deeper, sounder, and more meaningful way. All you need is

a coach, or even just a book, to teach you the rules and some basic tactics. I will be your coach. I will show you how.”

For reasons of time, this is seldom what I actually say in the classroom. But in a book, there’s room to stretch out a little more. I hope to back up the grand claims I just made by showing you that the problems we think about every day—problems of politics, of medicine, of commerce, of theology—are shot through with mathematics. Understanding this gives you access to insights accessible by no other means.

Even if I did give my student the full inspirational speech, she might—if she is really sharp—remain unconvinced.

“That sounds good, Professor,” she’ll say. “But it’s pretty abstract. You say that with mathematics at your disposal you can get things right you’d otherwise get wrong. But what kinds of things? Give me an *actual example*.”

And at that point I would tell her the story of Abraham Wald and the missing bullet holes.

## ABRAHAM WALD AND THE MISSING BULLET HOLES

This story, like many World War II stories, starts with the Nazis hounding a Jew out of Europe and ends with the Nazis regretting it. Abraham Wald was born in 1902 in what was then the city of Klausenburg in what was then the Austro-Hungarian Empire. By the time Wald was a teenager, one world war was in the books and his hometown had become Cluj, Romania. He was the grandson of a rabbi and the son of a kosher baker, but the younger Wald was a mathematician almost from the start. His talent for the subject was quickly recognized, and he was admitted to study mathematics at the University of Vienna, where he was drawn to subjects abstract and recondite even by the standards of pure mathematics: set theory and metric spaces.

But when Wald’s studies were completed, it was the mid-1930s, Austria was deep in economic distress, and there was no possibility that a foreigner could be hired as a professor in Vienna. Wald was rescued by a job offer from Oskar Morgenstern. Morgenstern would later immigrate

to the United States and help invent game theory, but in 1933 he was the director of the Austrian Institute for Economic Research, and he hired Wald at a small salary to do mathematical odd jobs. That turned out to be a good move for Wald: his experience in economics got him a fellowship offer at the Cowles Commission, an economic institute then located in Colorado Springs. Despite the ever-worsening political situation, Wald was reluctant to take a step that would lead him away from pure mathematics for good. But then the Nazis conquered Austria, making Wald's decision substantially easier. After just a few months in Colorado, he was offered a professorship of statistics at Columbia; he packed up once again and moved to New York.

And that was where he fought the war.

The Statistical Research Group (SRG), where Wald spent much of World War II, was a classified program that yoked the assembled might of American statisticians to the war effort—something like the Manhattan Project, except the weapons being developed were equations, not explosives. And the SRG was actually *in* Manhattan, at 401 West 118th Street in Morningside Heights, just a block away from Columbia University. The building now houses Columbia faculty apartments and some doctor's offices, but in 1943 it was the buzzing, sparking nerve center of wartime math. At the Applied Mathematics Group—Columbia, dozens of young women bent over Marchant desktop calculators were calculating formulas for the optimal curve a fighter should trace out through the air in order to keep an enemy plane in its gunsights. In another apartment, a team of researchers from Princeton was developing protocols for strategic bombing. And Columbia's wing of the atom bomb project was right next door.

But the SRG was the most high-powered, and ultimately the most influential, of any of these groups. The atmosphere combined the intellectual openness and intensity of an academic department with the shared sense of purpose that comes only with high stakes. "When we made recommendations," W. Allen Wallis, the director, wrote, "frequently things happened. Fighter planes entered combat with their machine guns loaded according to Jack Wolfowitz's\* recommendations about mixing

\* Paul's dad.

types of ammunition, and maybe the pilots came back or maybe they didn't. Navy planes launched rockets whose propellants had been accepted by Abe Girshick's sampling-inspection plans, and maybe the rockets exploded and destroyed our own planes and pilots or maybe they destroyed the target."

The mathematical talent at hand was equal to the gravity of the task. In Wallis's words, the SRG was "the most extraordinary group of statisticians ever organized, taking into account both number and quality." Frederick Mosteller, who would later found Harvard's statistics department, was there. So was Leonard Jimmie Savage, the pioneer of decision theory and great advocate of the field that came to be called Bayesian statistics.\* Norbert Wiener, the MIT mathematician and the creator of cybernetics, dropped by from time to time. This was a group where Milton Friedman, the future Nobel in economics, was often the fourth-smartest person in the room.

The *smartest* person in the room was usually Abraham Wald. Wald had been Allen Wallis's teacher at Columbia, and functioned as a kind of mathematical eminence to the group. Still an "enemy alien," he was not technically allowed to see the classified reports he was producing; the joke around SRG was that the secretaries were required to pull each sheet of notepaper out of his hands as soon as he was finished writing on it. Wald was, in some ways, an unlikely participant. His inclination, as it always had been, was toward abstraction, and away from direct applications. But his motivation to use his talents against the Axis was obvious. And when you needed to turn a vague idea into solid mathematics, Wald was the person you wanted at your side.

So here's the question. You don't want your planes to get shot down by enemy fighters, so you armor them. But armor makes the plane heavier, and heavier planes are less maneuverable and use more fuel. Armoring the planes too much is a problem; armoring the planes too little is a problem. Somewhere in between there's an optimum. The reason you

\* Savage was almost totally blind, able to see only out of one corner of one eye, and at one point spent six months living only on pemmican in order to prove a point about Arctic exploration. Just thought that was worth mentioning.

have a team of mathematicians socked away in an apartment in New York City is to figure out where that optimum is.

The military came to the SRG with some data they thought might be useful. When American planes came back from engagements over Europe, they were covered in bullet holes. But the damage wasn't uniformly distributed across the aircraft. There were more bullet holes in the fuselage, not so many in the engines.

Section of plane	Bullet holes per square foot
Engine	1.11
Fuselage	1.73
Fuel system	1.55
Rest of the plane	1.8

The officers saw an opportunity for efficiency; you can get the same protection with less armor if you concentrate the armor on the places with the greatest need, where the planes are getting hit the most. But exactly how much more armor belonged on those parts of the plane? That was the answer they came to Wald for. It wasn't the answer they got.

The armor, said Wald, doesn't go where the bullet holes are. It goes where the bullet holes *aren't*: on the engines.

Wald's insight was simply to ask: where are the missing holes? The ones that would have been all over the engine casing, if the damage had been spread equally all over the plane? Wald was pretty sure he knew. The missing bullet holes were on the missing planes. The reason planes were coming back with fewer hits to the engine is that planes that got hit in the engine weren't coming back. Whereas the large number of planes returning to base with a thoroughly Swiss-cheesed fuselage is pretty strong evidence that hits to the fuselage can (and therefore should) be tolerated. If you go to the recovery room at the hospital, you'll see a lot more people with bullet holes in their legs than people with bullet holes in their chests. But that's not because people don't get shot in the chest; it's because the people who get shot in the chest don't recover.

Here's an old mathematician's trick that makes the picture perfectly

clear: *set some variables to zero*. In this case, the variable to tweak is the probability that a plane that takes a hit to the engine manages to stay in the air. Setting that probability to zero means a single shot to the engine is guaranteed to bring the plane down. What would the data look like then? You'd have planes coming back with bullet holes all over the wings, the fuselage, the nose—but none at all on the engine. The military analyst has two options for explaining this: either the German bullets just happen to hit every part of the plane but one, or the engine is a point of total vulnerability. Both stories explain the data, but the latter makes a lot more sense. The armor goes where the bullet holes aren't.

Wald's recommendations were quickly put into effect, and were still being used by the navy and the air force through the wars in Korea and Vietnam. I can't tell you exactly how many American planes they saved, though the data-slinging descendants of the SRG inside today's military no doubt have a pretty good idea. One thing the American defense establishment has traditionally understood very well is that countries don't win wars just by being braver than the other side, or freer, or slightly preferred by God. The winners are usually the guys who get 5% fewer of their planes shot down, or use 5% less fuel, or get 5% more nutrition into their infantry at 95% of the cost. That's not the stuff war movies are made of, but it's the stuff wars are made of. And there's math every step of the way.

Why did Wald see what the officers, who had vastly more knowledge and understanding of aerial combat, couldn't? It comes back to his math-trained habits of thought. A mathematician is always asking, "What assumptions are you making? And are they justified?" This can be annoying. But it can also be very productive. In this case, the officers were making an assumption unwittingly: that the planes that came back were a random sample of all the planes. If that were true, you could draw conclusions about the distribution of bullet holes on all the planes by examining the distribution of bullet holes on only the surviving planes. Once you recognize that you've been making that hypothesis, it takes only a moment to realize it's dead wrong; there's no reason at all to expect the planes to have an equal likelihood of survival no matter where they get

hit. In a piece of mathematical lingo we'll come back to in chapter 15, the rate of survival and the location of the bullet holes are *correlated*.

Wald's other advantage was his tendency toward abstraction. Wolfowitz, who had studied under Wald at Columbia, wrote that the problems he favored were "all of the most abstract sort," and that he was "always ready to talk about mathematics, but uninterested in popularization and special applications."

Wald's personality made it hard for him to focus his attention on applied problems, it's true. The details of planes and guns were, to his eye, so much upholstery—he peered right through to the mathematical struts and nails holding the story together. Sometimes that approach can lead you to ignore features of the problem that really matter. But it also lets you see the common skeleton shared by problems that look very different on the surface. Thus you have meaningful experience even in areas where you appear to have none.

To a mathematician, the structure underlying the bullet hole problem is a phenomenon called *survivorship bias*. It arises again and again, in all kinds of contexts. And once you're familiar with it, as Wald was, you're primed to notice it wherever it's hiding.

Like mutual funds. Judging the performance of funds is an area where you don't want to be wrong, even by a little bit. A shift of 1% in annual growth might be the difference between a valuable financial asset and a dog. The funds in Morningstar's Large Blend category, whose mutual funds invest in big companies that roughly represent the S&P 500, look like the former kind. The funds in this class grew an average of 178.4% between 1995 and 2004: a healthy 10.8% per year.\* Sounds like you'd do well, if you had cash on hand, to invest in those funds, no?

Well, no. A 2006 study by Savant Capital shone a somewhat colder light on those numbers. Think again about how Morningstar generates its number. It's 2004, you take all the funds classified as Large Blend, and you see how much they grew over the last ten years.

But something's missing: *the funds that aren't there*. Mutual funds don't live forever. Some flourish, some die. The ones that die are, by and large, the ones that don't make money. So judging a decade's worth of

\* To be fair, the S&P 500 index itself did even better, gaining 212.5% over the same period.

mutual funds by the ones that still exist at the end of the ten years is like judging our pilots' evasive maneuvers by counting the bullet holes in the planes that come back. What would it mean if we never found more than one bullet hole per plane? Not that our pilots are brilliant at dodging enemy fire, but that the planes that got hit twice went down in flames.

The Savant study found that if you included the performance of the dead funds together with the surviving ones, the rate of return dropped down to 134.5%, a much more ordinary 8.9% per year. More recent research backed that up: a comprehensive 2011 study in the *Review of Finance* covering nearly 5,000 funds found that the excess return rate of the 2,641 survivors is about 20% higher than the same figure recomputed to include the funds that didn't make it. The size of the survivorship effect might have surprised investors, but it probably wouldn't have surprised Abraham Wald.

## MATHEMATICS IS THE EXTENSION OF COMMON SENSE BY OTHER MEANS

At this point my teenage interlocutor is going to stop me and ask, quite reasonably: Where's the math? Wald was a mathematician, that's true, and it can't be denied that his solution to the problem of the bullet holes was ingenious, but what's mathematical about it? There was no trig identity to be seen, no integral or inequality or formula.

First of all: Wald did use formulas. I told the story without them, because this is just the introduction. When you write a book explaining human reproduction to preteens, the introduction stops short of the really hydraulic stuff about how babies get inside Mommy's tummy. Instead, you start with something more like "Everything in nature changes; trees lose their leaves in winter only to bloom again in spring; the humble caterpillar enters its chrysalis and emerges as a magnificent butterfly. You are part of nature too, and . . ."

That's the part of the book we're in now.

But we're all adults here. Turning off the soft focus for a second, here's what a sample page of Wald's actual report looks like:



lower bound to the  $Q_i$  could be obtained. The assumption here is that the decrease from  $q_i$  to  $q_{i+1}$  lies between definite limits. Therefore, both an upper and lower bound for the  $Q_i$  can be obtained.

We assume that

$$\lambda_1 q_i \leq q_{i+1} \leq \lambda_2 q_i ,$$

where  $\lambda_1 < \lambda_2 < 1$  and such that the expression

$$\sum_{j=1}^n \frac{a_j}{\lambda_1^{\frac{j(j-1)}{2}}} < 1 - a_0 \quad (A)$$

is satisfied.

The exact solution is tedious but close approximations to the upper and lower bounds to the  $Q_i$  for  $i < n$  can be obtained by the following procedure. The set of hypothetical data used is

$$\begin{array}{ll} a_0 = .780 & a_3 = .010 \\ a_1 = .070 & a_4 = .005 \\ a_2 = .040 & a_5 = .005 \\ \lambda_1 = .80 & \lambda_2 = .90 \end{array}$$

Condition A is satisfied, since by substitution

$$.07 + \frac{.04}{.8} + \frac{.01}{(.8)^3} + \frac{.005}{(.8)^6} + \frac{.005}{(.8)^{10}} = .20529 ,$$

which is less than

$$1 - a_0 = .22 .$$

#### THE LOWER LIMIT OF $Q_i$

The first step is to solve equation 66. This involves the solution of the following four equations for positive roots  $g_0, g_1, g_2, g_3$ .

I hope that wasn't too shocking.

Still, the real *idea* behind Wald's insight doesn't require any of the formalism above. We've already explained it, using no mathematical notation of any kind. So my student's question stands. What makes that math? Isn't it just common sense?

Yes. Mathematics is common sense. On some basic level, this is clear.

How can you explain to someone why adding seven things to five things yields the same result as adding five things to seven? You can't: that fact is baked into our way of thinking about combining things together. Mathematicians like to give names to the phenomena our common sense describes: instead of saying, "*This* thing added to *that* thing is the *same* thing as *that* thing added to *this* thing," we say, "Addition is commutative." Or, because we like our symbols, we write:

For any choice of  $a$  and  $b$ ,  $a + b = b + a$ .

Despite the official-looking formula, we are talking about a fact instinctively understood by every child.

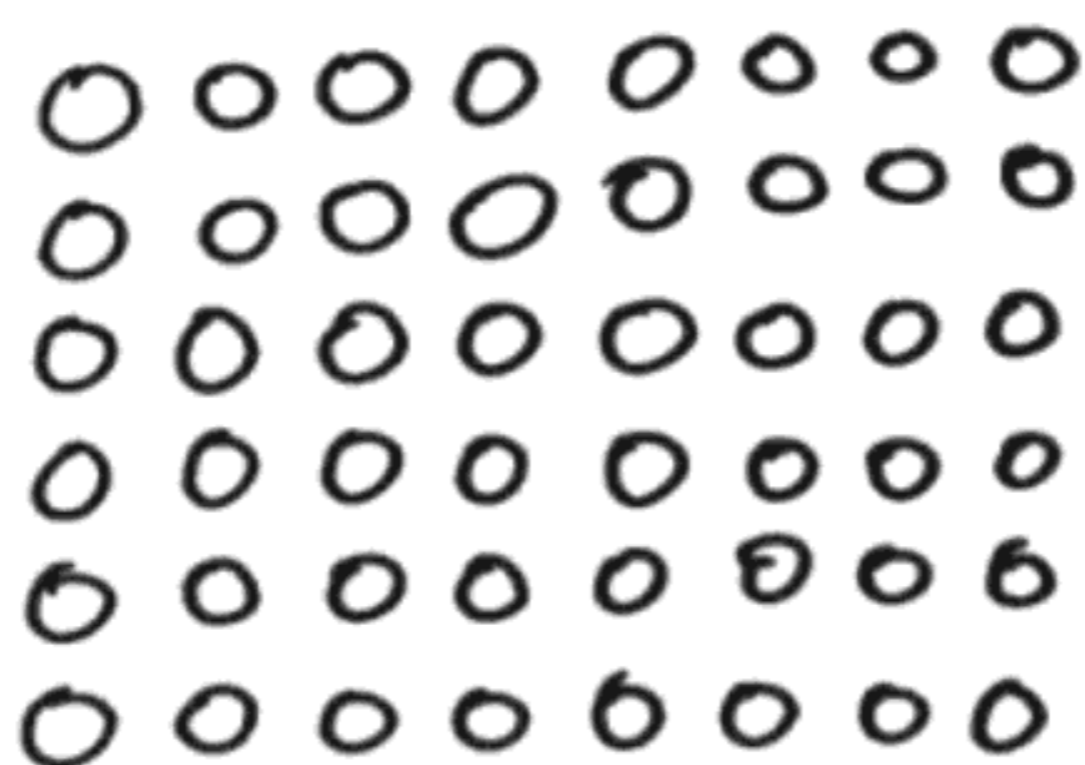
Multiplication is a slightly different story. The formula looks pretty similar:

For any choice of  $a$  and  $b$ ,  $a \times b = b \times a$ .

The mind, presented with this statement, does not say "no duh" quite as instantly as it does for addition. Is it "common sense" that two sets of six things amount to the same as six sets of two?

Maybe not; but it can *become* common sense. Here's my earliest mathematical memory. I'm lying on the floor in my parents' house, my cheek pressed against the shag rug, looking at the stereo. Very probably I am listening to side two of the Beatles' Blue Album. Maybe I'm six. This is the seventies, and therefore the stereo is encased in a pressed wood panel, which has a rectangular array of airholes punched into the side. Eight holes across, six holes up and down. So I'm lying there, looking at the airholes. The six rows of holes. The eight columns of holes. By focusing my gaze in and out I could make my mind flip back and forth between seeing the rows and seeing the columns. Six rows with eight holes each. Eight columns with six holes each.

And then I had it—eight groups of six were the same as six groups of eight. Not because it was a rule I'd been told, but because it could not be any other way. The number of holes in the panel was the number of holes in the panel, no matter which way you counted them.



MY PARENTS' STEREO,  
1977

We tend to teach mathematics as a long list of rules. You learn them in order and you have to obey them, because if you don't obey them you get a C-. *This is not mathematics.* Mathematics is the study of things that come out a certain way because there is no other way they could possibly be.

Now let's be fair: not everything in mathematics can be made as perfectly transparent to our intuition as addition and multiplication. You can't do calculus by common sense. But calculus is still *derived* from our common sense—Newton took our physical intuition about objects moving in straight lines, formalized it, and then built on top of that formal structure a universal mathematical description of motion. Once you have Newton's theory in hand, you can apply it to problems that would make your head spin if you had no equations to help you. In the same way, we have built-in mental systems for assessing the likelihood of an uncertain outcome. But those systems are pretty weak and unreliable, especially when it comes to events of extreme rarity. That's when we shore up our intuition with a few sturdy, well-placed theorems and techniques, and make out of it a mathematical theory of probability.

The specialized language in which mathematicians converse with one another is a magnificent tool for conveying complex ideas precisely and swiftly. But its foreignness can create among outsiders the impression of a sphere of thought totally alien to ordinary thinking. That's exactly wrong.

Math is like an atomic-powered prosthesis that you attach to your

common sense, vastly multiplying its reach and strength. Despite the power of mathematics, and despite its sometimes forbidding notation and abstraction, the actual mental work involved is little different from the way we think about more down-to-earth problems. I find it helpful to keep in mind an image of Iron Man punching a hole through a brick wall. On the one hand, the actual wall-breaking force is being supplied, not by Tony Stark's muscles, but by a series of exquisitely synchronized servomechanisms powered by a compact beta particle generator. On the other hand, from Tony Stark's point of view, what he is doing is punching a wall, exactly as he would without the armor. Only much, much harder.

To paraphrase Clausewitz: Mathematics is the extension of common sense by other means.

Without the rigorous structure that math provides, common sense can lead you astray. That's what happened to the officers who wanted to armor the parts of the planes that were already strong enough. But formal mathematics without common sense—without the constant interplay between abstract reasoning and our intuitions about quantity, time, space, motion, behavior, and uncertainty—would just be a sterile exercise in rule-following and bookkeeping. In other words, math would actually be what the peevish calculus student believes it to be.

That's a real danger. John von Neumann, in his 1947 essay "The Mathematician," warned:

As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality" it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art*. This need not be bad, if the field is surrounded by correlated subjects, which still have closer empirical connections, or if the discipline is under the influence of men with an exceptionally well-developed taste. But there is a grave danger that the subject will develop along the line of least resistance, that the stream, so far from its source, will separate into a multitude of insignificant branches, and that the discipline will become a disorga-

nized mass of details and complexities. In other words, at a great distance from its empirical source, or after much “abstract” inbreeding, a mathematical subject is in danger of degeneration.\*

## WHAT KINDS OF MATHEMATICS WILL APPEAR IN THIS BOOK?

If your acquaintance with mathematics comes entirely from school, you have been told a story that is very limited, and in some important ways false. School mathematics is largely made up of a sequence of facts and rules, facts which are certain, rules which come from a higher authority and cannot be questioned. It treats mathematical matters as completely settled.

Mathematics is not settled. Even concerning the basic objects of study, like numbers and geometric figures, our ignorance is much greater than our knowledge. And the things we do know were arrived at only after massive effort, contention, and confusion. All this sweat and tumult is carefully screened off in your textbook.

There are facts and there are facts, of course. There has never been much controversy about whether  $1 + 2 = 3$ . The question of *how and whether we can truly prove* that  $1 + 2 = 3$ , which wobbles uneasily between mathematics and philosophy, is another story—we return to that at the end of the book. But that the computation is correct is a plain truth. The tumult lies elsewhere. We’ll come within sight of it several times.

Mathematical facts can be simple or complicated, and they can be shallow or profound. This divides the mathematical universe into four quadrants:

\* Von Neumann’s view of the nature of math is solid, but it’s fair to feel a bit queasy about his characterization of mathematics carried out for purely aesthetic ends as “degenerate.” Von Neumann is writing this just ten years after the *entartete Kunst* (“degenerate art”) exhibition in Hitler’s Berlin, whose point was that “*l’art pour l’art*” was the sort of thing that Jews and Communists liked, and was designed to undercut the healthy “realist” art required by a vigorous Teutonic state. Under the circumstances, one feels a little defensive toward mathematics that serves no apparent purpose. A writer with different political commitments than my own would, at this point, bring up von Neumann’s energetic work on the development and delivery of nuclear weapons.

profound	YOU ARE HERE ↕	Fermat Poincaré Riemann Hypothesis derived algebraic geometry perfectoid spaces...
shallow	$1+2=3$ $\sin 2x =$ $2 \sin x \cos x$	$\int x^{-1} \sqrt{x^3-1} dx =$ $\frac{2}{3} \sqrt{x^3-1} \cdot$ $\left(1 - \frac{\operatorname{tanh}^{-1}(\sqrt{1-x^3})}{\sqrt{1-x^3}}\right)$ $+ C$
	simple	complicated

Basic arithmetic facts, like  $1 + 2 = 3$ , are simple and shallow. So are basic identities like  $\sin(2x) = 2 \sin x \cos x$  or the quadratic formula: they might be slightly harder to convince yourself of than  $1 + 2 = 3$ , but in the end they don't have much conceptual heft.

Moving over to complicated/shallow, you have the problem of multiplying two ten-digit numbers, or the computation of an intricate definite integral, or, given a couple of years of graduate school, the trace of Frobenius on a modular form of conductor 2377. It's conceivable you might, for some reason, need to know the answer to such a problem, and it's undeniable that it would be somewhere between annoying and impossible to work it out by hand; or, as in the case of the modular form, it might take some serious schooling even to understand what's being asked for. But knowing those answers doesn't really enrich your knowledge about the world.

The complicated/profound quadrant is where professional mathematicians like me try to spend most of our time. That's where the celebrity theorems and conjectures live: the Riemann Hypothesis, Fermat's Last Theorem,\* the Poincaré Conjecture, P vs. NP, Gödel's Theorem . . . Each

\* Which, among pros, is now called Wiles's Theorem, since Andrew Wiles proved it (with a critical assist from Richard Taylor) and Fermat did not. But the traditional name will probably never be dislodged.

one of these theorems involves ideas of deep meaning, fundamental importance, mind-blowing beauty, and brutal technicality, and each of them is the protagonist of books of its own.

But not this book. This book is going to hang out in the upper left quadrant: simple and profound. The mathematical ideas we want to address are ones that can be engaged with directly and profitably, whether your mathematical training stops at pre-algebra or extends much further. And they are not “mere facts,” like a simple statement of arithmetic—they are principles, whose application extends far beyond the things you’re used to thinking of as mathematical. They are the go-to tools on the utility belt, and used properly they will help you not be wrong.

Pure mathematics can be a kind of convent, a quiet place safely cut off from the pernicious influences of the world’s messiness and inconsistency. I grew up inside those walls. Other math kids I knew were tempted by applications to physics, or genomics, or the black art of hedge fund management, but I wanted no such *rumspringa*.\* As a graduate student, I dedicated myself to number theory, what Gauss called “the queen of mathematics,” the purest of the pure subjects, the sealed garden at the center of the convent, where we contemplated the same questions about numbers and equations that troubled the Greeks and have gotten hardly less vexing in the twenty-five hundred years since.

At first I worked on number theory with a classical flavor, proving facts about sums of fourth powers of whole numbers that I could, if pressed, explain to my family at Thanksgiving, even if I couldn’t explain how I proved what I proved. But before long I got enticed into even more abstract realms, investigating problems where the basic actors—“residually modular Galois representations,” “cohomology of moduli schemes,” “dynamical systems on homogeneous spaces,” things like that—were impossible to talk about outside the archipelago of seminar halls and faculty lounges that stretches from Oxford to Princeton to Kyoto to Paris to Madison, Wisconsin, where I’m a professor now. When I tell you this stuff is thrilling, and meaningful, and beautiful, and that

\* To be honest, I did spend some part of my early twenties thinking I might want to be a Serious Literary Novelist. I even finished a Serious Literary Novel, called *The Grasshopper King*, and got it published. But in the process I discovered that every day I devoted to Serious Literary Novel writing was a day half spent moping around wishing I were working on math problems.

I'll never get tired of thinking about it, you may just have to believe me, because it takes a long education just to get to the point where the objects of study rear into view.

But something funny happened. The more abstract and distant from lived experience my research got, the more I started to notice how much math was going on in the world outside the walls. Not Galois representations or cohomology, but ideas that were simpler, older, and just as deep—the northwest quadrant of the conceptual foursquare. I started writing articles for magazines and newspapers about the way the world looked through a mathematical lens, and I found, to my surprise, that even people who said they hated math were willing to read them. It was a kind of math teaching, but very different from what we do in a classroom.

What it has in common with the classroom is that the reader gets asked to do some work. Back to von Neumann on “The Mathematician”:

“It is harder to understand the mechanism of an airplane, and the theories of the forces which lift and which propel it, than merely to ride in it, to be elevated and transported by it—or even to steer it. It is exceptional that one should be able to acquire the understanding of a process without having previously acquired a deep familiarity with running it, with using it, before one has assimilated it in an instinctive and empirical way.”

In other words: it is pretty hard to *understand* mathematics without *doing* some mathematics. There's no royal road to geometry, as Euclid told Ptolemy, or maybe, depending on your source, as Menaechmus told Alexander the Great. (Let's face it, famous old maxims attributed to ancient scientists are probably made up, but they're no less instructive for that.)

This will not be the kind of book where I make grand, vague gestures at great monuments of mathematics, and instruct you in the proper manner of admiring them from a great distance. We are here to get our hands a little dirty. We'll compute some things. There will be a few formulas and equations, when I need them to make a point. No formal math beyond arithmetic will be required, though lots of math way beyond arithmetic will be explained. I'll draw some crude graphs and charts. We'll encounter some topics from school math, outside their usual habitat; we'll see how trigonometric functions describe the extent to which two



variables are related to each other, what calculus has to say about the relationship between linear and nonlinear phenomena, and how the quadratic formula serves as a cognitive model for scientific inquiry. And we'll also run into some of the mathematics that usually gets put off to college or beyond, like the crisis in set theory, which appears here as a kind of metaphor for Supreme Court jurisprudence and baseball umpiring; recent developments in analytic number theory, which demonstrate the interplay between structure and randomness; and information theory and combinatorial designs, which help explain how a group of MIT undergrads won millions of dollars by understanding the guts of the Massachusetts state lottery.

There will be occasional gossip about mathematicians of note, and a certain amount of philosophical speculation. There will even be a proof or two. But there will be no homework, and there will be no test.

# PART I

. . . . .

# Linearity

*Includes: the Laffer curve, calculus explained in one page, the Law of Large Numbers, assorted terrorism analogies, "Everyone in America will be overweight by 2048," why South Dakota has more brain cancer than North Dakota, the ghosts of departed quantities, the habit of definition*



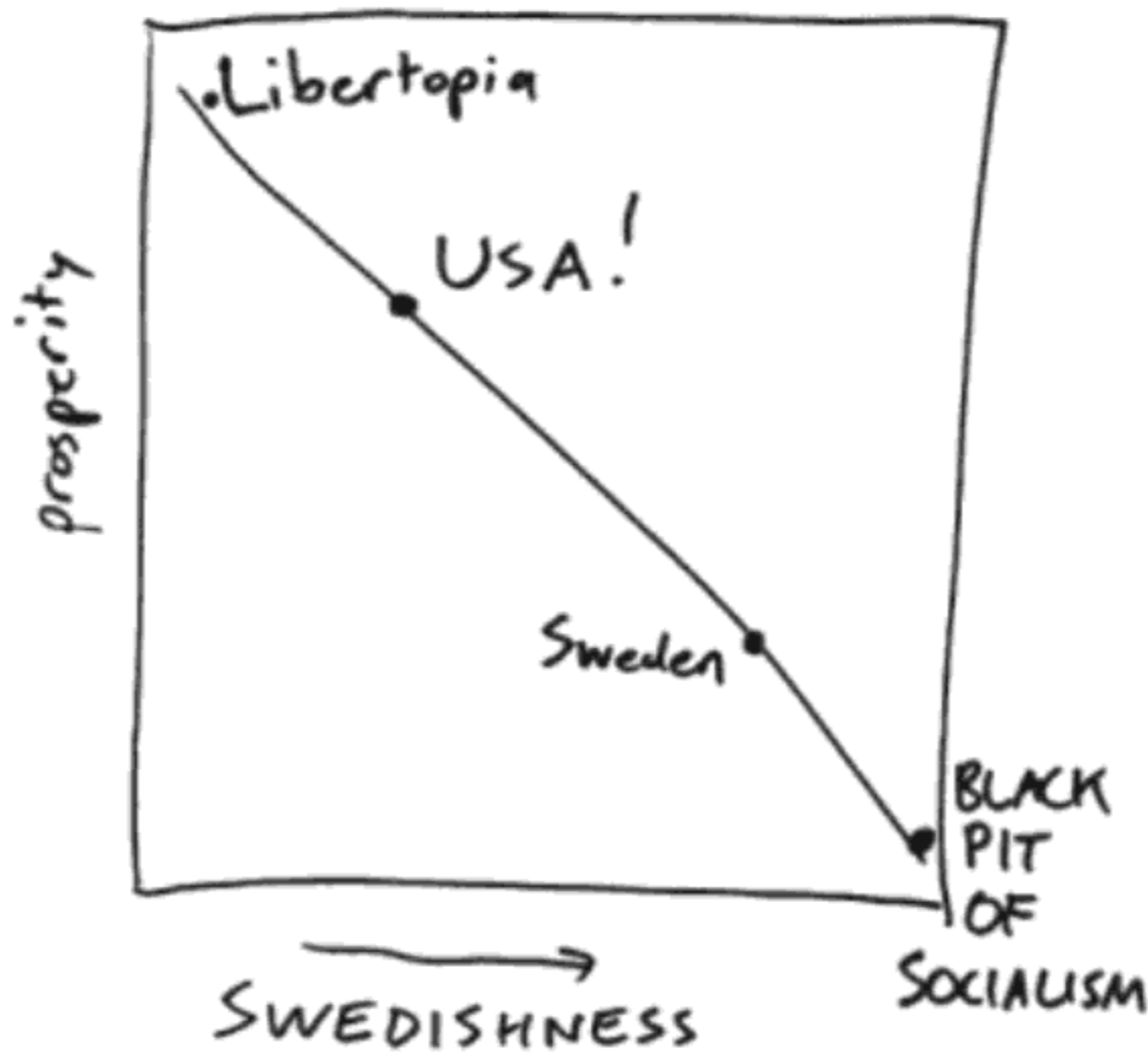
## ONE

# LESS LIKE SWEDEN

A few years ago, in the heat of the battle over the Affordable Care Act, Daniel J. Mitchell of the libertarian Cato Institute posted a blog entry with the provocative title “Why Is Obama Trying to Make America More Like Sweden when Swedes Are Trying to Be Less Like Sweden?”

Good question! When you put it that way, it does seem pretty perverse. Why, Mr. President, are we swimming against the current of history, while social welfare states around the world—even rich little Sweden!—are cutting back on expensive benefits and high taxes? “If Swedes have learned from their mistakes and are now trying to reduce the size and scope of government,” Mitchell writes, “why are American politicians determined to repeat those mistakes?”

Answering this question will require an extremely scientific chart. Here’s what the world looks like to the Cato Institute:



The x-axis represents Swedishness,\* and the y-axis is some measure of prosperity. Don't worry about exactly how we're quantifying these things. The point is just this: according to the chart, the more Swedish you are, the worse off your country is. The Swedes, no fools, have figured this out and are launching their northwestward climb toward free-market prosperity. But Obama's sliding in the wrong direction.

At the top of the facing page I've drawn the same picture from the point of view of people whose economic views are closer to President Obama's than to those of the Cato Institute.

This picture gives very different advice about how Swedish we should be. Where do we find peak prosperity? At a point more Swedish than America, but less Swedish than Sweden. If this picture is right, it makes perfect sense for Obama to beef up our welfare state while the Swedes trim theirs down.

The difference between the two pictures is the difference between linearity and nonlinearity, one of the central distinctions in mathematics. The Cato curve is a line;† the non-Cato curve, the one with the hump in

\* Here "Swedishness" refers to "quantity of social services and taxation," not to other features of Sweden such as "ready availability of herring in dozens of different sauces," a condition to which all nations should obviously aspire.

† Or a line segment, if you must. I won't make a big deal of this distinction.



the middle, is not. A line is one kind of curve, but not the only kind, and lines enjoy all kinds of special properties that curves in general may not. The highest point on a line segment—the maximum prosperity, in this example—has to be on one end or the other. That’s just how lines are. If lowering taxes is good for prosperity, then lowering taxes even more is even better. And if Sweden wants to de-Swede, so should we. Of course, an anti-Cato think tank might posit that the line slopes in the other direction, going southwest to northeast. And if that’s what the line looks like, then no amount of social spending is too much. The optimal policy is Maximum Swede.

Usually, when someone announces they’re a “nonlinear thinker” they’re about to apologize for losing something you lent them. But non-linearity is a real thing! And in this context, thinking nonlinearly is crucial, because not all curves are lines. A moment of reflection will tell you that the real curves of economics look like the second picture, not the first. They’re nonlinear. Mitchell’s reasoning is an example of *false linearity*—he’s assuming, without coming right out and saying so, that the course of prosperity is described by the line segment in the first picture, in which case Sweden stripping down its social infrastructure means we should do the same.

But as long as you believe there’s such a thing as too much welfare

state and such a thing as too little, you know the linear picture is wrong. Some principle more complicated than “More government bad, less government good” is in effect. The generals who consulted Abraham Wald faced the same kind of situation: too little armor meant planes got shot down, too much meant the planes couldn’t fly. It’s not a question of whether adding more armor is good or bad; it could be either, depending on how heavily armored the planes are to start with. If there’s an optimal answer, it’s somewhere in the middle, and deviating from it in either direction is bad news.

Nonlinear thinking means *which way you should go depends on where you already are*.

This insight isn’t new. Already in Roman times we find Horace’s famous remark “*Est modus in rebus, sunt certi denique fines, quos ultra citraque nequit consistere rectum*” (“There is a proper measure in things. There are, finally, certain boundaries short of and beyond which what is right cannot exist”). And further back still, in the *Nicomachean Ethics*, Aristotle observes that eating either too much or too little is troubling to the constitution. The optimum is somewhere in between, because the relation between eating and health isn’t linear, but curved, with bad outcomes on both ends.

## SOMETHING-DOO ECONOMICS

The irony is that economic conservatives like the folks at Cato used to understand this better than anybody. That second picture I drew up there? The extremely scientific one with the hump in the middle? I am not the first person to draw it. It’s called the *Laffer curve*, and it’s played a central role in Republican economics for almost forty years. By the middle of the Reagan administration, the curve had become such a commonplace of economic discourse that Ben Stein ad-libbed it into his famous soul-killing lecture in *Ferris Bueller’s Day Off*:

Anyone know what this is? Class? Anyone? . . . Anyone? Anyone seen this before? The Laffer curve. Anyone know what this says? It

says that at this point on the revenue curve, you will get exactly the same amount of revenue as at this point. This is very controversial. Does anyone know what Vice President Bush called this in 1980? Anyone? Something-doo economics. “Voodoo” economics.

The legend of the Laffer curve goes like this: Arthur Laffer, then an economics professor at the University of Chicago, had dinner one night in 1974 with Dick Cheney, Donald Rumsfeld, and *Wall Street Journal* editor Jude Wanniski at an upscale hotel restaurant in Washington, DC. They were tussling over President Ford’s tax plan, and eventually, as intellectuals do when the tussling gets heavy, Laffer commandeered a napkin\* and drew a picture. The picture looked like this:



The horizontal axis here is level of taxation, and the vertical axis represents the amount of revenue the government takes in from taxpayers. On the left edge of the graph, the tax rate is 0%; in that case, by definition, the government gets no tax revenue. On the right, the tax rate is 100%; whatever income you have, whether from a business you run or a salary you’re paid, goes straight into Uncle Sam’s bag.

\* Laffer disputes the napkin portion of the story, recalling that the restaurant had classy cloth napkins that he would never have vandalized with an economic doodle.



Which is empty. Because if the government vacuums up every cent of the wage you're paid to show up and teach school, or sell hardware, or middle-manage, why bother doing it? Over on the right edge of the graph, people don't work at all. Or, if they work, they do so in informal economic niches where the tax collector's hand can't reach. The government's revenue is zero once again.

In the intermediate range in the middle of the curve, where the government charges us somewhere between none of our income and all of it—in other words, in the real world—the government does take in some amount of revenue.

That means the curve recording the relationship between tax rate and government revenue cannot be a straight line. If it were, revenue would be maximized at either the left or right edge of the graph; but it's zero in both places. If the current income tax is really close to zero, so that you're on the left-hand side of the graph, then raising taxes increases the amount of money the government has available to fund services and programs, just as you might intuitively expect. But if the rate is close to 100%, raising taxes actually *decreases* government revenue. If you're to the right of the Laffer peak, and you want to decrease the deficit without cutting spending, there's a simple and politically peachy solution: lower the tax rate, and thereby increase the amount of taxes you take in. *Which way you should go depends on where you are.*

So where are we? That's where things get sticky. In 1974, the top income tax rate was 70%, and the idea that America was on the right-hand downslope of the Laffer curve held a certain appeal—especially for the few people lucky enough to pay tax at that rate, which only applied to income beyond the first \$200,000.\* And the Laffer curve had a potent advocate in Wanniski, who brought his theory into the public consciousness in a 1978 book rather self-assuredly titled *The Way the World Works*.† Wanniski was a true believer, with the right mix of zeal and political canniness to get people to listen to an idea considered fringy even by tax-cut advocates. He was untroubled by being called a nut. “Now, what does ‘nut’ mean?” he asked an interviewer. “Thomas Edison was a nut, Leibniz

\* Somewhere between a half million and a million dollars a year in today's income.

† Like I'm one to talk.

was a nut, Galileo was a nut, so forth and so on. Everybody who comes with a new idea to the conventional wisdom, comes with an idea that's so far outside the mainstream, that's considered nutty."

(Aside: it's important to point out here that people with out-of-the-mainstream ideas who compare themselves to Edison and Galileo are *never actually right*. I get letters with this kind of language at least once a month, usually from people who have "proofs" of mathematical statements that have been known for hundreds of years to be false. I can guarantee you Einstein did not go around telling people, "Look, I know this theory of general relativity sounds wacky, but that's what they said about Galileo!")

The Laffer curve, with its compact visual representation and its agreeably counterintuitive sting, turned out to be an easy sell for politicians with a preexisting hunger for tax cuts. As economist Hal Varian put it, "You can explain it to a Congressman in six minutes and he can talk about it for six months." Wanniski became an adviser first to Jack Kemp, then to Ronald Reagan, whose experiences as a wealthy movie star in the 1940s formed the template for his view of the economy four decades later. His budget director, David Stockman, recalls:

"I came into the Big Money making pictures during World War II," [Reagan] would always say. At that time the wartime income surtax hit 90 percent. "You could only make four pictures and then you were in the top bracket," he would continue. "So we all quit working after about four pictures and went off to the country." High tax rates caused less work. Low tax rates caused more. His experience proved it.

These days it's hard to find a reputable economist who thinks we're on the downslope of the Laffer curve. Maybe that's not surprising, considering top incomes are currently taxed at just 35%, a rate that would have seemed absurdly low for most of the twentieth century. But even in Reagan's day, we were probably on the left-hand side of the curve. Greg Mankiw, an economist at Harvard and a Republican who chaired the Council of Economic Advisors under the second President Bush, writes in his microeconomics textbook:

Subsequent history failed to confirm Laffer's conjecture that lower tax rates would raise tax revenue. When Reagan cut taxes after he was elected, the result was less tax revenue, not more. Revenue from personal income taxes (per person, adjusted for inflation) fell by 9 percent from 1980 to 1984, even though average income (per person, adjusted for inflation) grew by 4 percent over this period. Yet once the policy was in place, it was hard to reverse.

Some sympathy for the supply-siders is now in order. First of all, maximizing government revenue needn't be the goal of tax policy. Milton Friedman, whom we last met during World War II doing classified military work for the Statistical Research Group, went on to become a Nobel-winning economist and adviser to presidents, and a powerful advocate for low taxes and libertarian philosophy. Friedman's famous slogan on taxation is "I am in favor of cutting taxes under any circumstances and for any excuse, for any reason, whenever it's possible." He didn't think we should be aiming for the top of the Laffer curve, where government tax revenue is as high as it can be. For Friedman, money obtained by the government would eventually be money spent by the government, and that money, he felt, was more often spent badly than well.

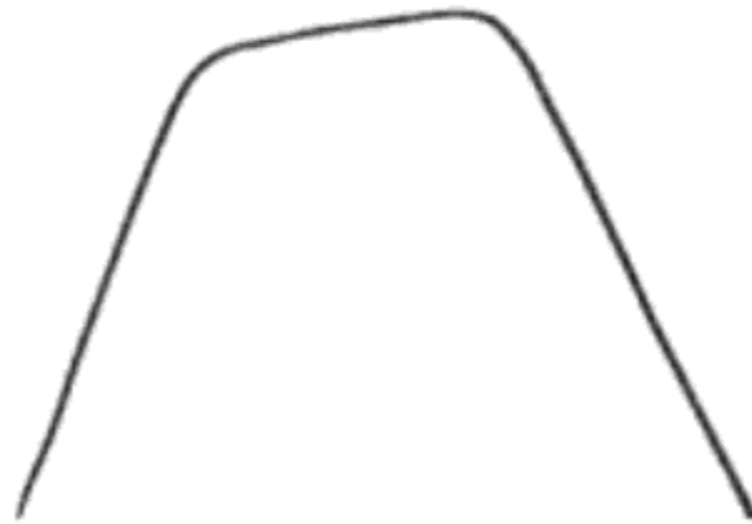
More moderate supply-side thinkers, like Mankiw, argue that lower taxes can increase the motivation to work hard and launch businesses, leading eventually to a bigger, stronger economy, even if the immediate effect of the tax cut is decreased government revenue and bigger deficits. An economist with more redistributionist sympathies would observe that this cuts both ways; maybe the government's diminished ability to spend means it constructs less infrastructure, regulates fraud less stringently, and generally does less of the work that enables free enterprise to thrive.

Mankiw also points out that the very richest people—the ones who'd been paying 70% on the top tranche of their income—*did* contribute more tax revenue after Reagan's tax cuts.\* That leads to the somewhat vexing possibility that the way to maximize government revenue is to

\* Whether the increased tax receipts are because the rich started working harder once less encumbered by income tax, as supply-side theory predicts, is more difficult to say for certain.

jack up taxes on the middle class, who have no choice but to keep on working, while slashing rates on the rich; those guys have enough stock-piled wealth to make credible threats to withhold or offshore their economic activity, should their government charge them a rate they deem too high. If that story's right, a lot of liberals will uncomfortably climb in the boat with Milton Friedman: maybe maximizing tax revenue isn't so great after all.

Mankiw's final assessment is a rather polite "Laffer's argument is not completely without merit." I would give Laffer more credit than that! His drawing made the fundamental and incontrovertible mathematical point that the relationship between taxation and revenue is necessarily nonlinear. It doesn't, of course, have to be a single smooth hill like the one Laffer sketched; it could look like a trapezoid



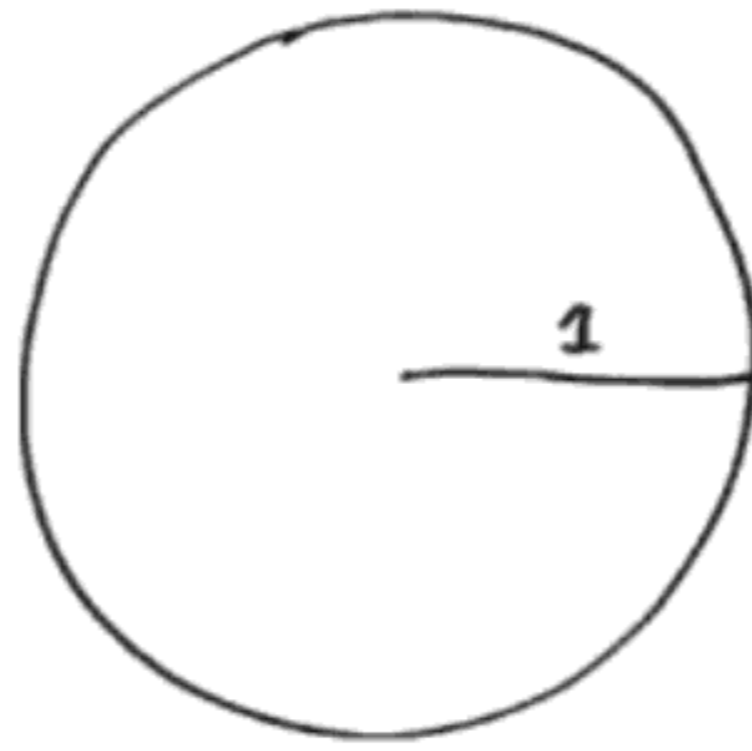
or a dromedary's back



or a wildly oscillating free-for-all\*



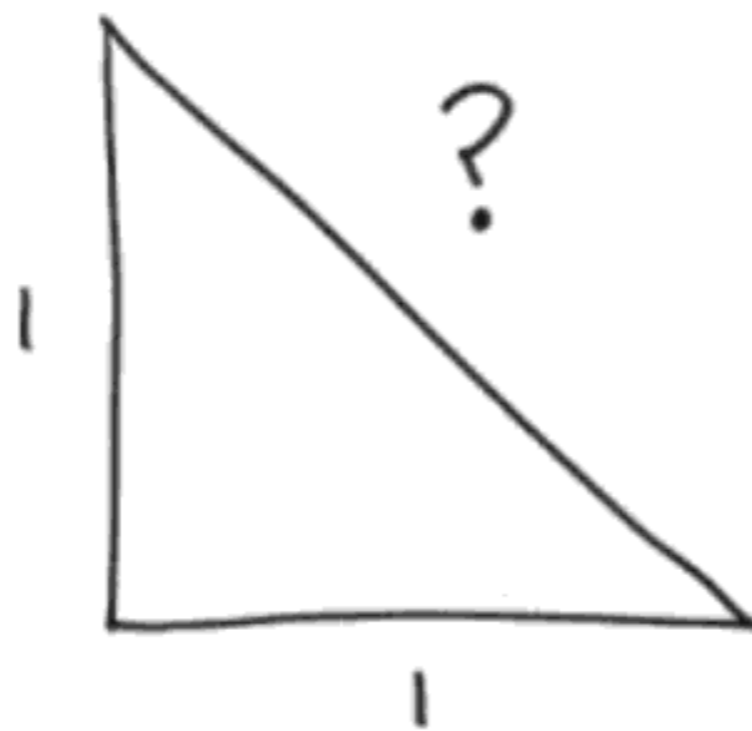
\* Or, more likely still, it might not be a single curve at all, as Martin Gardner illustrated by means of the snarly "Neo-Laffer curve" in his acid assessment of supply-side theory, "The Laffer Curve."



the area is  $\pi$ . But two thousand years ago this was a vexing open question, important enough to draw the attention of Archimedes.

Why was it so hard? For one thing, the Greeks didn't really think of  $\pi$  as a number, as we do. The numbers they understood were whole numbers, numbers that counted things: 1, 2, 3, 4 . . . But the first great success of Greek geometry—the Pythagorean Theorem\*—turned out to be the ruin of their number system.

Here's a picture:



The Pythagorean Theorem tells you that the square of the *hypotenuse*—the side drawn diagonally here, the one that doesn't touch the right angle—is the sum of the squares of the other two sides, or *legs*. In this picture, that says the square of the hypotenuse is  $1^2 + 1^2 = 1 + 1 = 2$ . In particular, the hypotenuse is longer than 1 and shorter than 2 (as you can check with your eyeballs, no theorem required). That the length is not a

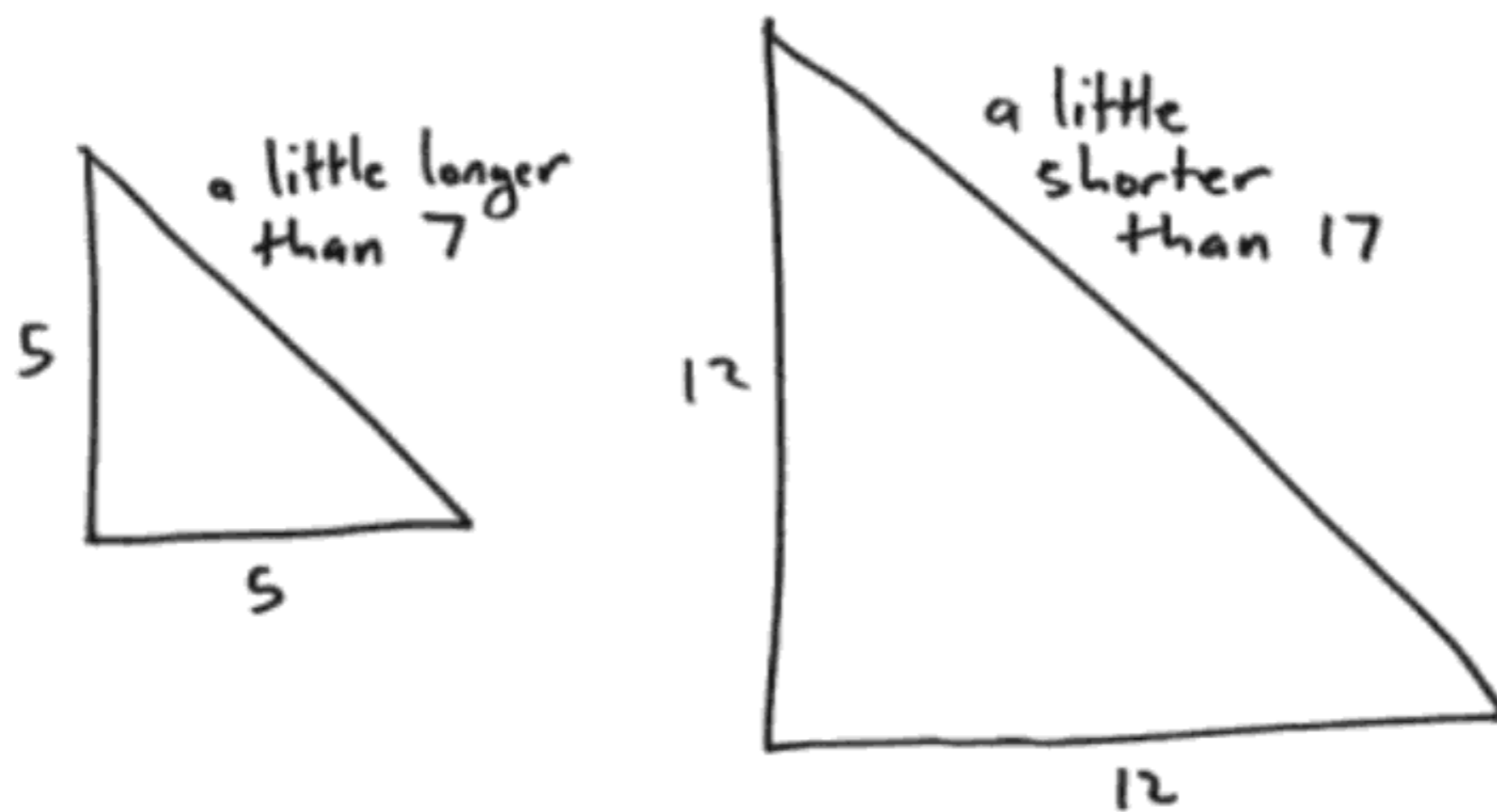
\* By the way, we don't know who first proved the Pythagorean Theorem, but scholars are almost certain it was not Pythagoras himself. In fact, beyond the bare fact, attested by contemporaries, that a learned man by that name lived and gained fame in the sixth century BCE, we know almost nothing about Pythagoras. The main accounts of his life and work date from almost eight hundred years after his death. By that time, Pythagoras the real person had been completely replaced by Pythagoras the myth, a kind of summing up in an individual of the philosophy of the scholars who called themselves Pythagoreans.

whole number was not, in itself, a problem for the Greeks. Maybe we just measured everything in the wrong units. If we choose our unit of length to make the legs 5 units long, you can check with a ruler that the hypotenuse is just about 7 units long. Just about—but a bit too long. For the square of the hypotenuse is

$$5^2 + 5^2 = 25 + 25 = 50$$

and if the hypotenuse were 7, its square would be  $7 \times 7 = 49$ .

Or if you make the legs 12 units long, the hypotenuse is almost exactly 17 units, but is tantalizingly too short, because  $12^2 + 12^2$  is 288, a smidgen less than  $17^2$ , which is 289.



And at some point around the fifth century BCE, a member of the Pythagorean school made a shocking discovery: there was *no way* to measure the isosceles right triangle so that the length of each side was a whole number. Modern people would say “the square root of 2 is irrational”—that is, it is not the ratio of any two whole numbers. But the Pythagoreans would not have said that. How could they? Their notion of quantity was built on the idea of proportions between whole numbers. To them, the length of that hypotenuse had been revealed to be *not a number at all*.

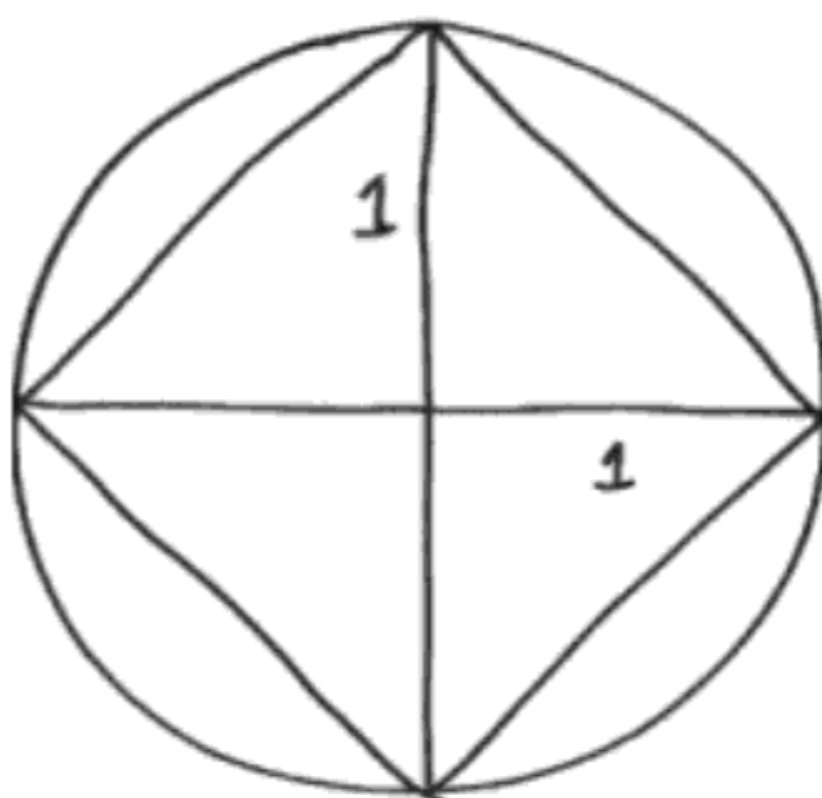
This caused a fuss. The Pythagoreans, you have to remember, were extremely weird. Their philosophy was a chunky stew of things we’d now call mathematics, things we’d now call religion, and things we’d now call mental illness. They believed that odd numbers were good and even numbers evil; that a planet identical to our own, the Antichthon, lay on

the other side of the sun; and that it was wrong to eat beans, by some accounts because they were the repository of dead people's souls. Pythagoras himself was said to have had the ability to talk to cattle (he told them not to eat beans) and to have been one of the very few ancient Greeks to wear pants.

The mathematics of the Pythagoreans was inseparably bound up with their ideology. The story (probably not really true, but it gives the right impression of the Pythagorean style) is that the Pythagorean who discovered the irrationality of the square root of 2 was a man named Hippasus, whose reward for proving such a nauseating theorem was to be tossed into the sea by his colleagues, to his death.

But you can't drown a theorem. The Pythagoreans' successors, like Euclid and Archimedes, understood that you had to roll up your sleeves and measure things, even if this brought you outside the pleasant walled garden of the whole numbers. No one knew whether the area of a circle could be expressed using whole numbers alone.\* But wheels must be built and silos filled;† so the measurement must be done.

The original idea comes from Eudoxus of Cnidus; Euclid included it as book 12 of the elements. But it was Archimedes who really brought the project to its full fruition. Today we call his approach *the method of exhaustion*. And it starts like this.

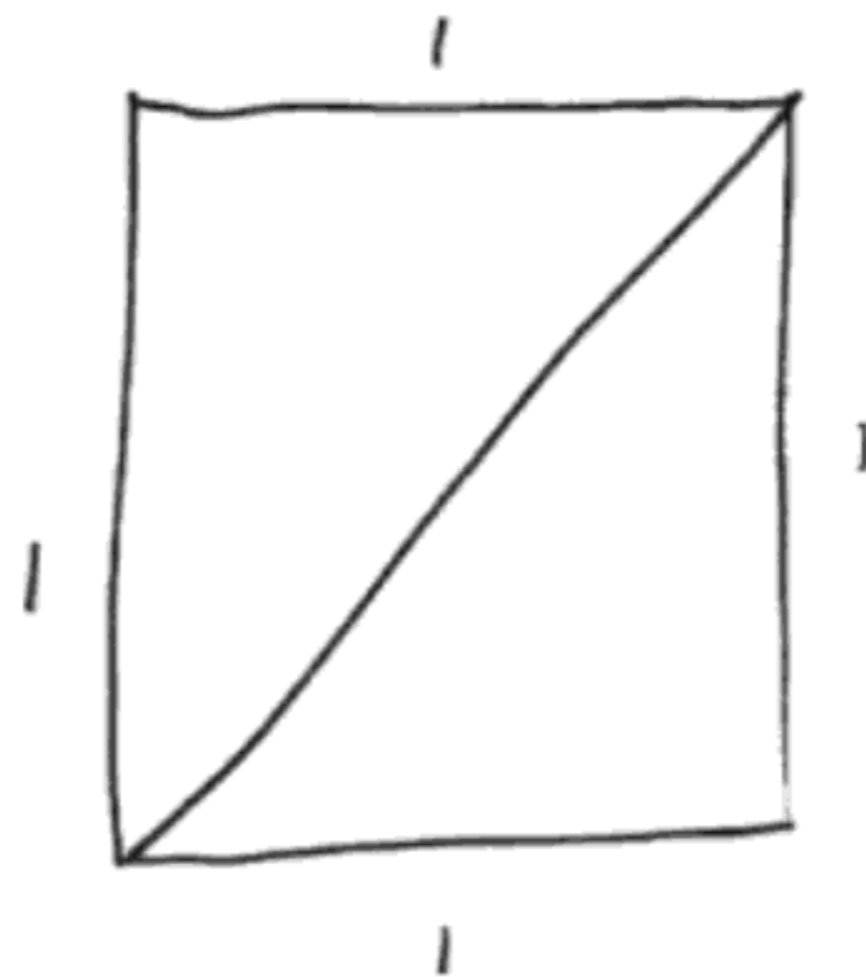


\* It can't, in fact, but nobody figured out how to prove this until the eighteenth century.

† Actually, silos weren't round until the early twentieth century, when a University of Wisconsin professor, F.H. King, invented the now-ubiquitous cylindrical design in order to solve the problem of spoilage in the corners.

The square in the picture is called the *inscribed square*; each of its corners just touches the circle, but it doesn't extend beyond the circle's boundary. Why do this? Because circles are mysterious and intimidating, and squares are easy. If you have before you a square whose side has length  $X$ , its area is  $X$  times  $X$ —indeed, that's why we call the operation of multiplying a number by itself squaring! A basic rule of mathematical life: if the universe hands you a hard problem, try to solve an easier one instead, and hope the simple version is close enough to the original problem that the universe doesn't object.

The inscribed square breaks up into four triangles, each of which is none other than the isosceles triangle we just drew.\* So the square's area is four times the area of the triangle. That triangle, in turn, is what you get when you take a  $1 \times 1$  square and cut it diagonally in half like a tuna fish sandwich.



The area of the tuna fish sandwich is  $1 \times 1 = 1$ , so the area of each triangular half-sandwich is  $1/2$ , and the area of the inscribed square is  $4 \times 1/2$ , or 2.

By the way, suppose you *don't* know the Pythagorean Theorem. Guess what—you do now! Or at least you know what it has to say about this particular right triangle. Because the right triangle that makes up the lower half of the tuna fish sandwich is exactly the same as the one that is the northwest quarter of the inscribed square. And its hypotenuse is the

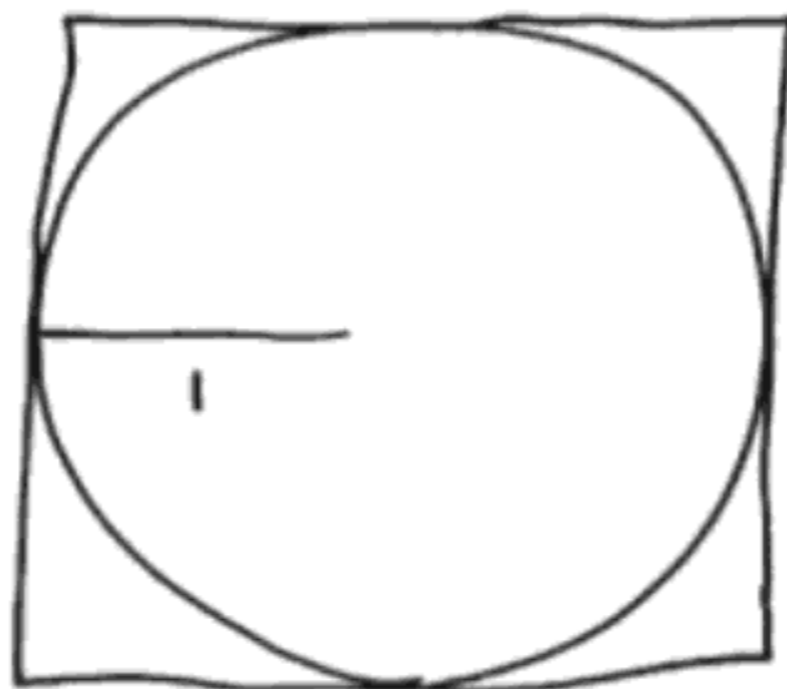
\* Rather, each of the four pieces can be obtained from the original isosceles right triangle by sliding and rotating it around the plane; we take as given that these manipulations don't change a figure's area.



inscribed square's side. So when you square the hypotenuse, you get the area of the inscribed square, which is 2. That is, the hypotenuse is that number which, when squared, yields 2; or, in the usual more concise lingo, the square root of 2.

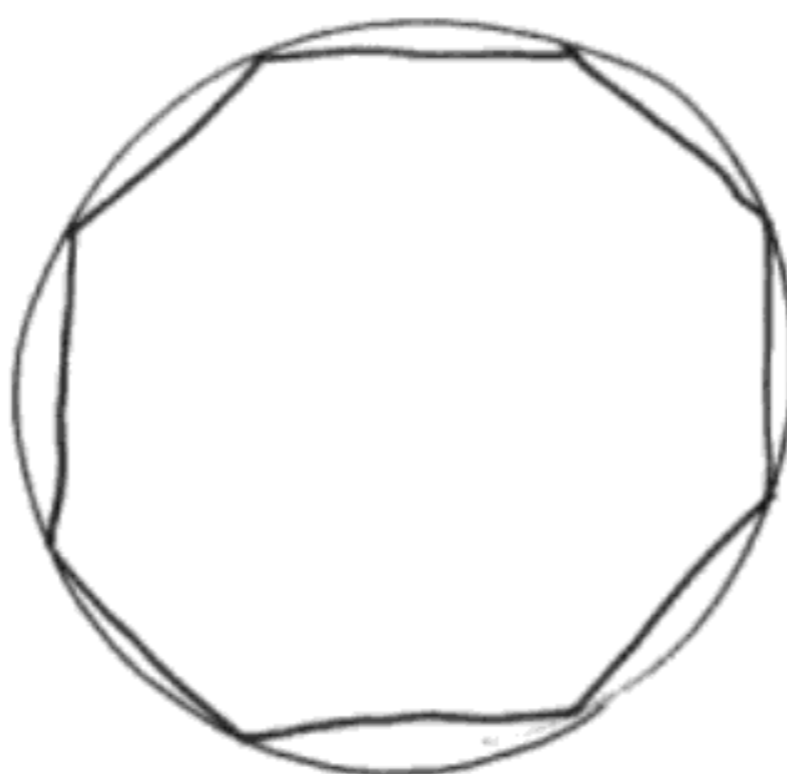
The inscribed square is entirely contained within the circle. If its area is 2, the area of the circle must be *at least* 2.

Now we draw another square.



This one is called the *circumscribed* square; it, too, touches the circle at just four points. But this square contains the circle. Its sides have length 2, so its area is 4; and so we know the area of the circle is at most 4.

To have shown that pi is between 2 and 4 is perhaps not so impressive. But Archimedes is just getting started. Take the four corners of your inscribed square and mark new points on the circle halfway between each adjacent pair of corners. Now you've got eight equally spaced points, and when you connect those, you get an inscribed octagon, or, in technical language, a "stop sign":



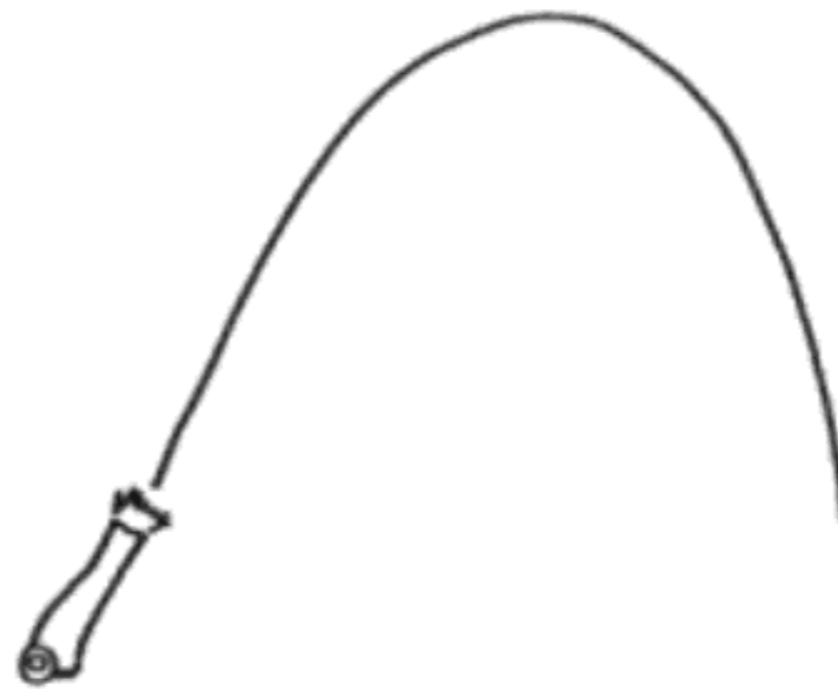
Computing the area of the inscribed octagon is a bit harder, and I'll spare you the trigonometry. The important thing is that it's about straight

Until, zooming in, and zooming in, what you see is pretty much indistinguishable from a line. An ant on the circle, aware only of his own tiny immediate surroundings, would think he was on a straight line, just as a person on the surface of the earth (unless she is clever enough to watch objects crest the horizon as they approach from afar) feels like she's standing on a plane.

## THE PAGE WHERE I TEACH YOU CALCULUS

I will now teach you calculus. Ready? The idea, for which we have Isaac Newton to thank, is that there's nothing special about a perfect circle. *Every* smooth curve, when you zoom in enough, looks just like a line. Doesn't matter how winding or snarled it is—just that it doesn't have any sharp corners.

When you fire a missile, its path looks like this:



The missile goes up, then down, in a parabolic arc. Gravity makes all motion curve toward the earth; that's among the fundamental facts of our physical life. But if we zoom in on a very short segment, the curve starts to look like this:



And then like this:



Just like the circle, the missile's path looks to the naked eye like a straight line, progressing upward at an angle. The deviation from straightness caused by gravity is too small to see—but it's still there, of course. Zooming in to an even smaller region of the curve makes the curve even more like a straight line. Closer and straighter, closer and straighter . . .

Now here's the conceptual leap. Newton said, look, let's go all the way. Reduce your field of view until it's *infinitesimal*—so small that it's smaller than any size you can name, but not zero. You're studying the missile's arc, not over a very short time interval, but at a single moment. What was *almost* a line becomes *exactly* a line. And the slope of this line is what Newton called the *fluxion*, and what we'd now call the *derivative*.

That's a kind of jump Archimedes wasn't willing to make. He understood that polygons with shorter sides got closer and closer to the circle—but he would never have said that the circle actually *was* a polygon with infinitely many infinitely short sides.

Some of Newton's contemporaries, too, were reluctant to go along for the ride. The most famous objector was George Berkeley, who denounced Newton's infinitesimals in a tone of high mockery sadly absent from current mathematical literature: "And what are these fluxions? The velocities of evanescent increments. And what are these same evanescent increments? They are neither finite quantities, nor quantities infinitely small, nor yet nothing. May we not call them the ghosts of departed quantities?"

And yet calculus *works*. If you swing a rock in a loop around your head and suddenly release it, it'll shoot off along a linear trajectory at constant

speed,\* exactly in the direction that calculus says the rock is moving at the precise moment you let go. That's yet another Newtonian insight; objects in motion tend to proceed in a straight-line path, unless some other force intercedes to nudge the object one way or the other. That's one reason linear thinking comes so naturally to us: our intuition about time and motion is formed by the phenomena we observe in the world. Even before Newton codified his laws, something in us knew that things like to move in straight lines, unless given a reason to do otherwise.

## EVANESCENT INCREMENTS AND UNNECESSARY PERPLEXITIES

Newton's critics had a point; his construction of the derivative didn't amount to what we'd call rigorous mathematics nowadays. The problem is the notion of the infinitely small, which was a slightly embarrassing sticking point for mathematicians for thousands of years. The trouble started with Zeno, a fifth-century-BCE Greek philosopher of the Eleatic school who specialized in asking innocent-seeming questions about the physical world that inevitably blossomed into huge philosophical brouhahas.

His most famous paradox goes like this. I decide to walk to the ice cream store. Now certainly I can't get to the ice cream store until I've gone halfway there. And once I've gone halfway, I can't get to the store until I've gone half the distance that remains. Having done so, I still have to cover half the remaining distance. And so on, and so on. I may get closer and closer to the ice cream store—but no matter how many steps of this process I undergo, I never actually *reach* the ice cream store. I am always some tiny but nonzero distance away from my two scoops with jimmies. Thus, Zeno concludes, to walk to the ice cream store is impossible. The argument works just as well for any destination: it's equally impossible to walk across the street, or to take a single step, or to wave your hand. All motion is ruled out.

\* Apart from the effects of gravity, air resistance, etc., etc. But on a short timescale, the linear approximation is good enough.

Diogenes the Cynic was said to have refuted Zeno's argument by standing up and walking across the room. Which is a pretty good argument that motion is actually possible; so something must be wrong with Zeno's argument. But where's the mistake?

Break down the trip to the store numerically. First you go halfway. Then you go half of the remaining distance, which is  $1/4$  of the total distance, and you've got  $1/4$  left to go. So half of what's left is  $1/8$ , then  $1/16$ , then  $1/32$ . Your progress toward the store looks like this:

$$1/2 + 1/4 + 1/8 + 1/16 + 1/32 + \dots$$

If you add up ten terms of this sequence you get about 0.999. If you add up twenty terms it's more like 0.999999. In other words, you are getting really, really, really close to the store. But no matter how many terms you add, you never get to 1.

Zeno's paradox is much like another conundrum: is the repeating decimal 0.99999. . . . . equal to 1?

I have seen people come nearly to blows over this question.\* It's hotly disputed on websites ranging from World of Warcraft fan pages to Ayn Rand forums. Our natural feeling about Zeno is "of course you eventually get your ice cream." But in this case, intuition points the other way. Most people, if you press them, say 0.9999 . . . doesn't equal 1. It doesn't *look* like 1, that's for sure. It looks smaller. But not much smaller! Like Zeno's hungry ice cream lover, it gets closer and closer to its goal, but never, it seems, quite makes it there.

And yet, math teachers everywhere, myself included, will tell them, "No, it's 1."

How do I convince someone to come over to my side? One good trick is to argue as follows. Everyone knows that

$$0.33333. . . . . = 1/3.$$

Multiply both sides by 3 and you'll see

\* Admittedly, these particular people were teenagers at a summer math camp.

$$0.99999 \dots = 3/3 = 1.$$

If that doesn't sway you, try multiplying  $0.99999 \dots$  by 10, which is just a matter of moving the decimal point one spot to the right.

$$10 \times (0.99999 \dots) = 9.99999 \dots$$

Now subtract the vexing decimal from both sides:

$$\begin{aligned} 10 \times (0.99999 \dots) - 1 \times (0.99999 \dots) \\ = 9.99999 \dots - 0.99999 \dots \end{aligned}$$

The left-hand side of the equation is just  $9 \times (0.99999 \dots)$ , because 10 times something minus that something is 9 times the aforementioned thing. And over on the right-hand side, we have managed to cancel out the terrible infinite decimal, and are left with a simple 9. So we end up with

$$9 \times (0.99999 \dots) = 9.$$

If 9 times something is 9, that something just has to be 1—doesn't it?

These arguments are often enough to win people over. But let's be honest: they lack something. They don't really address the anxious uncertainty induced by the claim  $0.99999 \dots = 1$ ; instead, they represent a kind of algebraic intimidation. "You believe that  $1/3$  is 0.3 repeating—don't you? *Don't you?*"

Or worse: maybe you bought my argument based on multiplication by 10. But how about this one? What is

$$1 + 2 + 4 + 8 + 16 + \dots ?$$

Here the " $\dots$ " means "carry on the sum forever, adding twice as much each time." Surely such a sum must be infinite! But an argument much like the apparently correct one concerning  $0.9999 \dots$  seems to suggest otherwise. Multiply the sum above by 2 and you get

the classroom definitions, worked out a fair chunk of the theory by himself, referring to his infinitesimal quantities as “Brian numbers.”

Brian was not actually the first to get there. There’s a whole field of mathematics that specializes in contemplating numbers of this kind, called *nonstandard analysis*. The theory, developed by Abraham Robinson in the mid-twentieth century, finally made sense of the “evanescent increments” that Berkeley found so ridiculous. The price you have to pay (or, from another point of view, the reward you get to reap) is a profusion of novel kinds of numbers; not only infinitely small ones, but infinitely large ones, a huge spray of them in all shapes and sizes.\*

As it happened, Brian was in luck—I had a colleague at Princeton, Edward Nelson, who was an expert in nonstandard analysis. I set up a meeting for the two of them so Brian could learn more about it. The meeting, Ed told me later, didn’t go well. As soon as Ed made it clear that infinitesimal quantities were not in fact going to be called *Brian numbers*, Brian lost all interest.

(Moral lesson: people who go into mathematics for fame and glory don’t stay in mathematics for long.)

But we’re no closer to settling our dispute. What is  $0.999\dots$ , *really*? Is it 1? Or is it some number infinitesimally less than 1, a crazy kind of number that hadn’t even been discovered a hundred years ago?

The right answer is to unask the question. What is  $0.999\dots$ , *really*? It appears to refer to a kind of sum:

$$.9 + .09 + .009 + .0009 + \dots$$

But what does that mean? That pesky ellipsis is the real problem. There can be no controversy about what it means to add up two, or three, or a hundred numbers. This is just mathematical notation for a physical process we understand very well: take a hundred heaps of stuff, mush them together, see how much you have. But infinitely many? That’s a different story. In the real world, you can never have infinitely many heaps.

\* The *surreal numbers*, developed by John Conway, are especially charming and weird examples, as their name suggests; they are strange hybrids between numbers and games of strategy and their depths have not yet been fully explored. The book *Winning Ways*, by Berlekamp, Conway, and Guy, is a good place to learn about these exotic numbers, and lots more about the rich mathematics of game playing besides.

What's the numerical value of an infinite sum? It doesn't have one—*until we give it one*. That was the great innovation of Augustin-Louis Cauchy, who introduced the notion of *limit* into calculus in the 1820s.\*

The British number theorist G. H. Hardy, in his 1949 book *Divergent Series*, explains it best:

It does not occur to a modern mathematician that a collection of mathematical symbols should have a “meaning” until one has been assigned to it by definition. It was not a triviality even to the greatest mathematicians of the eighteenth century. They had not the habit of definition: it was not natural to them to say, in so many words, “by  $X$  we *mean*  $Y$ .” . . . It is broadly true to say that mathematicians before Cauchy asked not, “How shall we *define*  $1 - 1 + 1 - 1 + \dots$ ” but “What *is*  $1 - 1 + 1 - 1 + \dots$ ?” and that this habit of mind led them into unnecessary perplexities and controversies which were often really verbal.

This is not just loosey-goosey mathematical relativism. Just because we *can* assign whatever meaning we like to a string of mathematical symbols doesn't mean we should. In math, as in life, there are good choices and there are bad ones. In the mathematical context, the good choices are the ones that settle unnecessary perplexities without creating new ones.

The sum  $.9 + .09 + .009 + \dots$  gets closer and closer to 1 the more terms you add. And it never gets any farther away. No matter how tight a cordon we draw around the number 1, the sum will eventually, after some finite number of steps, penetrate it, and never leave. Under those circumstances, Cauchy said, we should simply *define* the value of the infinite sum to be 1. And then he worked very hard to prove that committing oneself to his definition didn't cause horrible contradictions to pop up elsewhere. By the time this labor was done, he'd constructed a framework that made Newton's calculus completely rigorous. When we say a curve looks locally like a straight line at a certain angle, we now

\* Like all mathematical breakthroughs, Cauchy's theory of limits had precursors—for instance, Cauchy's definition was very much in the spirit of d'Alembert's bounds for the error terms of binomial series. But there's no question that Cauchy was the turning point; after him, analysis is modern.



mean more or less this: as you zoom in tighter and tighter, the curve resembles the given line more and more closely. In Cauchy's formulation, there's no need to mention infinitely small numbers, or anything else that would make a skeptic blanch.

Of course there is a cost. The reason the  $0.999\dots$  problem is difficult is that it brings our intuitions into conflict. We would like the sum of an infinite series to play nicely with arithmetic manipulations like the ones we carried out on the previous pages, and this seems to demand that the sum equal 1. On the other hand, we would like each number to be represented by a unique string of decimal digits, which conflicts with the claim that the same number can be called either 1 or  $0.999\dots$ , as we like. We can't hold on to both of these desires at once; one must be discarded. In Cauchy's approach, which has amply proved its worth in the two centuries since he invented it, it's the uniqueness of the decimal expansion that goes out the window. We're untroubled by the fact that the English language sometimes uses two different strings of letters (i.e., two words) to refer synonymously to the same thing in the world; in the same way, it's not so bad that two different strings of digits can refer to the same number.

As for Grandi's  $1 - 1 + 1 - 1 + \dots$ , it is one of the series outside the reach of Cauchy's theory: that is, one of the *divergent series* that formed the subject of Hardy's book. The Norwegian mathematician Niels Henrik Abel, an early fan of Cauchy's approach, wrote in 1828, "Divergent series are the invention of the devil, and it is shameful to base on them any demonstration whatsoever." Hardy's view, which is our view today, is more forgiving; there are some divergent series to which we ought to assign values and some to which we ought not, and some to which we ought or ought not depending on the context in which the series arises. Modern mathematicians would say that if we are to assign the Grandi series a value, it should be  $1/2$ , because, as it turns out, all interesting theories of infinite sums either give it the value  $1/2$  or decline, like Cauchy's theory, to give it any value at all.<sup>†</sup>

To write Cauchy's definitions down precisely takes a bit more work.

\* Ironic, considering Grandi's original theological application of his divergent series!

† In the famous words of Lindsay Lohan, "The limit does not exist!"

This was especially true for Cauchy himself, who had not quite phrased the ideas in their clean, modern form.\* (In mathematics, you very seldom get the clearest account of an idea from the person who invented it.) Cauchy was an unwavering conservative and a royalist, but in his mathematics he was proudly revolutionary and a scourge to academic authority. Once he understood how to do things without the dangerous infinitesimals, he unilaterally rewrote his syllabus at the *École Polytechnique* to reflect his new ideas. This enraged everyone around him: his mystified students, who had signed up for freshman calculus, not a seminar on cutting-edge pure mathematics; his colleagues, who felt that the engineering students at the *École* had no need for Cauchy's level of rigor; and the administrators, whose commands to stick to the official course outline he completely ignored. The *École* imposed a new curriculum from above that emphasized the traditional infinitesimal approach to calculus, and placed note takers in Cauchy's classroom to make sure he complied. Cauchy did not comply. Cauchy was not interested in the needs of engineers. Cauchy was interested in the truth.

It's hard to defend Cauchy's stance on pedagogical grounds. But I'm sympathetic with him anyway. One of the great joys of mathematics is the incontrovertible feeling that you've understood something the right way, all the way down to the bottom; it's a feeling I haven't experienced in any other sphere of mental life. And when you know how to do something the right way, it's hard—for some stubborn people, impossible—to make yourself explain it the wrong way.

\* If you've ever taken a math course that uses epsilons and deltas, you've seen the descendants of Cauchy's formal definitions.

## THREE

# EVERYONE IS OBESE

**T**he stand-up comic Eugene Mirman tells this joke about statistics. He says he likes to tell people, “I read that 100% of Americans were Asian.”

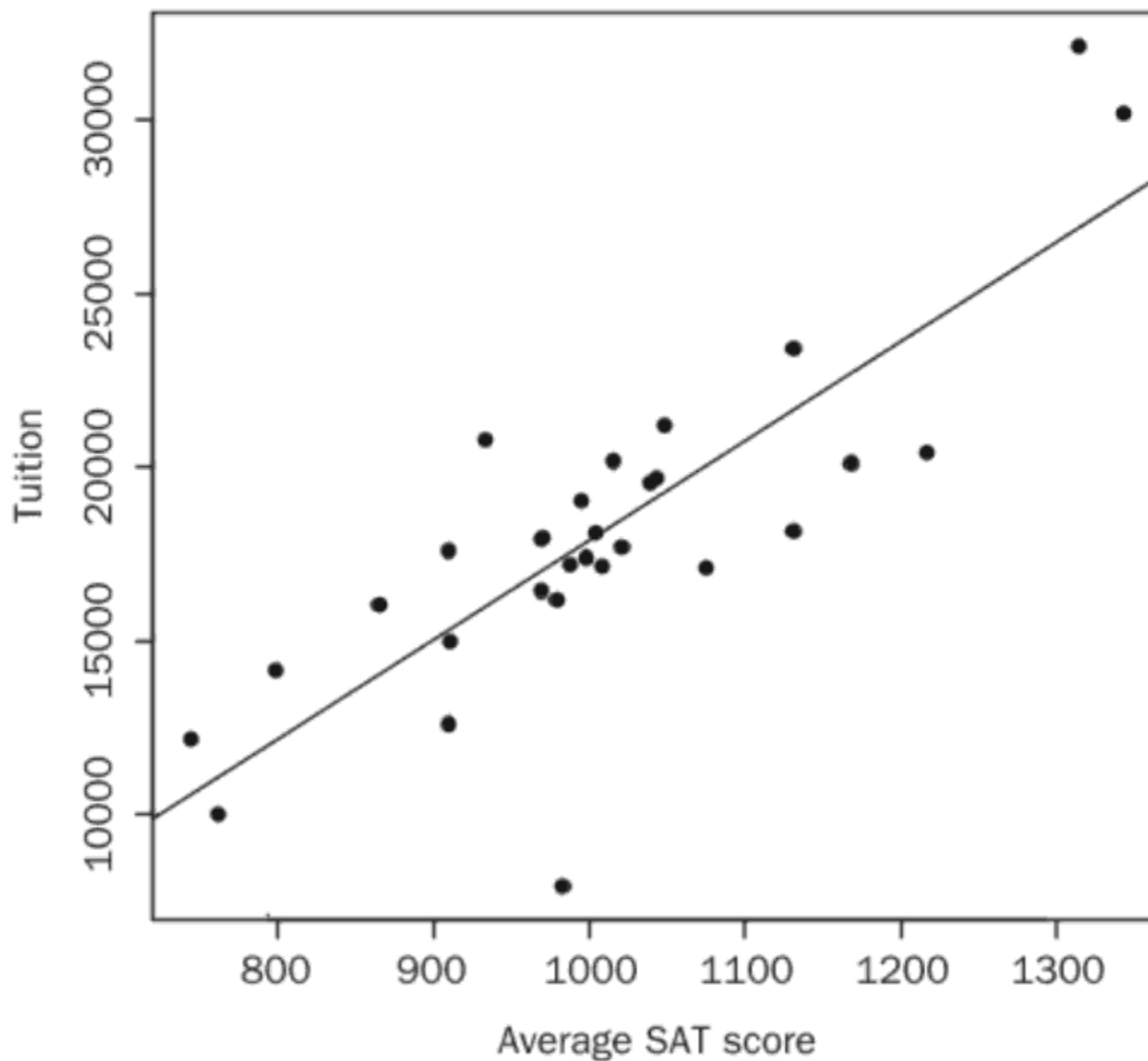
“But Eugene,” his confused companion protests, “*you’re* not Asian.”

And the punch line, delivered with magnificent self-assurance: “I read that I was!”

I thought of Mirman’s joke when I encountered a paper in the journal *Obesity* whose title posed the discomfiting question: “Will all Americans become overweight or obese?” As if the rhetorical question weren’t enough, the article supplies an answer: “Yes—by 2048.”

In 2048 I’ll be seventy-seven years old, and I hope not to be overweight. But I read I would be!

The *Obesity* paper got plenty of press, as you might imagine. ABC News warned of an “obesity apocalypse.” The *Long Beach Press-Telegram* went with the simple headline “We’re Getting Fatter.” The study’s results resonated with the latest manifestation of the fevered, ever-shifting anxiety with which Americans have always contemplated our national moral status. Before I was born, boys grew long hair and thus we were bound to get whipped by the Communists. When I was a kid, we played



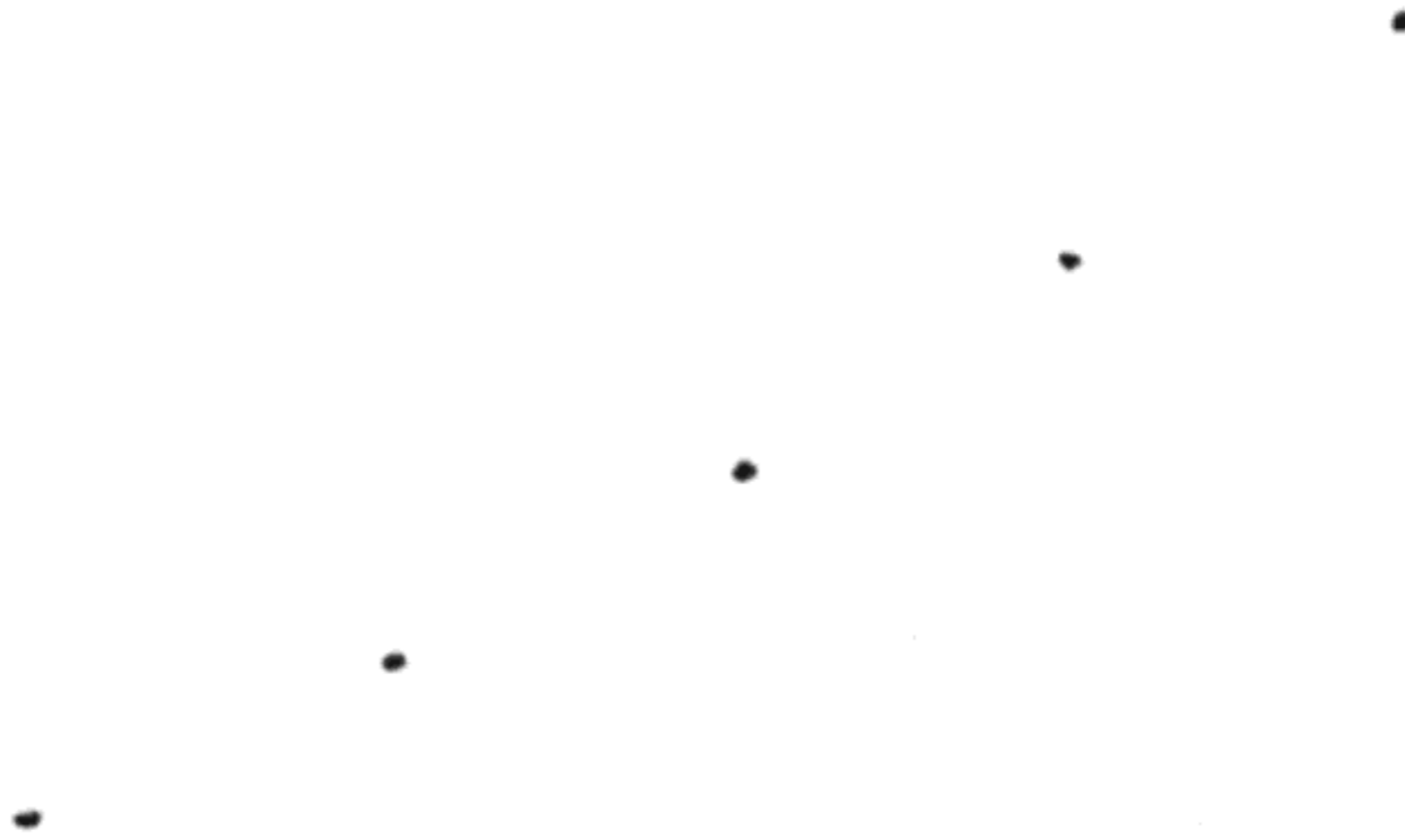
extra \$28 in tuition. If you can raise the average SAT score of your incoming first-years by 50 points on average, you can charge \$1,400 more in tuition. (Or, from the parents' point of view, your kid improving 100 points is going to cost you an extra \$2,800 a year. That test-prep course was more expensive than you thought!)

Linear regression is a marvelous tool, versatile, scalable, and as easy to execute as clicking a button on your spreadsheet. You can use it for data sets involving two variables, like the ones I've drawn here, but it works just as well for three variables, or a thousand. Whenever you want to understand which variables drive which other variables, and in which direction, it's the first thing you reach for. And it works on any data set at all.

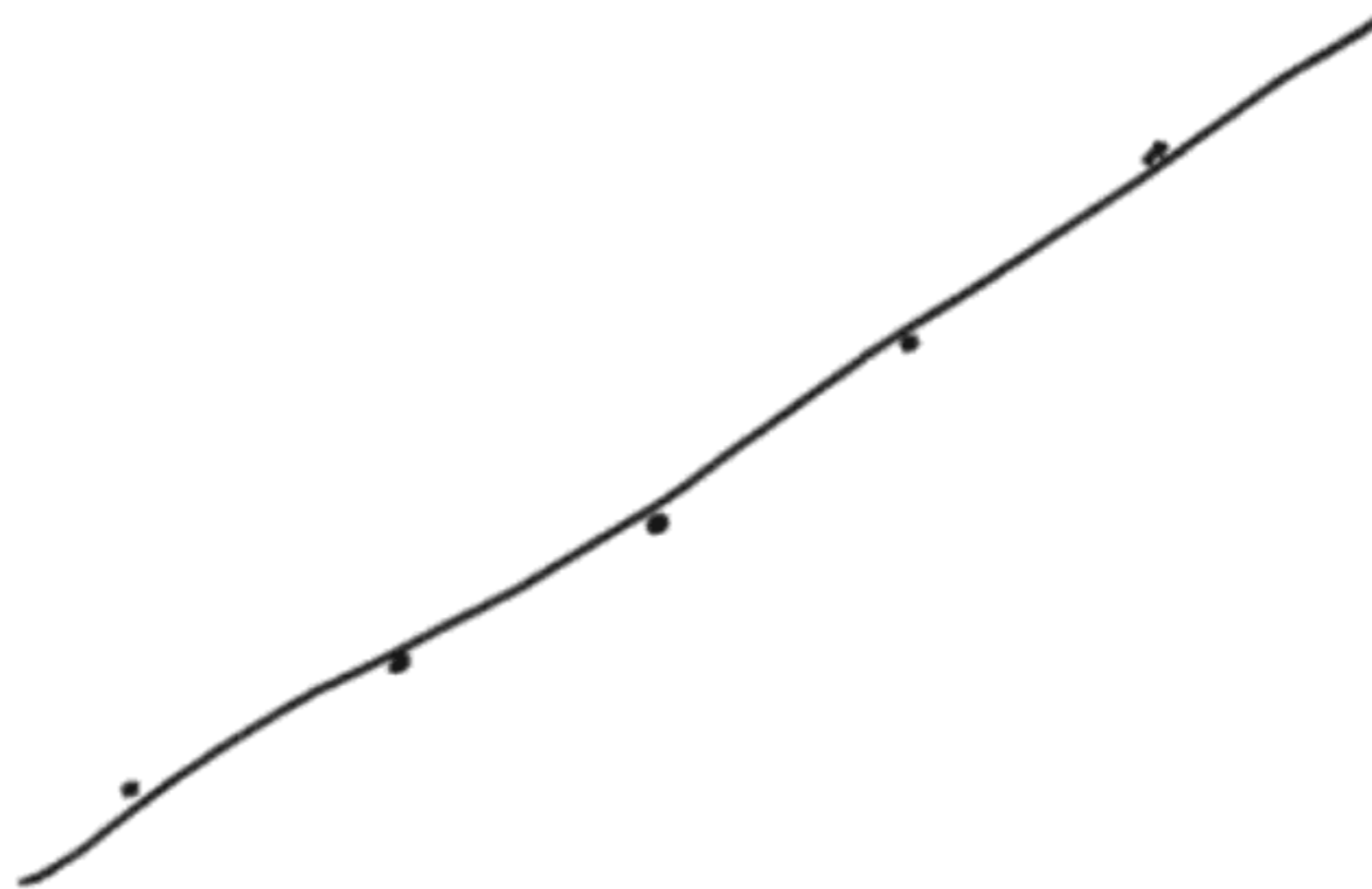
That's a weakness as well as a strength. You can do linear regression without thinking about whether the phenomenon you're modeling is actually close to linear. *But you shouldn't.* I said linear regression was like a screwdriver, and that's true; but in another sense, it's more like a table saw. If you use it without paying careful attention to what you're doing, the results can be gruesome.

Take, for instance, the missile we fired off in the last chapter. Perhaps you were not the one who fired the missile at all. Perhaps you are, instead, the missile's intended recipient. As such, you have a keen interest in analyzing the missile's path as accurately as possible.

Maybe you have plotted the vertical position of the missile at five points in time, and it looks like this:



Now you do a quick linear regression, and you get great results. There's a line that passes almost exactly through the points you plotted:



(This is where your hand starts to creep, unthinkingly, toward the table saw's keening blade.)

Your line gives a very precise model for the missile's motion: for every minute that passes, the missile increases its altitude by some fixed

amount: say, 400 meters. After an hour it's 24 km above the earth's surface. When does it come down? It never comes down! An upward sloping line just keeps on sloping upward. That's what lines do.

(Blood, gristle, screams.)

Not every curve is a line. And the curve of a missile's flight is *most emphatically* not a line; it's a parabola. Just like Archimedes's circle, it looks like a line close up; and that's why the linear regression will do a great job telling you where the missile is five seconds after the last time you tracked it. But an hour later? Forget it. Your model says the missile is in the lower stratosphere, when, in fact, it is probably approaching your house.

The most vivid warning I know against thoughtless linear extrapolation was set down not by a statistician but by Mark Twain, in *Life on the Mississippi*:

The Mississippi between Cairo and New Orleans was twelve hundred and fifteen miles long one hundred and seventy-six years ago. It was eleven hundred and eighty after the cut-off of 1722. It was one thousand and forty after the American Bend cut-off. It has lost sixty-seven miles since. Consequently its length is only nine hundred and seventy-three miles at present. . . . In the space of one hundred and seventy-six years the Lower Mississippi has shortened itself two hundred and forty-two miles. This is an average of a trifle over one mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oolitic Silurian Period, just a million years ago next November, the Lower Mississippi River was upward of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing-rod. And by the same token any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three-quarters long, and Cairo and New Orleans will have joined their streets together, and be plodding comfortably along under a single mayor and a mutual board of aldermen. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact.

## ASIDE: HOW TO GET PARTIAL CREDIT ON MY CALCULUS EXAM

The methods of calculus are a lot like linear regression: they're purely mechanical, your calculator can carry them out, and it is very dangerous to use them inattentively. On a calculus exam you might be asked to compute the weight of water left in a jug after you punch some kind of hole and let some kind of flow take place for some amount of time, blah blah blah. It's easy to make arithmetic mistakes when doing a problem like this under time pressure. And sometimes that leads to a student arriving at a ridiculous result, like a jug of water whose weight is  $-4$  grams.

If a student arrives at  $-4$  grams and writes, in a desperate, hurried hand, "I screwed up somewhere, but I can't find my mistake," I give them half credit.

If they just write " $-4g$ " at the bottom of the page and circle it, they get zero—even if the entire derivation was correct apart from a single misplaced digit somewhere halfway down the page.

Working an integral or performing a linear regression is something a computer can do quite effectively. Understanding whether the result makes sense—or deciding whether the method is the right one to use in the first place—requires a guiding human hand. When we teach mathematics we are supposed to be explaining how to be that guide. A math course that fails to do so is essentially training the student to be a very slow, buggy version of Microsoft Excel.

And let's be frank: that really is what many of our math courses are doing. To make a long, contentious story short (but still contentious), the teaching of mathematics to children has for decades now been the arena of the so-called math wars. On one side, you have teachers who favor an emphasis on memorization, fluency, traditional algorithms, and exact answers; on the other, teachers who think math teaching should be about learning meaning, developing ways of thinking, guided discovery, and approximation. Sometimes the first approach is called *traditional* and the second *reform*, although the supposedly nontraditional discovery approach has been around in some form for decades, and whether "reform"